PUBH 8341 Notes

***Association ⇔ Correlation is different from Causation ⇔ Inference: The traditional scientific advancement is to find correlation in observational studies then testing the assumptions to find a causal link in RCTs (when possible within funding and ethical concerns)

Counterfactual thinking

Effect identification: Theory that explains the data; the effect is identifiable if you can get statistics that support your theory after getting rid of "competing explanations". You need to have quality data & plausible assumptions.

Elements of effect identification: Positivity (comparable by exposure) + exchangeability (same effect when we swap E and unE) + consistency (Same treatment among all exposed)

** Counterfactual Approach Model; A counterfactual is an alternative world where we could assign an exposure to an individual, observe the outcome, then go back in time and observe their outcome without ever having the exposure, all other things held equal.

4 Individual Response to Exposures (Potential outcomes frame)			
	Exposure	No Exposure	
Type 1: No Effect / Doom	1	1	
Type 2: Exposure is Causative	1	0	Counterfactual of Type 1
Type 3: Exposure is Preventative	0	1	
Type 4: Immunity	0	0	Counterfactual of Type 3

Assumptions: Dichotomous E, Dichotomous O, perfectly measured, no sampling variability, deterministic outcomes, no interference, treatment variation irrelevance.

You cannot infer anything about E if you only have the results of Type 1 and 2, or Type 3 and Type 4. However, if you assume that the individual represented by Type 1 and Type 2 / Type 3 and Type 4 are comparable, then you could set up a 2x2 table. We cannot estimate the individual causal effect, but we can estimate average causal effects.

Notations

A if for exposure (A=1 for exposed, A=0 for unexposed)

Y is for the outcome (Y=1 for exposed, Y=0 for unexposed)

In the counterfactual model, we can theoretically define individual outcomes in the potential outcomes frame, but we cannot estimate them: Therefore we find $P[Y=1|A=1]$ proportion of individuals with the outcome if everybody was exposed and $P[Y=1|A=0]$ proportion of individuals with the outcome if everybody was unexposed and we can find the following measures by comparing the two groups.

Other approaches to epidemiologic studies besides the potential outcomes counterfactual models

- potential outcomes counterfactual models (Looks at effects -> intervention)
- Bradford Hill/Sufficient Component Model (Looks at the outcome ->effects). Good to conceptualize interaction, impact of an intervention, but it is hard to measure these components directly.

Basic causal effects assumptions

- Effects with A
- Effects without A
- All other things held equal

Why are causal effects in randomized experiments a good substitute for the counterfactual?

Association RR = Pr[Y(a=1)=1|A=1] / Pr[Y(a=1)=1|A=0] ⇔ Pr[Y(a=1)]/ Pr[Y(a=0)] = Counterfactual RR

These 2 RR are comparable thanks to randomization which makes the concept of exchangeability acceptable on average in large populations / samples

Exchangeability under the counterfactual = Pr[Y(a=1)=1|A=1] (What we see) ⇔ Pr[Y(a=1)=1|A=0] (What we cannot observe)

Likewise: Pr[Y(a=0)=1|A=0] (What we see) ⇔ Pr[Y(a=0)=1|A=1] (What we cannot observe)

And furthermore: Pr[Y(a)=1|A=0] (among exposed) ⇔ Pr[Y(a)=1|A=0] (among unexposed) ⇔ Pr[Y(a)=1] (counterfactual)

What if A=1 Depends on another variable L? (which is a risk factor for Y)

In this case, Pr[Y(a=1)=1|A=0] (among exposed)=/= Pr[Y(a=0)=1|A=0] (among unexposed) so we need to stratify by L

Therefore: Pr[Y(a=1)=1|L=1] (among exposed) ⇔ Pr[Y(a=0)=1|L=0] (among unexposed)

In order to get to Pr[Y(a)=1] (counterfactual), we have to weight L so that it represents P[L] in the actual population: This is conditional exchangeability

Pr[Y(a)=1] = Pr[Y(a=1)=1|L=1] * P [L=1] (among exposed) ⇔ Pr[Y(a=0)=1|L=0] (among unexposed) * P [L=0]

In summary

Exchangeability: Pr[Y(a=1)=1|A=1] (What we see) ⇔ Pr[Y(a=1)=1|A=0] (What we cannot observe)

Positivity: Must have both Exposed and Unexposed participant in each level of measured confounders in the population otherwise we will not be able to have comparisons (i.e. estimate an effect)

Consistency: If A=a the Y=Y|a so the treatment is the same among ALL study participants

Modern Epidemiology Readings

Chapter 2

There are 2 meanings to randomization: Random assignment of an individual to a treatment (internal validity to avoid confounding and ensure comparability) + Random sampling of a group within a population (external validity for generalizability): Randomization in a large sample mechanically and statistically produces exchangeability

Ya U A is independence between the observed outcome and its counterfactual: Aka if you have randomization and you flipped treatment groups, you would see the same effects (exchangeability). We need this to be true to accurately estimate an effect (exchangeability under randomization).

Y U A is independence between the observed exposure and the observed outcome. If this is true, there can be no causality because the exposure does NOT affect the outcome when they are independent.

Chapter 3

Conditional randomization: Stratified analysis on another value L, so the exchangeability clause may not hold true, since the proportion of exposed and unexposed may be different within each strata.

In observational studies, you see the outcome first, then you go back to find the exposure.

In randomized experiments, you pick your exposure first, then you wait to see your outcome.

There is no causal effect without manipulation: after all, we want to see an effect that we could modify for our public health purposes.

Causal effects in experimental studies review

- What we want to observe: Pr[Y(a=1)=1|A=1]
- What is unobservable due to the counterfactual Pr[Y(a=1)=1|A=0]
- What we observed P[Y=1|A=1] ⇔ Pr[Y(a=1)=1|A=1]
- Thanks to randomization, we have exchangeability between Pr[Y(a=1)=1|A=1] ⇔ Pr[Y(a=1)=1|A=0]
- Potential Outcomes under exchangeability: Ya |_| A
- If P(A=1) depends on another variable L:
- P(A=1|L=1) =/= P(A=1|L=0) therefore P(Y(a=1)=1|A=1) =/=P(Y(a=1)=1|A=0)
- so you have to stratify by levels of L to have
- Pr[Y(a=1)=1|A=1, L=1] ⇔ Pr[Y(a=1)=1|A=0, L=0]

Causation vs association

P[Y=1|A=1] – P[Y=1|A=0] = association Risk Difference

P[Y=1|A=1] / P[Y=1|A=0] = association Relative Risk

[P[Y=1|A=1] / P[Y=0|A=1]] / [P[Y=1|A=0] / P[Y=1|A=0]] = association Odds Ratio

Enter Randomization, a mechanism which yields exchangeability : It generates data with missing values (unobservable counterfactual) but since the missingness is at random, there is exchangeability

Direct Acyclic Graphs

- A= Exposure
- Y= Outcome
- L= Covariate
- B= Stratifying

Causal Inference in a counterfactual notation (A is the only cause of Y)

$P[Y(a=0)=1] =/= P[Y(a=1)=1$ ⇔ Causality under the counterfactual

$P[Y=1|A=1] =/= P[Y=1|A=0]$ ⇔ Association under the observed data

Causal Inference in a counterfactual notation (A does not cause Y, but L is a cause of both A & Y)

$P[Y(a=0)=1] = P[Y(a=1)=1$ ⇔ No Causality under the counterfactual

$P[Y=1|A=1] =/= P[Y=1|A=0]$ ⇔ Association under the observed data

Causal Inference in a counterfactual notation (A doesn't cause Y, but L is a *collider* caused by A & Y)

$P[Y(a=0)=1] = P[Y(a=1)=1$ ⇔ No Causality under the counterfactual

$P[Y=1|A=1] = P[Y=1|A=0]$ ⇔ No Association under the observed data

Conditional Independence (A causes B which in turns causes Y)

$P[Y(a=0)=1] =/= P[Y(a=1)=1$ ⇔ Causality under the counterfactual with or without stratification on B

$P[Y=1|A=1] = P[Y=1|A=0]$ ⇔ No Association under the observed data if we stratify by B

Standardization VS IPW

Age	Stroke	N	Dementia	N	Dementia
65-74	No	15000	750	35000	3500
65-74	Yes	4000	1400	6000	1800
75+	No	18000	3600	12000	3000
75+	Yes	8000	4400	2000	1000
Total		45000	10150	55000	9300

A = Stroke, L = Age, Y = Dementia

Question 1

Standardization: ∑Pr[Y=1, A=a & L=l] * Pr[L=l] (there are 4 strata)

Standardization weight per strata = (n in treatment group of the strata + n in control group of the strata) / N in all study

For example, in the 65-74 strata, it will be (15000 + 35000)/(45000 + 55000) = 0.5

Age	Stroke	N	Dementia	N	Dementia	Standardization Weight	Risk Difference	∑Pr[Y=1, A=a & L=l] * Pr[L=l]
65-74	No	15000	750	35000	3500	0.5	-0.05	-0.025
65-74	Yes	4000	1400	6000	1800	0.1	0.05	0.005
75+	No	18000	3600	12000	3000	0.3	-0.05	-0.015
75+	Yes	8000	4400	2000	1000	0.1	0.05	0.005
Total		45000	10150	55000	9300			-0.03

The risk difference in dementia between the brain boost group and the placebo group is -0.03. The Brain boost group will have 3 fewer cases per 100 cases of dementia, compared to the placebo group.

Question 2

Inverse Probability Weighting = P(A=a given L=l) for cases and controls separately to create pseudo population. within a strata, it is n1/(n1+n2)
For example, in the 65-74 strata, it will be (15000)/ (15000 + 35000) = 0.3 weight for treatment group
For example, in the 65-74 strata, it will be (35000)/(15000 + 35000) = 0.7 weight for the control group
Then take the inverse of the weights (hence the inverse probability weighting name) to multiply the number of cases and participants in each strata (creates a pseudo population) which you can then use to calculate your measures of association regularly.

Age	Stroke	N	Dementia	N	Dementia	Weight for treatment	Weight for Placebo	Inverse Weight for Treatment	Inverse Weight for Placebo	Pseudo N for treatment group	Dementia in Treatment Group	Pseudo N for placebo group	Dementia in placebo Group
65-74	No	15000	750	35000	3500	0.3	0.7	3.333333333	1.428571429	50000	2500	50000	5000
65-74	Yes	4000	1400	6000	1800	0.4	0.6	2.5	1.666666667	10000	3500	10000	3000
75+	No	18000	3600	12000	3000	0.6	0.4	1.666666667	2.5	30000	6000	30000	7500
75+	Yes	8000	4400	2000	1000	0.8	0.2	1.25	5	10000	5500	10000	5000
Total		45000	10150	55000	9300					100000	17500	100000	20500

The risk difference in dementia between the brain boost group and the placebo group is -0.03. The Brain boost group will have 3 fewer cases per 100 cases of dementia, compared to the placebo group.

Question 3

Standardization: ∑Pr[Y=1, A=a & L=l] * Pr[L=l] (there are 4 strata)
Standardization weight per strata = n for treatment group in the strata + n for control group in the strata / N in all study

Age	Stroke	N	Dementia	N	Dementia	Standardization Weight	Risk Difference	∑Pr[Y=1, A=a & L=l] * Pr[L=l]
65-74	No	15000	750	35000	3500	0.5	-0.05	-0.025
65-74	Yes	4000	1400	6000	1800	0.1	0.05	0.005
75+	No	18000	3600	12000	3000	0.3	-0.05	-0.015
75+	Yes	8000	4400	2000	1000	0.1	0.05	0.005
Total		45000	10150	55000	9300			

For the 65 to 74 and No stroke strata and the over 75 and no stroke strata, the risk difference between the brainboost group and the placebo group was -0.05. The Brain boost group would have 5 fewer cases per 100 cases of dementia, compared to the placebo group. However for the 65 to 74 and Stroke strata and the over 75 and Stroke strata, the risk difference between the brainboost group and the placebo group was 0.05. The Brain boost group would have 5 more cases per 100 cases of dementia, compared to the placebo group.

Question 4

The effects estimated in question 1 is the absolute risk difference in dementia averaged over all four strata in our study, using the population as a standard. In the presence of exchangeability this measure can be used to estimate the (counterfactual) risk that would have been observed had everyone in the population taken brainboost. The effects estimated in question 2 represent the absolute risk difference in by creating a pseudo population in which we observed the outcomes if all participants had taken brainboost, and then we saw their outcomes if they had all taken placebo. Both standardization in question 1 and IPW in question 2 allowed us to simulate what would have been observed if we did not have different probability of being assigned in to strata based on Age and Stroke history in this study, and thus compute the average causal effects (assuming positivity, consistency and exchangeability) and gave us the same answer. In question three, we evaluated the absolute risk difference in each individual strata, taking into account that the vector composed of age and stroke status variables affected the probability of treatment.

Question 5

Based on the previous results, I would only recommend Brainboost among those 65 years and older if they have no previous history of stroke. Even though we observed an overall benefit in the absolute risk in dementia in our population, we consistently observed benefits (-0.05 *50000 + -0.05 * 30000 = 4000 fewer cases of dementia thanks to brain boost) among those without a stroke history, but no benefits for those with a history of a stroke (the positive risk difference implies that brain boost had no protective effect in the 65-74 and Stroke or the 75+ and Stroke groups). The only caveat would be that the observed risk differences would have to be collapsible over the age group categories so that they would remain the same if we were to look at Stroke history alone.

Measures of Incidence

Cumulative Incidence = # of new cases / # at risk = a+c / a+b+c+d

Incidence Rate = # of new cases / person-time at risk = a / PY1

Prevalence = # new cases + existing cases / Defined population

Mathematically, Prevalence = Incidence Rate * Duration

When the disease in rare (a and c are small) so:

CIR = (a/a+b) / (c/c+d) ~ (a/b) / (c/d) = OR

Measures of association

Absolute measures of association are differences, and are valuable for public health decision making and policy

- Absolute Risk = cumulative incidence in exposed - cumulative incidence in unexposed = CID
- Absolute Rate = incidence rate in exposed - incidence rate in unexposed = IRD

Relative measures of association are ratios, and are indicators of the strength of an association (etiological epi, where 1 exposure > 1 disease). Incidence Rates are better for dynamic populations, whereas cumulative incidence is better for fixed populations and is easier to interpret

- Relative Risk = cumulative incidence in exposed / cumulative incidence in unexposed = CIR
- Relative Rate = incidence rate in exposed / incidence rate in unexposed = IRR

Excess Risk: CI(in exposed) – CI(in unexposed) / CI(in unexposed) = (a/a+b) – (c/c+d) / (c+c+d) = Percentage in risk from baseline i.e. unexposed population as a result of the exposure

Measures of impact

Measures of impact are important to predict the impact of removing a particular exposure on the risk of developing an outcome.

- Rate Difference = CI1 – CI0 = CID. Therefore (CI1 – CI0)*(a+b) = CID*n(exposed) = number of new cases <u>among the exposed</u> due to the exposure

- Population Rate Difference: proportion exposed = (a+b) / (a+b+c+d) = prevalence of an exposure. PRD = (CI1 – CI0)*(proportion exposed) = CID*p1. Therefore PRD*N = CID*P1*N = number of new cases in the whole population that were due to the exposure
- Attributable proportion among the exposed = (CI1 – CI0)/CI1 = CIR – 1 / CIR. APe is the percentage of cases due to the exposure, among the exposed
- Attributable proportion among the total population = (CI - CI0)/CI = p(CIR – 1) / 1+p(CIR-1). APt is the percentage of cases due to the exposure, among the whole population

Randomization

Randomization helps us control for measured and unmeasured confounders by making their distribution among our exposed and unexposed groups similar.

- Simple randomization allows individual allocation to treatment or control group using simple mechanisms such as a coin flip or a randomization table
- Block randomization creates small groups (blocks) in which there are equal number or people in both treatment and control blocks. Balances number of people in each arm during recruitment
- Stratification followed by block randomization balances the stratification factor (age, severity of the disease) across treatment arms

Sensitivity

Sensitivity = a / a+c = proportion of people with the classification who are correctly identified

Specificity = b / b+d = proportion of people without the classification who are correctly identified

Population Standardization

	Sweden				Panama				Standardized Population			
Age	Deaths	Population	Proportion of Total Population	Rate / 100 k	Deaths	Population	Proportion of Total Population	Rate / 100 k	New Population	Adjusted Proportion of Pseudo Population	Adjusted Rate from Sweden	Adjusted Rate from Panama
<29	3523	3145000	0.42	112	3904	741000	0.689	527	3886000	0.453389336	1597.290631	1770.0319
29 - 59	10928	3057000	0.408	357	1421	275000	0.256	517	3332000	0.388752771	4248.290281	552.41768
>60	59104	1294000	0.172	4568	2456	59000	0.055	4163	1353000	0.157857893	9330.032902	387.69898
total	73555	7496000	1	981	7781	1075000	1	724	8571000	1	15175.61381	2710.1486

Crude Rate = # deaths * proportion f the total population in the stratum

We cannot directly compare the death rates of Sweden and Panama because the populations are different. Therefore, we have to go from crude rates to adjusted rate by creating weights from a standardized population

Standardized population = sum of the individual population

Adjusted Stratum specific weights = # of people in strata / total population

Adjusted rates = #death from individual population * adjusted stratum specific weight

Mantel Haenzel Estimates for Effect Modification

Crude Analysis			
Stratified Analysis			
Evaluate Effect Modification			
Yes	No		
Give stratum specific estimate	Evaluate Confounding		
	Yes	No	
	Give stratified or standardized estimates	Give crude estimate	

How to evaluate if there is effect modification

Calculate the stratum specific OR for each strata

- Ln(ORi) = ad/bc so OR = e^(ad/bc)
- Var(Ln(ORi) = (1/a + 1/b + 1/c + 1/d)

Calculate the mantel-haenszel summary estimator for all combined strata

- Mantel Haenszel Estimator = ∑(Wi * RR) / (Wi)
- Wi = c * (a+b) / (a+b+c+d) = ci *n1i / Ni
- RRi = (a/a+b) / (c/c+d)

Calculate the test for homogeneity across strata

- X^2 = ∑[Ln(ORi) – Ln(Mantel-Haenszel Estimator)] / Var(Ln(ORi)
- If P-value is statistically significant, there is effect modification.

Counterfactuals & Causal Inference

Michael Oakes
Division of Epidemiology and Community Health
School of Public Health
University of Minnesota
PubH 8341

Rothman I - Sufficient-component cause model

- A cause of a disease is an event, condition, or characteristics that plays an *essential role* in producing an occurrence of the disease.
- Constellation of components that *act in concert*.
- The Rothman pie!
- Sufficient cause: A set of minimal conditions and events that inevitably produce disease.
- Component cause: An individual event, condition, or characteristic required by a given sufficient cause.
- Necessary cause: A component cause present in every sufficient cause.
- Typically, there are many sufficient causes for any given effect.

Sufficient-component cause model

- A.k.a. Sufficient-cause model or Rothman's model
- Its premise:
  - "Diseases" have multiple causes, and each cause has different components

3

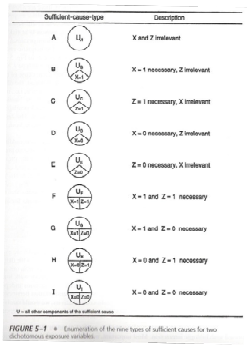


FIGURE 5–1 • Enumeration of the nine types of sufficient causes for two dichotomous exposure variables.

Effect Identification

- *Identification* means, loosely, that effects are discernable… no other explanation could explain the data
- Identification is not estimation; assume infinite sample size in identification analysis
- An effect is identifiable if it is theoretically possible to learn the true value of the parameter when the sample size approaches infinity.
- Imagining that your sample size is infinitely large eliminates all problems of statistical uncertainty, confidence intervals and p-values.

Identification is about ruling out competing hypotheses or explanations.

Apart from statistical imprecision, if more than one explanation for your effect (eg, difference b/w treatment and control) exists, then you have an identification problem. You cannot say X_1 caused Y because the cause might have been X_2, or X_3, or X_n .

It's all always about data and assumptions

Got quality data?

Got plausible assumptions?

The real task

Rule out competing explanations!

Done with well-executed experiments

&

Done with well done case studies

Elements of Identification

- *Positivity*
- *Exchangeability*
- *Consistency*

So causal inference is all about finding the best counterfactual substitute for the unobservable counterfactual scenario.

The best ones are said to be *exchangeable*.

Randomization is a good mechanism to produce a group of subjects that are exchangeable.

Simply, it's all about the comparison group!

Causal Effect

$$\Delta = \overline{Y}_{Treatment} - \overline{Y}_{Control}$$

Potential Outcomes

(causal effects defined across rows)

Condition Assigned	Outcome if Treated	Outcome if not Treated
Treatment	Observed	Unobservable Counterfactual
Control	Unobservable Counterfactual	Observed

Potential Outcomes

(a missing data problem)

Condition Assigned	Outcome if Treated, Y^T	Outcome if NOT Treated, Y^C
Treatment	$Y^T \mid T$	$Y^C \mid T$
Control	$Y^T \mid C$	$Y^C \mid C$

Intuitive definition of cause

- Basic causal thinking:
  1. What happened when A was present?
  2. What happened when A was not present?
  3. All other things being equal
- If 1 and 2 are different, we say that A has a (causal) effect

13

Assumptions

1. Dichotomous exposures
2. Dichotomous outcomes
3. Perfectly measured
4. No sampling variability (infinite populations)
5. Deterministic outcomes
6. No interference
7. Treatment variation irrelevance
   (More later on assumptions 6 and 7)

14

Some notation

- We use A for the exposure
  - A=1 if exposed (taking the pill), A=0 otherwise
- Y for the outcome
  - Y=1 if outcome present (dying five days later), Y=0 otherwise
- A^i is the exposure level for individual *i* and Y^i is the outcome for that same individual

15

Individual causal effects

- We say that A has a (causal) effect on Y in the individual i if:
  - $Y^{i,a=0} \neq Y^{i,a=1}$
- We say that A does not have a (causal) effect on Y in the individual i if:
  - $Y^{i,a=0} = Y^{i,a=1}$

16

Observed and potential (counterfactual) outcomes for Donald and Jeb

	A^i	Y^i	$Y^{i,a=0}$	$Y^{i,a=1}$
Donald	1	1	0	1
Jeb	0	0	0	0

17

Individual causal effects

- Problem:
  - In general, we are not able to see the outcome in the same individual after two different exposures
- Is there any solution?

	A^i	Y^i	$Y^{i,a=0}$	$Y^{i,a=1}$
Donald	1	1	?	1
Jeb	0	0	0	?

18

Interference

- Assumption we have been making: a subject's counterfactual outcome does not depend on other subject's treatment or outcome
  - Rubin called it 'stable unit treatment value assumption' (SUTVA)
- Problematic in some settings
- In the presence of multiple versions of treatment, the counterfactual Y_i^a for an individual *i* is not well-defined because it depends on the version of treatment a.
  - Considered part of SUTVA
  - Weaker assumption: "treatment variation irrelevance", i.e. multiple versions of treatment A=*a* exist but all result in the same outcome Y_i^a

19

Average causal effects

- We define **$Pr[Y^{a=0}=1]$** as the proportion of individuals who would have developed the outcome in the population had everybody been unexposed.
- We define **$Pr[Y^{a=1}=1]$** as the proportion of individuals who would have developed the outcome in the population had everybody been exposed.
- Alternative notation:
  - Pr(Y=1|SET[A=1]) and Pr(Y=1|SET[A=0])

20

Average causal effects

- We say that A does not have an effect on Y in the population if:
  - $Pr[Y^{a=0}=1] = Pr[Y^{a=1}=1]$
- Similarly, we say that A does have an effect on Y in the population if:
  - $Pr[Y^{a=0}=1] \neq Pr[Y^{a=1}=1]$
- No average causal effect doesn't imply absence of individual causal effects
  - If $Y^{i,a=0} = Y^{i,a=1}$ is true for all i, we say that the sharp null hypothesis is true

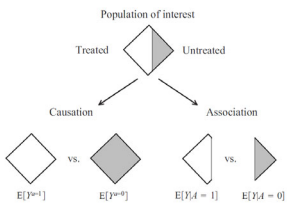
21

Average causal effects: effect measures

- Causal risk difference:
  - $Pr[Y^{a=1}=1] - Pr[Y^{a=0}=1] \neq 0$
- Causal risk ratio:
  - $Pr[Y^{a=1}=1] / Pr[Y^{a=0}=1] \neq 1$
- Causal odds ratio:
  - $(Pr[Y^{a=1}=1]/(1-Pr[Y^{a=1}=1])) / (Pr[Y^{a=0}=1]/(1-Pr[Y^{a=0}=1])) \neq 1$

22

To summarize…



23

Causal effects in observational studies

Alvaro Alonso
Division of Epidemiology and Community Health
School of Public Health
University of Minnesota
PubH 8341 Fall 2015

24

Outline

- Estimation of causal effects in randomized experiments
- Estimation of causal effects in observational studies
- The concepts of interference, positivity and consistency

25

Goal

- Randomize a very large population to treatment (A=1) or placebo (A=0)
- Determine the outcome (e.g. death) during follow-up (Y=1 or Y=0)
- Interested in the effect of A on Y in this population

$$aRR = \frac{\Pr[Y=1|A=1]}{\Pr[Y=1|A=0]} \qquad cRR = \frac{\Pr[Y^{a=1}=1]}{\Pr[Y^{a=0}=1]}$$

26

Exchangeability

- Because there is randomization, those receiving treatment or placebo, on average, have the same characteristics but for the assigned exposure
- Therefore, the mortality among those exposed to treatment would be the same as the mortality if those exposed to placebo had been exposed to treatment
- Using counterfactual notation:

  $Pr[Y^{a=1}=1|A=1]=Pr[Y^{a=1}=1|A=0]$

  and

  $Pr[Y^{a=0}=1|A=0]=Pr[Y^{a=0}=1|A=1]$
- That is, because of **randomization**, exposed and unexposed are **exchangeable**

27

Putting all together…

- By definition:
  - $Pr[Y=1|A=1]=Pr[Y^{a=1}=1|A=1]$
- Because of exchangeability:
  - $Pr[Y^{a=1}=1|A=1]= Pr[Y^{a=1}=1|A=0]= Pr[Y^{a=1}=1]$
- Therefore:
  - $Pr[Y=1|A=1] = Pr[Y^{a=1}=1]$
- (And the same for A=0)
  - $Pr[Y=1|A=0] = Pr[Y^{a=0}=1]$

28

Implications

- In an ideal randomized trial, association is causation:

$$aRR = \frac{\Pr[Y=1|A=1]}{\Pr[Y=1|A=0]} = cRR = \frac{\Pr[Y^{a=1}=1]}{\Pr[Y^{a=0}=1]}$$

29

Exchangeability

- In counterfactual notation, exchangeability defined as:
- If treated ar $Y^a \coprod A \text{ for all } a$ e, the treatment is exogenous.
  - Exchangeability also known as *exogeneity*

30

A different kind of randomized experiment

- Probability of receiving A=1 is different according to levels of another variable L (which is a risk factor for Y, the outcome)
  - If L=1, then 75% are assigned to A=1
  - If L=0, then 50% are assigned to A=1
  - Pr[A=1|L=1] = 0.75 ≠ Pr[A=1|L=0] = 0.5

31

Conditional exchangeability

- By definition:
  - $Pr[Y=1|A=1,L=1] = Pr[Y^{a=1}=1|A=1,L=1]$
- Because of exchangeability:
  - $Pr[Y^{a=1}=1|A=1,L=1] = Pr[Y^{a=1}=1|A=0,L=1] = Pr[Y^{a=1}=1|L=1]$
- And the same for A=0
  - $Pr[Y=1|A=0,L=1] = Pr[Y^{a=0}=1|A=0,L=1] = Pr[Y^{a=0}=1|A=1,L=1]=Pr[Y^{a=0}=1|L=1]$
- We could do this for L=0, too:

  $Pr[Y=1|A=1,L=0] = Pr[Y^{a=1}=1|A=1,L=0] = Pr[Y^{a=1}=1|A=0,L=0] = Pr[Y^{a=1}=1|L=0]$

  and

  $Pr[Y=1|A=0,L=0] = Pr[Y^{a=0}=1|A=0,L=0] = Pr[Y^{a=0}=1|A=1,L=0] = Pr[Y^{a=0}=1|L=0]$

32

Conditional exchangeability

- In general:

$$\Pr[Y^a = 1] = \sum_l \Pr\left[Y^a = 1 \middle| L = l\right] \times \Pr[L = l]$$

- By conditional exchangeability:

$$\Pr\left[Y^a = 1 \middle| L = l\right] = \Pr\left[Y^a = 1 \middle| A = a, L = l\right]$$

- By definition of counterfactual:

$$\Pr\left[Y^a = 1 \middle| A = a, L = l\right] = \Pr\left[Y = 1 \middle| A = a, L = l\right]$$

33

Conditional exchangeability

- We have $Pr[Y^{a=1}=1|L=1]$ and $Pr[Y^{a=1}=1|L=0]$ but can we estimate $Pr[Y^{a=1}=1]$?
- From probability theory (law of total probability):

  $Pr[Y^{a=1}=1]=Pr[Y^{a=1}=1|L=1] \times Pr[L=1]+Pr[Y^{a=1}=1|L=0] \times Pr[L=0]$ (Weighted average)
- From the previous slide it infers that we can use observed data to estimate a counterfactual risk:

$$\Pr[Y^a = 1] = \sum_l \Pr\left[Y = 1 \middle| A = a, L = l\right] \times \Pr[L = l]$$

34

Conditional exchangeability in observational studies

- Because we have conditional exchangeability within levels of L, we can estimate causal effects in this particular type of randomized experiment
- Now, let's assume that we do not randomize exposure according to levels of L, but we observe the probabilities of exposure within levels of L
- We could assume that L is the only factor determining exposure in an observational study
- If we are willing to **assume** conditional exchangeability within levels of L, we could estimate causal effects in observational studies
- Strongly ignorable treatment assignment (Rosenbaum and Rubin)

35

Positivity

- Positivity requires that there be both exposed and unexposed at every combination of the observed confounders in the population under study
- Example:
  - Clinical trial of HIV patients in which those with CD4<200 always receive treatment
  - We couldn't estimate causal effect among those with CD4<200

36

Consistency

- If observed treatment A=a then $Y=Y^a$
- If $Y=Y^a$, we say that the observed outcome and the counterfactual outcome under the observed exposure are *consistent*
- Isn't this always the case?

37

Interference

- Assumption we have been making: a subject's counterfactual outcome does not depend on other subject's treatment or outcome
  - Rubin called it 'stable unit treatment value assumption' (SUTVA)
- Problematic in some settings

38

Causal diagrams in epidemiology

Alvaro Alonso
Division of Epidemiology and Community Health
School of Public Health
University of Minnesota
PubH 8341 Fall 2015

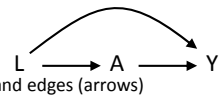
39

Why causal diagrams?

- The counterfactual approach provides formal definitions of causal effects and its identifying conditions, and is useful to derive methods to identify those effects
  - This can get cumbersome
- Causal diagrams are better to conceptualize problems
- Simple way to encode subject-matter *knowledge* and a priori *assumptions* about the *qualitative* structure of a problem
- Common language to represent causal relations
  - Enhance communication among investigators

40

Which diagrams?

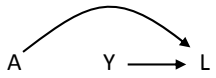


- Nodes (variables) and edges (arrows)
- [Time flows from left to right]
- Absence of arrow: no effect
- Presence of arrow: effect (or unwilling to assume no effect)

41

Causal diagrams and marginal independence / association

$Pr[Y^{a=0}=1] = Pr[Y^{a=1}=1]$
$Pr[Y=1|A=1] = Pr[Y=1|A=0]$

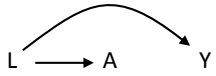


L: common effect of A and Y or *collider*

42

Confounding

$Pr[Y^{a=0}=1] = Pr[Y^{a=1}=1]$
$Pr[Y=1|A=1] \neq Pr[Y=1|A=0]$

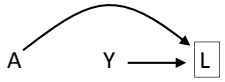


Confounding as presence of backdoor paths

43

Selection bias

$Pr[Y^{a=0}=1] = Pr[Y^{a=1}=1]$
$Pr[Y=1|A=1,L=\ell] \neq Pr[Y=1|A=0,L=\ell]$



44

Confounding and confounders

- What is a confounder?
- Standard definition:
  1. A confounder must be a risk factor (associated) for the disease conditional on the treatment
  2. A confounder must be associated with the exposure in the source population
  3. A confounder must not be affected by exposure or outcome, can't be an intermediate

45

What is a confounder?

- "Extraneous factors responsible for differences in disease frequency between the exposed and unexposed (or factors associated with those extraneous factors)" (Rothman)
- "Any variable that can be used to help eliminate confounding bias" (Hernán)

46

Methods to adjust for confounding

- In the design:
  - Randomization
  - Restriction, matching
- In the analysis:
  - Stratification-based methods
  - G-methods: standardization, IPW, g-estimation

47

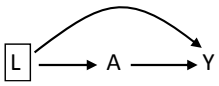
Randomization

- Removes measured and unmeasured confounding

48

Stratification-based methods

- Stratification (including multivariable modeling)
- Restriction
  - $Pr[Y_a=1|L]$



49

Folic acid suppl and neural tube defects

TABLE 3. Periconceptional folic acid supplementation (*E* = 1) and neural tube defects (*D* = 1), Slone Epidemiology Unit Birth Defects Study, 1992–1997

	D = 1	*D* = 0
E = 1	43	239
E = 0	194	704

TABLE 4. Periconceptional folic acid supplementation (*E* = 1) and neural tube defects (*D* = 1), stratified by the covariate *C*, Slone Epidemiology Unit Birth Defects Study, 1992–1997

	C = 1		*C* = 0	
	D = 1	*D* = 0	*D* = 1	*D* = 0
E = 1	19	8	24	231
E = 0	100	46	94	658

- OR_{ED}=0.65
- $OR_{ED|C}$=0.80

1. Is C a confounder?
   1. 10% rule
   2. Stepwise model
   3. Standard definition:
      1. Risk factor for D in unexposed?
      2. Associated with E in the source population?
      3. Intermediate variable?
2. Should we adjust for C?

Am J Epidemiol 2002;155:176–84.

Measurement bias

Alvaro Alonso
Division of Epidemiology and Community Health
School of Public Health
University of Minnesota
PubH 8341 Fall 2015

51

Nondifferential misclassification of exposure

- Expected towards the null with dichotomous exposures
- Might bias away from the null or even reverse a trend if more than two levels

52

Nondifferential misclassification of outcome

- In most situations (binary outcomes), it is expected to bias toward the null, provided that the misclassification is independent of other errors
  - Exception: risk ratio scale, with perfect specificity

53

Dependent errors

- When both exposure and disease are nondifferentially misclassified but the classification errors are dependent, it is possible to be biased away from the null
- Possible in information obtained by interviews or self-administered questionnaires

54

Misclassification of confounders

- Independent nondifferential misclassification of a dichotomous confounder will lead to residual confounding
- If the misclassification is differential or dependent, there is additional distortion
- Misclassification of a confounder may result in apparent effect modification

55

Non-differential misclassification: expectation vs observation

- Even when non-differential misclassification is expected towards the null, actual results might be away from the null

56

Selection bias

Alvaro Alonso
Division of Epidemiology and Community Health
School of Public Health
University of Minnesota
Fall 2015

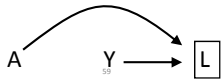
Definition

- "Selection biases are distortions that result from procedures used to select subjects and from factors that influence study participation" (Rothman)
- "Biases that arise from the procedure by which individuals are selected into the analysis" (Hernán)
- Lack of generalizability is not selection bias

58

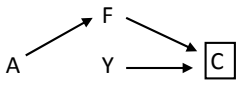
Selection bias

- Structural definition:
  - "Biases that arise from conditioning on a common effect of 2 variables, 1 of which is either the treatment or a cause of treatment, and the other is the outcome or a cause of the outcome" (Hernán)
  - Collider-stratification bias



59

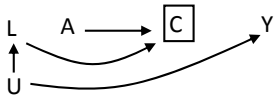
Inappropriate selection of controls in case-control studies



- A: postmenopausal hormones
- Y: coronary heart disease
- F: hip fracture
- C: selection into study (1=yes, 0=no)

60

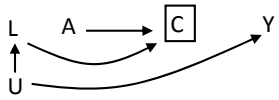
Differential loss to follow-up in cohort studies



- A: statins
- Y: dementia
- C: censoring (1=yes, 0=no)
- U: cognitive function
- L: cognitive symptoms

61

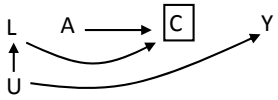
Volunteer bias / self-selection bias



- A: participation in community service
- Y: intimate partner violence
- C: participation in the study (1=yes, 0=no)
- U: suffering childhood victimization
- L: "trauma" awareness

62

Healthy worker bias (I)



- A: occupational exposure
- Y: death
- L: results from physical exam, blood test
- U: health status
- C: being at work (1: yes, 0: no)

63

Healthy worker bias (II)

Study comparing mortality in a group of workers with the general population:

- A: being a lumberjack
- Y: death
- L: health status

This is not selection bias, but confounding (common cause)

64

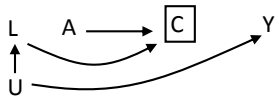
Control of selection bias

- Design
  - Appropriate selection of controls
  - Increase retention
- Analysis
  - Stratification
  - Multiple imputation
  - IPW
  - Use of external information

65

Stratification to control for selection bias

- We could block the path opened when conditioning on the collider C
  - If we have information on L or U



$$\Pr[Y=1|A=1,C=0,L=l] / \Pr[Y=1|A=0,C=0,L=l]$$

IPW to control for selection bias

- Using IPW to control for confounding, we created a pseudo-population in which exposed and unexposed were exchangeable

67

Interaction / effect modification

Alvaro Alonso
Division of Epidemiology and Community Health
School of Public Health
University of Minnesota
PubH 8341 Fall 2015

68

Effect modification vs interaction

- Interaction
  - Effect of a treatment depends on the presence of one or more other conditions.
  - The causal effects of two exposures together are of interest
- Effect (measure) modification
  - When no bias is present, statistical interactions correspond to effect measure modification
  - The causal effect of one exposure within strata of another exposure is of interest

Rothman, Greenland, Lash. Modern Epidemiology 2008

69

Interactions in epidemiology

- Two dichotomous exposures (X, Z)
- Dichotomous outcome Y
- R_{11} counterfactual outcome when X=1 and Z=1 (i.e. $Y_{x=1,z=1}$)
- $R_{11} - R_{01}$ and $R_{10} - R_{00}$ are effects of changing from X=0 to X=1
- $R_{11} - R_{10}$ and $R_{01} - R_{00}$ are effects of changing from Z=0 to Z=1
- $(R_{11} - R_{01}) - (R_{10} - R_{00}) = (R_{11} - R_{10}) - (R_{01} - R_{00})$
- Interaction contrast (IC) = $R_{11} - R_{10} - R_{01} + R_{00}$

70

Sufficient cause interaction

- Unlike the counterfactual definition of interaction, sufficient cause interaction makes explicit reference to the causal mechanisms involving the treatments X and Z
- Different combinations of susceptibilities to sufficient causes may produce the same response type, so that the sufficient cause model is a "finer" model than the potential outcome model
  - A person at risk of sufficient causes B and C, but no other, will be type 2
  - A person at risk of sufficient causes F, G and H, but no other, will be also type 2

71

Analysis of interaction

- Choose an effect measure
- Look at possible modifiers for which you have some prior support
- Test for interactions
- Look at interaction contrasts

72

Analysis of interaction

- A departure from additivity implies causal interaction
- Presence of additivity doesn't imply absence of causal interaction
- You don't get to find out what type of interaction is occurring

73

Direct Effects, Indirect Effects and Mediation

What can we do?

What should we do?

Part I.

Decomposing additive effects when effect measure modification on the additive scale is not present and all confounders are controlled

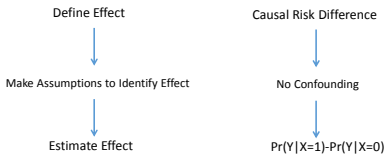
What is Mediation Analysis?

- An effect decomposition technique that allows a researcher to estimate what amount of the effect of an exposure, X, on outcome, Y, is mediated by M
- Steps in Baron Kenny Decomposition:
  1. Make a bunch of assumptions
  2. Estimate Total Effect (TE), Direct Effect (DE), Indirect Effect (IE)
  3. Mediation Proportion=IE/(IE+DE)=IE/TE
- If 100% of the effect is mediated, X has no effect on Y except through M
- If 0% of the effect is mediated, X only has an effect on Y that is not through M

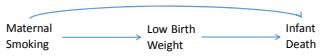
Mediation in Epidemiology

- A large part of public health research involves etiologic investigations
  - Epidemiologists want to say that an exposure is associated with disease
    - How much higher are heart disease rates among smokers than among non-smokers
  - But just as often, we want to identify "pathways" through which an exposure works
    - How much of the increase in heart disease rates is due to the increase in hypertension among smokers
- The latter are mediation questions!
  - Is the effect of family meals on obesity because families who eat together frequently have better diet quality?
  - How much of the effect of smoking on neonatal mortality is because smoking leads to low birth weight infants?

Inference in Public Health



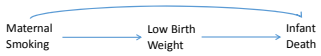
Total Effect



- Average causal effect:

  Pr(Death=1|Set[Smoke=1])-Pr(Death=1|Set[Smoke=0])
  - This is a counterfactual notion of causation: What happens to the probability of death in this cohort if I first force all mothers to smoke and then travel back in time to force all mothers not to smoke
  - Amazingly, this is something I can estimate given the right data and the right set of assumptions

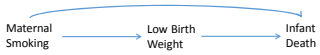
Controlled Direct Effect



CDE(m)=Pr(Death=1|Set[Smoke=1], Set[LBW=m])-
Pr(Death=1|Set[Smoke=0], Set[LBW=m])

- The difference in the probability of neonatal death between smokers and non smokers if we force everyone in the population to a certain LBW status.
- Notice there are TWO direct effects, one in each stratum of low birth weight
  - Also note that CDE(0) is impossible to imagine

Controlled Indirect Effect



- A manipulative definition of CIE is not straightforward
- Loosely, the proportion of infants who would die if mothers smoked but not if their mothers didn't smoke, but only through pathways involving LBW
  - A definition of exclusion: the indirect effects are all the effects that aren't direct effects.

Assumption 1: No Effect Modification

- Baron & Kenny required 3 effects (total, direct, indirect)
- We just defined 4: TE, CDE(0), CDE(1), CIE
- The first BIG assumption that allows Barron and Kenny to work: No effect measure modification on the additive scale: CDE(1)=CDE(0)
- That leaves us with: TE, CDE, CIE

No EMM for the risk difference

- Suppose CDE(1)= 0.25, CDE(0)= -0.25, CIE=0.25
- TE=CDE(1)+CDE(0)+CIE=0.25
  - Mediation proportion=IE/TE=0.25/0.25
    - Implies no direct effect, which isn't true!
- Sensible results are only obtained with no EMM for the risk difference:
- TE=CDE+CIE

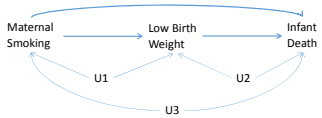
Important Caveat

- More precisely, there can be no *unit level* interactions between exposure and mediator
- There are no ***individuals*** in the dataset, for example, who
  - get the disease only if exposure=1 and mediator=1
  - don't get the disease only if exposure=1 and mediator=1
- This is something we don't get to observe!
- Instead, we observe average interactions. Because the unit level interactions can average out, observing no interaction in the data does not necessarily imply there are not unit level interactions
  - However observing interactions does imply that unit level interactions are present

Estimation

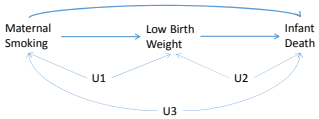
- For this decomposition to work, we need to be able to estimate the TE, CDE, CIE without bias
  - They need to be identified in our dataset
- We identify parameters in the standard way:
  - make guesses about the true state of nature
- These identifying assumptions are the standard 'no confounding' assumptions we're used to but with a few twists

Assumption 2&3: Identifying the TE



- No unmeasured confounding between exposure (smoking) and outcome (death)
  - U1 and U3 don't exist, or you have controlled for them

Assumption 4: Identifying the CDE



- No unmeasured confounding between intermediate (LBW) and outcome (death)
  - U2 doesn't exist, or you have controlled for it
  - The assumption regarding U2 was initially pointed out in the psychometrics literature, but subsequently ignored
    - Judd and Kenny (1981) vs Baron and Kenny (1986)

Identifying the CIE

- This comes for free!
- After assuming no effect measure modification

  TE=CDE+CIE
- Once I can identify TE and CDE, I automatically know the CIE.
- Now we can proceed with decomposition
  - Mediation proportion=CIE/TE

Baron-Kenny Decomposition, I

- $E(Y)=\beta_0+ \beta_1X + \beta_2M$
- $E(Y)=\alpha_0+ \alpha_1X$
- Total Effect: α_1
- Direct Effect: β_1
- Indirect Effect: $\alpha_1-\beta_1$
- Mediation proportion: $(\alpha_1-\beta_1)/(\alpha_1)$

Baron-Kenny Decomposition, II

- $E(Y)=\beta_0+ \beta_1X + \beta_2M$
- $E(M)=\delta_0+ \delta_1X$
- Direct Effect: β_1
- Indirect Effect: $\delta_1\beta_2$
- Total Effect: $\beta_1+\delta_1\beta_2$
- Mediation proportion: $\delta_1\beta_2/(\beta_1+\delta_1\beta_2)$

Example

	LBW=0 (>2500g)		LBW=1 (<2500g)	
	Smoker	Non-Smoker	Smoker	Non-Smoker
Live Birth	353,335	2,453,633	40,383	137,154
Infant Death	1729	5838	2192	9387
Risk	.0049	.0024	.0515	.0641

1997 birth certificate data linked to Infant mortality files from the National Center for Health Statistics

Reproduced from VanderWeele et al "Conditioning on intermediates in Perinatal epidemiology" Epidemiology 2012

Example – Calculate Effects

	LBW=0 (>2500g)		LBW=1 (<2500g)	
	Smoker	Non-Smoker	Smoker	Non-Smoker
Live Birth	353,335	2,453,633	40,383	137,154
Infant Death	1729	5838	2192	9387
Risk	.0049	.0024	.0515	.0641

TE = Pr(Death|Set[Smoke=1]-Pr(Death|Set[Smoke=0]
= 3921/393,718-15225/2,590,787
= 0.004

Example – Calculate Effects

	LBW=0 (>2500g)		LBW=1 (<2500g)	
	Smoker	Non-Smoker	Smoker	Non-Smoker
Live Birth	353,335	2,453,633	40,383	137,154
Infant Death	1729	5838	2192	9387
Risk	.0049	.0024	.0515	.0641

CDE(1)=CDE(0)=CDE
= Pr(Death|Set[Smoke=1],Set[LBW=m])-Pr(Death|Set[Smoke=0],Set[LBW=m])
= 0.001*

*Using the Mantel-Haenszel formula to combine stratum specific estimates

Example- Calculate Effects

- TE=0.004
- CDE=0.001
- CIE=TE-CDE=0.004-0.001=0.003
- Mediation proportion: 0.003/0.004=0.75

- 75% of the effect of smoking on infant mortality is due to the fact that infants of smokers are more likely to be LBW and LBW infants are more likely to die

Part I Summary

- You can decompose effects without anyone trying to kill you if:
  1. You work on an additive scale (Risk Difference; Difference in means)
  2. There is no effect measure modification on the additive scale
  3. There is no unmeasured confounding of the exposure-disease
  4. There is no unmeasured confounding intermediate-disease relationships.

P-values, hypothesis testing, confidence intervals and all that stuff

Michael Oakes

PubH 8341 – Fall 2015

96

Background…

- Fisher used p-value as an index of discrepancy between data (ie, effect estimate) and null hypothesis;
- Smaller the p-value, the greater the discrepancy
- Size of p-value indicated strength of evidence
- Fisher advocated P=0.05 as a good threshold for rejecting null

Background…

- Neyman & Pearson thought Fisher's approach too subjective…
  They wanted and produced an objective decision-theoretic approach
- By fixing error rates (I & II) in advance the number of mistakes over repeated experiments would be limited
- It's an all or nothing approach; results are significant or not; no strength of evidence in a given trial
- To properly interpret results one must know *a priori* power calculations and sample size
- Type I error is to be fixed in advance and does not depend in anything, including sample size

P-values today

- Probability under (hypothetical) repeated experiments of obtaining an teststatistic at least a far from a specified value (eg, 0.00) if that value were in fact true.
- It is a term used in frequentist/classical statistics
- p-value is *not the probability* that a given (alternative) hypothesis is true

Frequentist probability

- Based on repeated / repeatable events
- Aims to be "objective", divorced from any consideration of personal factors and amenable to practical demonstration through experimentation (ie, sample data)
- Invariant to circumstance or context, or time
- It's the 'Law of Large Numbers' and CLT

Frequentist probability

- Based on repeated / repeatable events
- Aims to be "objective", divorced from any consideration of personal factors and amenable to practical demonstration through experimentation (ie, sample data)
- Invariant to circumstance or context, or time
- It's the 'Law of Large Numbers' and CLT
- We very much want to know the probability that our hypothesis is true. But p-values cannot tell us this. We must use Bayesian methods to answer such questions.
- To interpret 95% CIs as having 95% probability of including the parameter of interest is a mistake; again, this mixes Bayesian interpretations with frequentist.

Criticisms

- For over 50 years every major methodologist, it seems, has rejected the use of p-values and significance testing
- Statistical significance says nothing about importance, utility or meaningfulness; effects with small p-values may be meaningless and imply precise but negligible effects

If you reject under a 2-sided test you get to make the statement that your test statistic is greater than *or* less than the null, but you have no information about which
A large p-value does not mean null is true; it means you cannot reject it; the effect best supported by the data is always the observed effect, regardless of its significance

Cautions

- P-values don't imply the strength of relationships; magnitude and precision must matter
- Don't compare p-values from tests across experiments and say one is more significant than the other
- Don't compare p-values from tests in same study and say one is more significant than the other

Confidence intervals

- If all other assumptions are met, we can say that under repeated runs of the experiment, the 95% CI will contain the true population parameter (ie, effect) in 95% of the time
  - Again, what is true is that if we run experiment over and over, we can expect the true value of the parameter to be included in the computer interval 95% of the times… This is not the probability of the hypothesis
  - When interpreting CIs, it is essential to realize that the parameter values in the center of the interval are no more likely than those near the bounds; all are equally likely

Table 1 Twelve P-Value Misconceptions

1	*If P = .05, the null hypothesis has only a 5% chance of being true.*
2	*A nonsignificant difference (eg, P ≥.05) means there is no difference between groups.*
3	*A statistically significant finding is clinically important.*
4	*Studies with P values on opposite sides of .05 are conflicting.*
5	*Studies with the same P value provide the same evidence against the null hypothesis.*
6	*P = .05 means that we have observed data that would occur only 5% of the time under the null hypothesis.*
7	*P = .05 and P ≤.05 mean the same thing.*
8	*P values are properly written as inequalities (eg, "P ≤.02" when P = .015)*
9	*P = .05 means that if you reject the null hypothesis, the probability of a type I error is only 5%.*
10	*With a P = .05 threshold for significance, the chance of a type I error will be 5%.*
11	*You should use a one-sided P value when you don't care about a result in one direction, or a difference in that direction is impossible.*
12	*A scientific conclusion or treatment policy should be based on whether or not the P value is significant.*

105

Randomized Controlled Trials

Michael Oakes
Division of Epidemiology and Community Health
School of Public Health
University of Minnesota
PubH 8341

One of the first published RCT in medicine appeared in the 1948 paper entitled

"Streptomycin treatment of pulmonary tuberculosis",

One of the authors of that paper was Sir Austin Bradford Hill, who is credited as having conceived the modern RCT… and a founding father of modern epidemiology.

Randomized Controlled Trials

- Randomize individuals to treatment or control conditions
- Intervene in treatment condition(s)
- Measure endpoints at given time periods

Advantages

- With LARGE numbers we expect no confounding between conditions.
- Helpful for causal claims (generally more tenable assumptions)
- Typically manipulate one factor and thus answer one basic question, though more sophisticated approaches are possible.

Disadvantages

- Usually expensive ($$$ and Time)
- Ethical concerns
- Not always possible (e.g., nutrition, social policy)
- Tend to screen out "problem" subjects, such as the very young, the elderly and pregnant and lactating women

Stages of Clinical Trials for Drug/Device Evaluation

- Basic science
- Animal models
- Phase I Trial: dose-finding
- Phase II Trial: preliminary evidence of efficacy
- Phase III Trial: comparisons to standard therapy
- Phase IV Trial: post-marketing surveillance

Efficacy & Effectiveness

RCTs test efficacy in a research setting with highly selected participants and under highly controlled conditions.

Pragmatic RCTs test effectiveness in everyday practice with relatively unselected participants and under flexible conditions

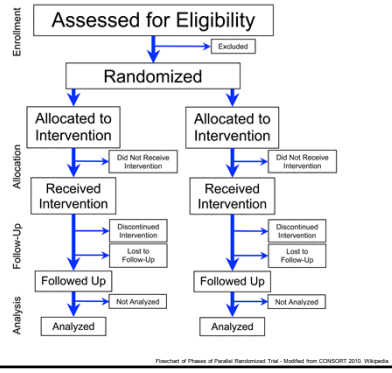
Superiority Trials – Treatment A is better than B

Non-Inferiority Trials – Treatment A is no worse than B

Equivalence Trials – Treatment A is the same as B

A Basic Trial Timeline

1. Protocol development and codification
   - Inclusion/Exclusion Criteria
   - Detailed procedures
   - Design, Power, Stopping Rules
2. Baseline measures
3. Randomization
4. Intervention
5. Follow up
6. Analysis



Flowchart of Phases of Parallel Randomized Trial - Modified from CONSORT 2010, Wikipedia

Control Condition?

- Placebo (sugar pill, sham treatment, etc)
- Active control
  - ☐ Most widely accepted treatment
  - ☐ Most accepted prevention intervention
  - ☐ Other website, etc

Randomization

The purpose of randomization is to achieve balance with respect to known and unknown risk factors in the allocation of participants to treatment arms in a study

Randomization

By probability, a simple randomization scheme may allocate a different number of participants to each study group. This may reduce the power of a statistical procedure to reject the null hypothesis as statistical power is maximized for equal sample sizes. Additionally, an imbalance of treatment groups within confounding factors may occur. This is especially true for small sample sizes.

Efird 2011 Int J Environ Res Public Health

Block Randomization

- Ensure that # of patients assigned to each treatment is not far out of balance
- Random block size reducing "good" guessing

Stratified randomization

- A priori certain factors likely important (e.g. Age, Gender)
- Randomize so different levels of the factor are BALANCED between treatment groups
- Cannot evaluate the stratification variable

Stratified randomization

- Stratify, then do block randomization

Male; 25-44 yrs	**ABBA**	**BBAA**	**BABA**	**ABAB**	**BAAB**
Female; 45-60 yrs	**AABB**	**ABBA**	**BBAA**	**BABA**	**ABAB**

Blinding

Apparently, "blinding" terminology emerged when Benjamin Franklin et al. actually blindfolded participants when evaluating Mesmerism

Lasagna used the term "double blindfold" in 1955

Slide based on Schultz "RCT" PPT

Inflating the Importance of Blinding

- Some investigators, readers, and editors overstate the importance of blinding
- Indeed, some consider RCTs as high quality if "double-blind," i.e. *sine qua non* of an RCT
- A randomized trial can be methodologically sound and not be double-blind or, conversely, double-blind and not methodologically sound

Slide based on Schultz "RCT" PPT

Loss to Follow-up

Induce selection bias?

Totally lost?

Why?

Any data collection?

Careful with imputation

MULTICENTER TRIALS

Often required in order to recruit necessary number of subjects within time frame

May enhance external validity

Enables investigators with similar interest and skills to work together on a common problem

Not a cluster trial ((but be careful!)

125

ITT vs ATT/TOT

ITT – As randomized

ATT – As treated

Answer DIFFERENT questions

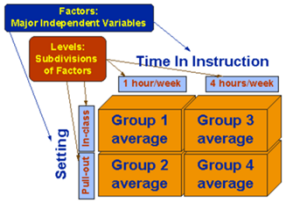
Sequential trials

- Not for a fixed sample size/ period
- Terminates when
  - ☐One treatment shows a clear superiority or
  - ☐It is highly unlikely any important difference will be seen
  - ☐Special statistical design methods

Factorial Experiments

- Imagine intervention to assess efforts to improve instruction for math. Manipulate classroom and time.
- Classroom: In or Out
- Time: 1 hour or 4 hours
- 2x2 = 2^2 = 4 cell experiment

Factorial Experiments



Cross-over Trial

- Two treatments, two period cross-overs
- Use each patient as her own control
- Must eliminate carryover effects (ie, need sufficient washout) and this is not always possible

Adaptive designs

- Smaller overall sample size (potential)
- Run-in; then analyze data continuously or fixed intervals
- Act like group sequential design
- Close an arm early
- Re-estimate sample size based on variance

Instrumental Variables for Epidemiologists

J. Michael Oakes, PhD
UMN Epidemiology

Outline

1. Review of Core Ideas
2. The Problem
3. IV Model Intuition
4. Examples
5. LATE
6. Worked Example
7. Issues & Assumptions

Analysis of Experimental Data

$$Y = \alpha + \beta_1 T + \varepsilon$$

$\hat{\beta}_1 \Rightarrow \bar{\Delta} =$ average causal effect

T is (0,1) treatment indicator which, for large samples, is independent of background characteristics by study design (ie, randomization)

Absent Randomization

$$Y = \alpha + \beta_1 T + \mathbf{\beta Z} + \varepsilon$$

Covariates, *Z*, serve to
adjust groups for confounding…

Absent Randomization

Unless specification of the model, including Z, is perfect, bias results

$$\hat{\beta}_1 = \bar{\Delta} + \text{BIAS}$$

If RCT

$$\text{Mortality} = \alpha + \beta_1 T + \varepsilon$$

If No RCT

$$\text{Mortality} = \alpha + \beta_1 T + \mathbf{\beta Z} + \varepsilon$$

Again, what goes in Z?

Patient's age, medical history, time to ED…
Cardiologist's skill, tools available…
Weather, politics, ????

$$\text{Mortality} = \alpha + \beta_1 T + \mathbf{\beta Z} + \varepsilon$$

Got Z?

If you can confidently say you measured (accurately!) all the relevant confounding variables you could claim conditional independence and conduct a propensity score analysis or employ a multiple regression model.

Really got Z?

What of unmeasured or unmeasurable variables?

Doctor's fatigue, Docs ability, Unexpected delays, Hospital reimbursement policy, Hospital equipment, ED culture, Patient's choice…????

$$\text{Mortality} = \alpha + \beta_1 T + \mathbf{\beta Z} + \varepsilon$$

If we don't (or can't!) get everything in Z then there will be a correlation between T and the error term. When this happens the estimated treatment effect is biased.

If no CIA, then

$$E(\varepsilon \mid T) \neq 0$$

$$\hat{\beta}_1 = \overline{\Delta} + \text{BIAS}$$

More generally…

$$\text{Y} = \alpha + \beta_1 T + \mathbf{\beta Z} + \varepsilon$$

If, even after adjustment for Z, there is a correlation between the treatment indicator, T, and the error term, we say that the treatment is ENDOGENOUS.

In the jargon of econometrics, T would be considered an endogenous regressor.

This is bad since to estimate valid treatment effects regression models require EXOGENOUS regressors.

IV models?

The instrumental variables technique has been appreciated for 50+ years. It's a staple of every econometrics class and textbook.

The technique seems to be *de rigueur* in economics.

Not well understood or adopted in epidemiology, where controlled experiments are also difficult. I say this is so because few epidemiologists believe in human choice.

Simply…

The IV technique fixes the problem of endogenous regressors (ie, omitted variables!)

It uses other exogenous regressors (ie, instrumental variables) to remove the endogenous part of the offending variable, which is then used in a "regular" analysis.

It's helpful to conceive of the approach as a two-step process: (1) fix problem regressor, (2) use fixed regressor in desired model.

The Good Ol' Two Step

If **Y = a + BT + GZ + e** and we miss some variable in Z that makes T correlated with e (bad news).

IV solution

1. Find an IV (predict T but not Y, unless thru T)
2. **T = a + B(IV) + e** (Regression #1, fix T)
3. Get predicted value of T, called T*
4. Fit **Y = a + BT* + GZ + e** (Regression #2, get better B)

IV Models

Can overcome problems of

- Omitted variables (often results in endogenous regressor)
- Measurement error in regressors
- Reverse causality

Can not only assess discrete Tx effects, but can also model the less biased effects of other predictors on outcome. Experiments and propensity score models cannot do this.

Imagine a XS study to estimate the effect of public health spending on incidence of some disease. We might consider equation

Disease Rates = a + B(Spending) + e

We hope that B is less than zero: as spending goes up, disease goes down. But it may be spending goes up because disease has gone up. Disease rates cause spending, which is reverse causation in this context. Regular regression estimates will always be biased no matter how many controls are added.

We need to model the causes of spending in order to estimate effect of spending on disease rates. An IV regression could be used.

Derived from Prof. Kit Baum

What is an IV?

Intuitively, instruments are variables that move around the probability of treatment but do not affect outcomes except through their effect on treatments.

Put more statistically, instruments are variables that are correlated with the endogenous variable – in this context the treatment indicator – but not correlated with the unobservable in the outcome equation.

What is an IV?

One can think of the instrumental variable as a device that achieves a pseudorandomization.

Indeed, the actual randomization in an RCT is a special case of IV. Imagine, for example, that one tosses an unbiased coin to assign people to treatment or control groups at random. The outcome of the coin toss, heads or tails, is the IV -- a variable that induces variation in the treatment variable

Parts is Parts

- For valid estimate of B in Y= a + BT + e, where T is correlated with e.
- The variation in T can be divided into two parts:
  - Part correlated with e (bad)
  - Part uncorrelated with e (good, just like RCT)
- Want to use the second part, which is what IV does.

What makes a good IV?

Imagine $Y = a + BT + e$ with potential instrument "Z"

Z is a good instrument for endogenous T if Z is

(1) Exogenous: uncorrelated with the errors
(2) Correlated with the endogenous T (predicts T)
(3) Correlated with the Y only through T

Some say "A good instrument should not be correlated with the dependent variable". This is incorrect. Z has to be correlated with Y, otherwise it is useless as an instrument. But it must be correlated with Y only through T.

Natural Experiments are when mother nature randomizes persons to one condition or another are a useful form of exogenous variation and thus can serve as excellent IVs.

Moreover, the effects of natural experiments should probably be estimated through IV methods.

IVs are not just another covariate

Assume that Z is both a valid IV and a valid covariate of a True regression model.

If Z is a covariate, the true regression model would be given by $y = X + Z + e$. The estimated model, however, is given by
$y = X+e$; Z is excluded from the regression equation (as an IV).

So the estimated error is $e = Z + e$. This error term, e, is obviously correlated with the IV, Z. Therefore, the assumption that a valid IV is also a valid covariate the model leads to a contradiction. Thus, an IV is not a valid covariate in the model.

I. Catheterization (CATH) after AMI on mortality?

- t-test shows CATH reduced four-year mortality by 37%.
- But no RCT: no exchangeability or balance
- Multiple regression (age, sex, race…) shows 28% mortality reduction due to CATH.
- Use distance to nearest hospital as IV
  - Correlated with CATH but not otherwise to mortality
- A simple IV estimator showed just 7% decrease in mortality.
- IV with more controls (high-volume hospital, rural area, etc) reduce effect of CATH on mortality to 5%.
- CATH matters but by far less than naïve methods suggested
- 37% is far different from 5%

From McClellan & Newhouse 1998

II. Compulsory school attendance effect income?

$Income = a + b_1(educ\_attain) + b_2(covariates) + e$

- But what of ability and/or motivation? How to measure it?
- It's important because ability affects educational attainment and thus leaving it out of our model creates a correlation between *educ_attain* and *e* and this biases the effect of interest, b1.
- An omitted variable problem.

From class notes of Prof Haley Fisher citing Angrist & Krueger (1991)

What instrument? Month born!

Child can enter kindergarten if 5 years old by September 1. Thus, children born on August 30 can enter kindergarten but their friend born on Sept 2 cannot. The first child will be young for her grade; the second child will be old for his grade.

Further, most states prevent youth from dropping out of school until they are 16 years old. This creates a natural experiment where children with arbitrary birthdays can drop out when they are in different grades.

Thus, month of birth is correlated to educational attainment (or drop out grade) at 16 but there is no reason to think it is related to income.

$Income = a + b_1(educ\_attain) + b_2(covariates) + e$

- Education is endogenous (correlated with error due to missing measures of "ability" and stuff) but we will use birth month as instrument to break the correlation and get a better estimate of the effect of education on income.
- Regular regression show educ_attain increases income by 7%
- IV methods
  - Step 1: educ_attain = birth_month + sex + race + hhinc + e
  - Step 2: income = educ_attain* + e
- The IV estimate shows a statistically insignificant effect of education on income!

Recall, ideally, we'd like the treatment effect for each individual in our study. If we could observe every person and their counterfactual we could just take the average across all persons as an estimate of delta.

$$\tau_i = Y_i(1) - Y_i(0)$$

$$\overline{\tau} \Rightarrow \Delta$$

Recall ATE:

The average treatment effect (ATE) is the difference in the average of the outcome variable in the treatment group minus the average of the outcome variable in the control group. ATE is the same as the average causal effect (ACE).

$$\text{ATE} = \text{ACE} = E[Y(1) - Y(0)]$$

Recall ATT:

The average treatment effect on the treated (ATT) is the mean difference between those actually treated or exposed and their counterfactuals. ATT is the same as the treatment effect on the treated (TOT).

$$\text{ATT} = \text{TOT} = E[Y(1) - Y(0) \mid \text{T=1}]$$

LATE

Local Average Treatment Effect

The average treatment effect for individuals "who can be induced to change their treatment by a change in the instrument". For example, those who comply with treatment assignment in an RCT.

LATE is the average causal effect of X on Y for "compliers," as opposed to "always takers" or "never takers". LATE does not incorporate those who refuse to be in the "RCT" or will take the "pill" regardless of assignment.

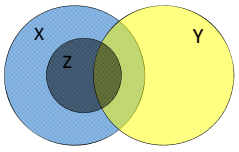
Who are the "marginal" patients (or "compliers"), whose treatment is effected by the instrument?

IV estimates treatment effect among these "marginal" patients, which is why the term "local" is used.

LATE rarely a good estimate of the treatment effect in the general population.

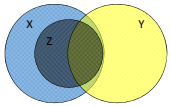
IV methods estimate average treatment effects, with the average depending on the instruments.

In IV, not all of the available variation in X is used.
Only that portion of X which is "explained" by Z is used to explain Y

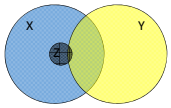


X = Endogenous variable
Y = Response variable
Z = Instrumental variable

Based on Prof. Robert Apel's PPT



Best-case scenario: A lot of X is explained by Z, and most of the overlap between X and Y is accounted for



Realistic scenario: Very little of X is explained by Z, or what is explained does not overlap much with Y

The Good & Bad

Good: IV methods can account for unmeasured factors correlated with the outcome

Good: IV methods also permit analysis of other factors that explain outcome besides treatment itself

Bad: IVs are hard to find and defend

Bad: IV standard errors tend to be large, especially when corr(x,z) is very small, which can lead to type II errors

Must have valid IV

Consistent instrumental variables (IV) estimation requires instruments which are valid, which is to say, uncorrelated with the error term in the regression equation. In practice, however, this condition is unlikely to be satisfied.

Moreover, this assumption is virtually impossible to test since the relevant error term is not directly observable.

Consequently, the validity of IV-based parameter inference largely rests on a statistical assumption which is both suspect and often untestable.

Cannot have weak IV

IV estimates using instruments which are weak (i.e., only weakly correlated with the endogenous variables) are known to yield unreliable parameter inference even when these instruments are valid, in the sense of being asymptotically uncorrelated with the model error term.

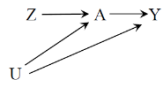
IV > RCT?

In principle, RCTs can be designed to answer any health or policy question. In practice, however, random assignment experiments have important limitations.

- Not everything of interest can or should be randomized
- Participation varies and attrition is a big problem
- RCTs black-box mechanisms

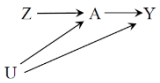
- IV techniques do not require randomization
- Varying participation rates can be studied and explained
- Mechanisms of effects can be investigated

- Z is considered an *instrumental variable* or *instrument* if:
  - Z has a nonzero causal effect on A
  - Z affects Y only through its effect on A
  - Z and Y do not share causes



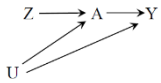
Examples

Effect of catheterization on mortality in patients suffering heart attacks
- Z: distance to nearest hospital with cath lab
- A: catheterization
- Y: mortality



Effect of blood cholesterol on heart attack risk
- Z: genes affecting blood cholesterol levels
- A: blood cholesterol levels
- Y: heart attack

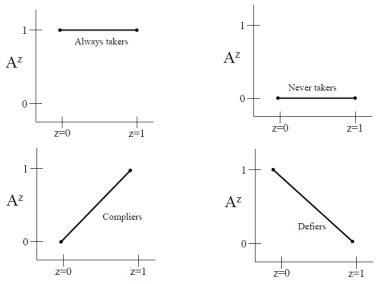
"Mendelian randomization"



Issues

- Untestable assumptions:
  - No direct effect of Z on Y
  - No common causes of Z and Y
- Additional assumptions required:
  - Homogeneity → effect of A on Y same for everyone
  - Monotonicity → no 'defiers', effect in the compliers

Monotonicity assumption



In summary

- IV methods are useful but not a panacea
- Require strong assumptions
- Good IVs are hard to find
- Estimated effect may not be the effect of interest
- Weak instruments can lead to large biases