Causal Inference & Causal Learning

Effect modification & interaction

Xiaolei Lin

School of Data Science
Fudan University

Mar 29, 2024

Recap

- Observational study vs conditional randomized experiment
- g-formula / standardization
- Inverse probability weighting

Today's plan

- Effect modification
- Interaction
- Sufficient cause

Heterogeneity of treatment effects

	V	Y^0	Y^1
Rheia	1	0	1
Demeter	1	0	0
Hestia	1	0	0
Hera	1	0	0
Artemis	1	1	1
Leto	1	0	1
Athena	1	1	1
Aphrodite	1	0	1
Persephone	1	1	1
Hebe	1	1	0
Kronos	0	1	0
Hades	0	0	0
Poseidon	0	1	0
Zeus	0	0	1
Apollo	0	1	0
Ares	0	1	1
Hephaestus	0	0	1
Polyphemus	0	0	1
Hermes	0	1	0
Dionysus	0	1	0

Heterogeneity of treatment effects

- Average causal effect of heart transplantation on mortality is null, since $Pr(Y^{a=1} = 1) = \frac{10}{20} = 0.5$ and $Pr(Y^{a=0} = 1) = \frac{10}{20} = 0.5$.
- Now, consider the causal question: what is the average causal effects of heart transplantation on mortality in men and women?
- Indicator $V = 1$ for females and $V = 0$ for males.

Heterogeneity of treatment effects

- For females $V = 1$
  - $Pr(Y^{a=1} = 1 \mid V = 1) = \frac{6}{10} = 0.6$
  - $Pr(Y^{a=0} = 1 \mid V = 1) = \frac{4}{10} = 0.4$
  - Causal risk ratio=1.5, causal risk difference = 0.2
  - on average, heart transplant increases the risk of death in women
- For males $V = 0$
  - $Pr(Y^{a=1} = 1 \mid V = 0) = \frac{4}{10} = 0.4$
  - $Pr(Y^{a=0} = 1 \mid V = 0) = \frac{6}{10} = 0.6$
  - Causal risk ratio=0.67, causal risk difference = -0.2
  - on average, heart transplant decreases the risk of death in men

Heterogeneity of treatment effects

- A null average causal effect in the population does not imply a null average causal effect in a particular subset of the population!
- The average causal effects in men and in women are of equal magnitude but in opposite direction.
- Because the proportion of each sex is 50%, both effects cancel out exactly when considering the entire population
- Heterogeneity of the individual causal effects of treatment is often expected because of variations in individual susceptibilities to treatment
- Exact cancellation of effects is probably rare

Effect modifier

V is a modifier of the effect of A on Y when the average causal effect of A on Y varies across levels of V.

- Since the average causal effect can be measured using different effect measures (e.g., risk difference, risk ratio), the presence of effect modification depends on the effect measure being used.
- Sex V is an effect modifier of the effect of heart transplant A on mortality Y on both the additive and multiplicative scales, since both causal risk ratio and risk difference vary across levels of V
- Qualitative effect modification: the average causal effects in the subsets $V = 1$ and $V = 0$ are in the opposite direction.

Effect modifier

- In the presence of qualitative effect modification, additive effect modification implies multiplicative effect modification, and vice versa.

- In the absence of qualitative effect modification, however, one can find effect modification on one scale (e.g., multiplicative) but not on the other (e.g., additive)
  - $Pr(Y^{a=0} = 1 \mid V = 1) = 0.8$, $Pr(Y^{a=1} = 1 \mid V = 1) = 0.9$
  - $Pr(Y^{a=0} = 1 \mid V = 0) = 0.1$, $Pr(Y^{a=1} = 1 \mid V = 0) = 0.2$
  - multiplicative, but no additive, effect modification by V
  - one cannot state that there is, or there is not, effect modification without referring to the effect measure being used

Identify Effect modification

- A stratified analysis is the natural way to identify effect modification
  - compute the causal effect of A on Y in each level (stratum) of the variable V
  - if causal effects differed between strata (on both the additive and the multiplicative scale), conclude there is additive or multiplicative effect modification
- For dichotomous V , the stratified causal risk differences are:
  - $Pr(Y^{a=1} = 1 \mid V = 1)$ - $Pr(Y^{a=0} = 1 \mid V = 1)$, and
  - $Pr(Y^{a=1} = 1 \mid V = 0)$ - $Pr(Y^{a=0} = 1 \mid V = 0)$

Identify Effect modification

For marginally randomized experiment

- if treatment assignment was random and unconditional, exchangeability is expected in every subset of the population
  - The causal risk difference in women, $Pr(Y^{a=1} = 1 \mid V = 1) - Pr(Y^{a=0} = 1 \mid V = 1) = Pr(Y = 1 \mid A = 1, V = 1) - Pr(Y = 1 \mid A = 0, V = 1)$
  - similarly for men

For conditionally randomized experiment

- conditional exchangeability is expected in each level of the observed covariates C

Identify Effect modification

- In a population of 40 people, transplant A has been randomly assigned with probability 0.75 to those in severe condition ($C = 1$), and with probability 0.50 to the others ($C = 0$)
- The 40 individuals can be classified into two nationalities according to their passports: 20 are Greek ($V = 1$) and 20 are Roman ($V = 0$)
- Does nationality modify the causal effect of transplant on mortality?

Greeks $V = 1$

	L	A	Y
Rheia	0	0	0
Kronos	0	0	1
Demeter	0	0	0
Hades	0	0	0
Hestia	0	1	0
Poseidon	0	1	0
Hera	0	1	0
Zeus	0	1	1
Artemis	1	0	1
Apollo	1	0	1
Leto	1	0	0
Ares	1	1	1
Athena	1	1	1
Hephaestus	1	1	1
Aphrodite	1	1	1
Polyphemus	1	1	1
Persephone	1	1	1
Hermes	1	1	0
Hebe	1	1	0
Dionysus	1	1	0

Greeks $V = 1$

- $Pr(Y^{a=1} = 1 \mid V = 1) = Pr(Y = 1 \mid A = 1, V = 1, C = 1) \times Pr(C = 1) + Pr(Y = 1 \mid A = 1, V = 1, C = 0) \times Pr(C = 0) = 0.5$
- $Pr(Y^{a=0} = 1 \mid V = 1) = Pr(Y = 1 \mid A = 0, V = 1, C = 1) \times Pr(C = 1) + Pr(Y = 1 \mid A = 0, V = 1, C = 0) \times Pr(C = 0) = 0.5$
- Causal risk ratio=1
- causal risk difference =0

Romans $V = 0$

	L	A	Y
Cybele	0	0	0
Saturn	0	0	1
Ceres	0	0	0
Pluto	0	0	0
Vesta	0	1	0
Neptune	0	1	0
Juno	0	1	1
Jupiter	0	1	1
Diana	1	0	0
Phoebus	1	0	1
Latona	1	0	0
Mars	1	1	1
Minerva	1	1	1
Vulcan	1	1	1
Venus	1	1	1
Seneca	1	1	1
Proserpina	1	1	1
Mercury	1	1	0
Juventas	1	1	0
Bacchus	1	1	0

Romans $V = 0$

- $Pr(Y^{a=1} = 1 \mid V = 0) = Pr(Y = 1 \mid A = 1, V = 0, C = 1) \times Pr(C = 1) + Pr(Y = 1 \mid A = 1, V = 0, C = 0) \times Pr(C = 0) = 0.6$
- $Pr(Y^{a=0} = 1 \mid V = 0) = Pr(Y = 1 \mid A = 0, V = 0, C = 1) \times Pr(C = 1) + Pr(Y = 1 \mid A = 0, V = 0, C = 0) \times Pr(C = 0) = 0.3$
- Causal risk ratio =2
- causal risk difference =0.3

Identify Effect modification

Compute the conditional risks $Pr[Y^{a=1} = 1 \mid V = v]$ and $Pr[Y^{a=0} = 1 \mid V = v]$ in each stratum v, involves

- stratification by V
- standardization (g-formula) or inverse probability weighting by C

There is both additive and multiplicative effect modification by nationality of the effect of heart transplant on mortality.

Possible causal mechanisms involved in such effect modification?

Identify Effect modification

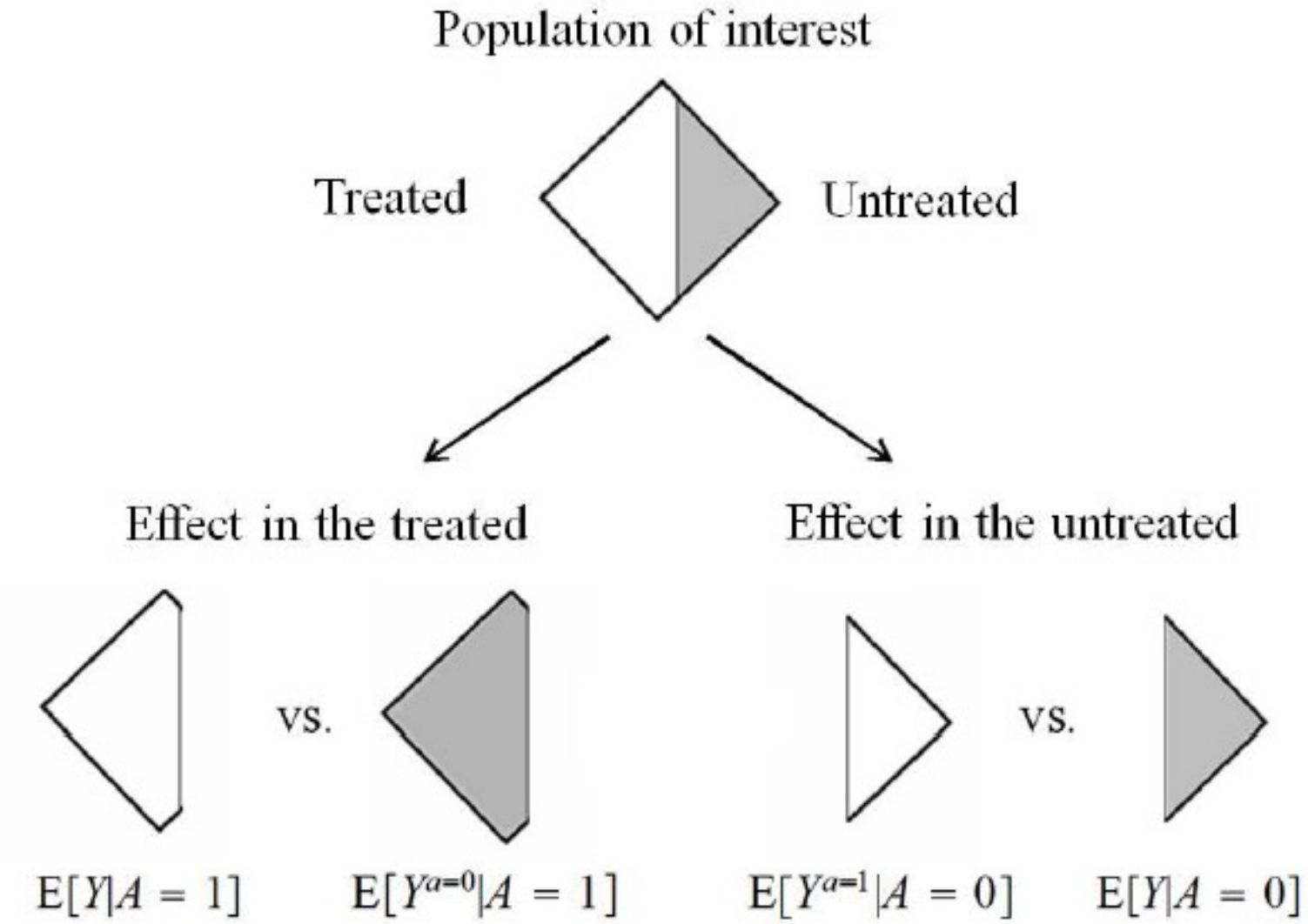
- The qualiy of heart surgery is better in Greece than in Rome?
- An intervention to improve the quality of heart surgery in Rome could eliminate the modification of the causal effects by nationality?
  - passport-defined nationality: surrogate effect modifier
  - quality of care: causal effect modifier
- Effect modifier V does not necessarily imply that V plays a causal role in modifying the effect

Causal effect in the treated

Sometimes, one would be particularly interested in estimating the causal effects in the treated

- Average causal difference in the treated
  $Pr(Y^{a=1} \mid A = 1) - Pr(Y^{a=0} \mid A = 1)$, or by consistency,
  $Pr(Y = 1 \mid A = 1) - Pr(Y^{a=0} = 1 \mid A = 1)$
- Average causal risk ratio in the treated
  $Pr(Y^{a=1} \mid A = 1)/Pr(Y^{a=0} \mid A = 1)$, or by consistency,
  $Pr(Y = 1 \mid A = 1)/Pr(Y^{a=0} = 1 \mid A = 1)$
- Risk in the treated compared to the counterfactual risk had the treated individuals been untreated
- The average effect in the treated will differ from the average effect in the population if the distribution of individual causal effects varies between the treated and the untreated

Causal effect in the treated & untreated



Causal effect in the treated

Computing the average causal effect in the treated only requires partial exchangeability $Y^{a=0} \perp A \mid C$.

It is irrelevant whether the risk in the untreated, had they been treated, equals the risk in those who were actually treated

- via standardization:
  $E(Y^a \mid A = a') = \sum_l E(Y \mid A = a, C = c) Pr(C = c \mid A = a')$
- via IP weighting: $E(Y^a \mid A = a') = E(\frac{I(A=a)Y}{f(A|C)} Pr(A = a' \mid C))/E(\frac{I(A=a)}{f(A|C)} Pr(A = a' \mid C))$, which is the IP weighted mean with weight $\frac{Pr(A=a'|C)}{f(A|C)}$

Why care about effect modification

- Average causal effect in a population depends on the distribution of individual causal effects in the population
- Average causal effect of treatment A on outcome Y cannot fit in the presence of heterogeneous treatment effect
- Identify sub-populations where treatment is the most beneficial, as well as sub-populations where treatment can be harmful
- Additive (not multiplicative) effect modification is the appropriate scale to identify groups that will benefit most from interventions

Why care about effect modification

- Transportability: extrapolation of causal effects computed in one population to a second population
- Conditional causal effects in the strata defined by the effect modifiers may be more transportable than the causal effect in the entire population
- However, there could be other unmeasured, or unknown, causal effect modifiers whose conditional distributions vary between the two populations
- Therefore, transportability of causal effects across populations is a more difficult problem than the identification of causal effects in a single population
  - need to stratify all those things required to achieve exchangeability

Transportability

Transportability of causal effects across populations may be justified if the following characteristics are similar between the two populations

- Effect modification
- Versions of treatment
- Interference.
  - Interference exists when treating one individual affect the outcome of others in the population.
  - e.g., A socially active individual may convince his friends to join him while exercising, and thus an intervention on that individual's physical activity may be more effective than an intervention on a socially isolated individual

Adjustment via stratification

In the previous heart transplant example

- standardization (or IP weighting) is used to adjust for C
- stratification is used to identify effect modification by V
- However, stratification can also be used to adjust for C
- Stratification results in multiple stratum-specific effect measures, each of them quantifies the average causal effect in a nonoverlapping subset of the population

Stratification challenges

- stratification forces one to evaluate effect modification by all variables required to achieve conditional exchangeability: Romans with critical condition, Romans with non-critical condition, Greeks with critical condition, Greeks with non-critical.
- stratification requires posivity; causal effects cannot be estimated in stratum with only treated, or untreated
- noncollapsibility of odds ratio
- adjustment for time-varying covariates

Collapsibility

- In the absence of multiplicative effect modification by V, the causal risk ratio in the entire population $Pr(Y^{a=1} = 1)/Pr(Y^{a=0} = 1)$ would be equal to the conditional risk ratio in each stratum $Pr(Y^{a=1} = 1 \mid V = v)/Pr(Y^{a=0} = 1 \mid V = v)$
- In the presence of effect modification by V, the causal risk ratio would be a weighted average of the stratum-specific risk ratios
- e.g., if the causal risk ratios in the strata $V = 1$ and $V = 0$ were equal to 2 and 3, then the causal risk ratio in the entire population would be between 2 and 3
- collapsibility is a desirable characteristics of an effect measure

Non-collapsibility of OR

	V	A	Y
Rheia	1	0	0
Demeter	1	0	0
Hestia	1	0	0
Hera	1	0	0
Artemis	1	0	1
Leto	1	1	0
Athena	1	1	1
Aphrodite	1	1	1
Persephone	1	1	0
Hebe	1	1	1
Kronos	0	0	0
Hades	0	0	0
Poseidon	0	0	1
Zeus	0	0	1
Apollo	0	0	0
Ares	0	1	1
Hephaestus	0	1	1
Polyphemus	0	1	1
Hermes	0	1	0
Dionysus	0	1	1

Non-collapsibility of OR

- Assume marginal exchangeability, in the entire population
  - causal risk ratio $Pr(Y = 1 \mid A = 1)/Pr(Y = 1 \mid A = 0) = 2.3$
  - causal odds ratio $\frac{Pr(Y=1|A=1)/Pr(Y=0|A=1)}{Pr(Y=1|A=0)/Pr(Y=0|A=0)} = 5.4$
- For men
  - causal risk ratio
    $Pr(Y = 1 \mid A = 1, V = 0)/Pr(Y = 1 \mid A = 0, V = 0) = 2$
  - causal odds ratio $\frac{Pr(Y=1|A=1,V=0)/Pr(Y=0|A=1,V=0)}{Pr(Y=1|A=0,V=0)/Pr(Y=0|A=0,V=0)} = 6$
- For women
  - causal risk ratio
    $Pr(Y = 1 \mid A = 1, V = 1)/Pr(Y = 1 \mid A = 0, V = 1) = 3$
  - causal odds ratio $\frac{Pr(Y=1|A=1,V=1)/Pr(Y=0|A=1,V=1)}{Pr(Y=1|A=0,V=1)/Pr(Y=0|A=0,V=1)} = 6$

Non-collapsibility of OR

- Risk ratio and risk difference are collapsible effect measures
- Odds ratio and odds difference (rarely used) are non-collapsible effect measures
  - odds ratio is collapsible under the sharp null hypothesis
  - approximately collapsible—and approximately equal to the risk ratio when the outcome is rare (say, $< 10\%$) in every stratum
  - The population causal odds ratio can be closer to the null value than the non-null stratum-specific causal odds ratio when V is an independent risk factor for Y and, as in our randomized experiment, A is independent of V

Pooling of stratum-specific effect measures

When dealing with stratum-specific effect measures estimated from samples, a common approach to reduce the variability of the estimates is to combine all stratum-specific effect measures into one pooled stratum-specific effect measure

- if there is no effect-measure modification, the pooled effect measure will be a more precise estimate of the common effect measure than each of the stratum-specific effect measures
- Pooling methods (e.g., Woolf, Mantel-Haenszel, maximum likelihood) sometimes compute a weighted average of the stratum-specific effect measures with weights chosen to reduce the variability of the pooled estimate
- if there are significant effect modifier, stratum specific effect measures should be provided rather than pooled effect measures

Adjustment via matching

- Goal: construct a subset of the population in which the variables C have the same distribution in both the treated and the untreated
- For each untreated individual in non critical condition $(A = 0, C = 0)$, randomly select a treated individual in non critical condition $(A = 1, C = 0)$
- For each untreated individual in critical condition $(A = 0, C = 1)$, randomly select a treated individual in critical condition $(A = 1, C = 1)$.
- matched pair: each untreated individual and corresponding treated individual
- matching factor: whether or not in critical condition C

Adjustment via matching

- All the untreated, but only a sample of the treated, in the population were selected
- In this subset of the population comprised of matched pairs, the proportion of individuals in critical condition ($C = 1$) is the same, by design, in the treated and in the untreated
- Under the assumption of conditional exchangeability given C, the result of this procedure is marginal exchangeability of the treated and the untreated in the matched population
- Due to the marginal exchangeability, average outcomes can be directly compared in the matched population across the treated and untreated
- matching ensures positivity in the matched population

Interaction

Joint intervention

- receiving either a multivitamin complex ($E = 1$) or no vitamins ($E = 0$)
- either heart transplant ($A = 1$) or no heart transplant ($A = 0$)

Four counterfactual outcomes: $Y^{a=0,e=0}$, $Y^{a=1,e=0}$, $Y^{a=0,e=1}$, $Y^{a=1,e=1}$

Interaction: there is interaction between two treatments A and E if the causal effect of A on Y after a joint intervention that set E to 1 differs from the causal effect of A on Y after a joint intervention that set E to 0

Interaction

When the causal effect is measured on the risk difference scale, there is interaction on the additive scale if

- $Pr(Y^{a=1,e=1} = 1) - Pr(Y^{a=0,e=1} = 1) \neq Pr(Y^{a=1,e=0} = 1) - Pr(Y^{a=0,e=0} = 1)$
- e.g., causal risk difference for transplant A when everybody receives vitamins$Pr(Y^{a=1,e=1} = 1) - Pr(Y^{a=0,e=1} = 1) = 0.1$
- causal risk difference for transplant A when nobody receives vitamins$Pr(Y^{a=1,e=0} = 1) - Pr(Y^{a=0,e=0} = 1) = 0.2$
- there is additive interaction between A and E on Y

Interaction

Using similar algebra, it can be easily calculated that

- the causal risk difference for vitamins E when everybody receives a transplant, $Pr(Y^{a=1,e=1} = 1) - Pr(Y^{a=1,e=0} = 1)$
- is smaller than
- the causal risk difference for vitamins E when nobody receives a transplant, $Pr(Y^{a=0,e=1} = 1) - Pr(Y^{a=0,e=0} = 1)$

Therefore, interaction can be equivalently defined by

$$Pr(Y^{a=1,e=1} = 1) - Pr(Y^{a=1,e=0} = 1) \neq Pr(Y^{a=0,e=1} = 1) - Pr(Y^{a=0,e=0}$$

Difference between Interaction and effect modification

- effect modification: a variable V is a modifier of the effect of A on Y when the average causal effect of A on Y varies across levels of V
- The concept of effect modification refers to the causal effect of A, not to the causal effect of V
- V and A don't have the same status in effect modification
- Effect modification does not involve counterfactual outcome $Y^{a,e}$, but only Y^a
- However, A AND V have equal status in interaction

Identify interaction

- Because interaction is concerned with the joint effect of two (or more) treatments A and E, identifying interaction requires exchangeability, positivity, and consistency for both treatments.

- Suppose that vitamins E were randomly, and unconditionally, assigned by the investigators
  - the treated $E = 1$ and the untreated $E = 0$ are exchangeable
  - marginal risk $Pr(Y^{a=1,e=1} = 1)$ = conditional risk $Pr(Y^{a=1} = 1 \mid E = 1)$

- Interaction between A and E on the additive scale

$$\begin{aligned} &Pr(Y^{a=1} = 1 \mid E = 1) - Pr(Y^{a=0} = 1 \mid E = 1) \\ \neq &Pr(Y^{a=1} = 1 \mid E = 0) - Pr(Y^{a=0} = 1 \mid E = 0) \end{aligned}$$

Identify interaction

- When treatment E is randomly assigned, interaction and effect modification coincide.
  - Methods for identifying modification of the effect of A by V can now be applied to identify interaction of A and E
- When treatment E is not randomly assigned by investigators, use g-formula / standardization or IP weighting of the marginal risks $Pr[Y^{a,e} = 1]$.
  - view AE as a combined treatment with four possible levels $(11, 01, 10, 00)$

Additive and multiplicative interaction

The null hypothesis of no interaction between A and E on additive scale

$$Pr(Y^{a=1,e=1} = 1) - Pr(Y^{a=0,e=1} = 1) = Pr(Y^{a=1,e=0} = 1) - Pr(Y^{a=0,e=0}$$

can be written as

$$\begin{aligned} & Pr(Y^{a=1,e=1} = 1) - Pr(Y^{a=0,e=0} = 1) \\ =& \{Pr(Y^{a=1,e=0} = 1) - Pr(Y^{a=0,e=0} = 1)\} \\ +& \{Pr(Y^{a=0,e=1} = 1) - Pr(Y^{a=0,e=0} = 1)\} \end{aligned}$$

The causal risk difference $Pr(Y^{a=1,e=1} = 1) - Pr(Y^{a=0,e=0} = 1)$ is additive because it can be written as the sum of the causal risk difference that measure the effect of A in the absence of E and the effect of E in the absence A

Additive and multiplicative interaction

Similarly, interaction between A and E on multiplicative scale

$$\frac{Pr(Y^{a=1,e=1}=1)}{Pr(Y^{a=0,e=0}=1)} = \frac{Pr(Y^{a=1,e=0}=1)}{Pr(Y^{a=0,e=0}=1)} \times \frac{Pr(Y^{a=0,e=1}=1)}{Pr(Y^{a=0,e=0}=1)}$$

The causal risk ratio $\frac{Pr(Y^{a=1,e=1}=1)}{Pr(Y^{a=0,e=0}=1)}$ is super multiplicative if left side is greater than the right side, and is submultiplicative if the left side is smaller than the right side.

Counterfactual response types and interaction

Type	$Y^{a=0}$	$Y^{a=1}$
Doomed	1	1
Helped	1	0
Hurt	0	1
Immune	0	0

Individuals can be classified in terms of their counterfactual responses when there is only one intervention A:

- $Y^{a=0} = 1$ & $Y^{a=1} = 1$: doomed
- $Y^{a=0} = 0$ & $Y^{a=1} = 0$: immune
- $Y^{a=0} = 1$ & $Y^{a=1} = 0$: helped
- $Y^{a=0} = 0$ & $Y^{a=1} = 1$: hurt

Counterfactual response types and interaction

	$Y^{a,e}$ for each a, e value			
Type	$1,1$	$0,1$	$1,0$	$0,0$
1	1	1	1	1
2	1	1	1	0
3	1	1	0	1
4	1	1	0	0
5	1	0	1	1
6	1	0	1	0
7	1	0	0	1
8	1	0	0	0
9	0	1	1	1
10	0	1	1	0
11	0	1	0	1
12	0	1	0	0
13	0	0	1	1
14	0	0	1	0
15	0	0	0	1
16	0	0	0	0

Counterfactual response types and interaction

Individuals can be classified in terms of their counterfactual responses when there are joint intervention A & E

- 1: $Y^{a=1,e=1} = 1,\ Y^{a=0,e=1} = 1,\ Y^{a=1,e=0} = 1,\ Y^{a=0,e=0} = 1$
  - neither A nor E has causal effect on Y
- 16: $Y^{a=1,e=1} = 0,\ Y^{a=0,e=1} = 0,\ Y^{a=1,e=0} = 0,\ Y^{a=0,e=0} = 0$
  - neither A nor E has causal effect on Y

Counterfactual response types and interaction

Individuals can be classified in terms of their counterfactual responses when there are joint intervention A & E

- 4 : $Y^{a=1,e=1} = 1,\ Y^{a=0,e=1} = 1,\ Y^{a=1,e=0} = 0,\ Y^{a=0,e=0} = 0$
  - die when given vitamin, survive when no vitamin
- 6: $Y^{a=1,e=1} = 1,\ Y^{a=0,e=1} = 0,\ Y^{a=1,e=0} = 1,\ Y^{a=0,e=0} = 0$
  - die when transplant, survive when no transplant
- 11: $Y^{a=1,e=1} = 0,\ Y^{a=0,e=1} = 1,\ Y^{a=1,e=0} = 0,\ Y^{a=0,e=0} = 1$
  - survive when transplant, die when no transplant
- 13: $Y^{a=1,e=1} = 0,\ Y^{a=0,e=1} = 0,\ Y^{a=1,e=0} = 1,\ Y^{a=0,e=0} = 1$
  - survive when vitamin, die when no vitamin

Counterfactual response types and interaction

Individuals can be classified in terms of their counterfactual responses when there are joint intervention A & E

- 8 : $Y^{a=1,e=1} = 1,\ Y^{a=0,e=1} = 0,\ Y^{a=1,e=0} = 0,\ Y^{a=0,e=0} = 0$
  - would only die when given transplant and vitamin
- 12: $Y^{a=1,e=1} = 0,\ Y^{a=0,e=1} = 1,\ Y^{a=1,e=0} = 0,\ Y^{a=0,e=0} = 0$
  - would only die when no transplant and vitamin
- 14: $Y^{a=1,e=1} = 0,\ Y^{a=0,e=1} = 0,\ Y^{a=1,e=0} = 1,\ Y^{a=0,e=0} = 0$
  - would only die when transplant and no vitamin
- 15: $Y^{a=1,e=1} = 0,\ Y^{a=0,e=1} = 0,\ Y^{a=1,e=0} = 0,\ Y^{a=0,e=0} = 1$
  - would only die when no transplant and no vitamin

Counterfactual response types and interaction

Individuals can be classified in terms of their counterfactual responses when there are joint intervention A & E

- 2 : $Y^{a=1,e=1} = 1,\ Y^{a=0,e=1} = 1,\ Y^{a=1,e=0} = 1,\ Y^{a=0,e=0} = 0$
  - would only survive when given no transplant and no vitamin
- 3: $Y^{a=1,e=1} = 1,\ Y^{a=0,e=1} = 1,\ Y^{a=1,e=0} = 0,\ Y^{a=0,e=0} = 1$
  - would only survive when transplant and no vitamin
- 5: $Y^{a=1,e=1} = 1,\ Y^{a=0,e=1} = 0,\ Y^{a=1,e=0} = 1,\ Y^{a=0,e=0} = 1$
  - would only survive when no transplant and vitamin
- 9: $Y^{a=1,e=1} = 0,\ Y^{a=0,e=1} = 1,\ Y^{a=1,e=0} = 1,\ Y^{a=0,e=0} = 1$
  - would only survive when transplant and vitamin

Counterfactual response types and interaction

Individuals can be classified in terms of their counterfactual responses when there are joint intervention A & E

- 7 : $Y^{a=1,e=1} = 1,\ Y^{a=0,e=1} = 0,\ Y^{a=1,e=0} = 0,\ Y^{a=0,e=0} = 1$
  - would die when transplant and vitamin, and when no transplant and no vitamin
  - would survive when no transplant and vitamin, and when transplant and no vitamin
- 10: $Y^{a=1,e=1} = 0,\ Y^{a=0,e=1} = 1,\ Y^{a=1,e=0} = 1,\ Y^{a=0,e=0} = 0$
  - would survive when transplant and vitamin, and when no transplant and no vitamin
  - would die when no transplant and vitamin, and when transplant and no vitamin

Counterfactual response types and interaction

- If all individuals in the population have response types 1, 4, 6, 11, 13 and 16, there will be no interaction between A and E on the additive scale.

- In the presence of interaction between A and E, there must be individuals
  - who would develop the outcome under only one of the four treatment types (8, 12, 14, and 15)
  - who would develop the outcome under two treatment combinations, the effect of each treatment is exactly the opposite under each level of the other treatment (7 and 10)
  - who would develop the outcome under three of the four treatment combinations (2, 3, 5, and 9)

Counterfactual response types and interaction

Absence of additive interaction between A and E implies

- No individual in the population belong to one of the three classes
- Perfect cancellation of equal deviations from additivity of opposite sign
  - equal proportion of individuals in 7 and 10
  - equal proportion of individuals in 8 and 12

Monotonicity of causal effects

Type	$Y^{a=0}$	$Y^{a=1}$
Doomed	1	1
Helped	1	0
Hurt	0	1
Immune	0	0

For individuals with $Y^{a=1} \geq Y^{a=0}$, where counterfactual outcomes are monotonically increasing (i.e., nondecreasing) in a.

When the treatment cannot prevent any individual's outcome (i.e., there are no individuals with $Y^{a=1} = 0$ & $Y^{a=0} = 1$), all individuals' counterfactual response types are monotonically increasing in a.

We then simply say that the causal effect of A on Y is monotonic.

Monotonicity of causal effects

The concept of monotonicity can be generalized to joint intervention A and E.

- The causal effects of A and E on Y are monotonic if every individual's counterfactual outcomes $Y^{a,e}$ are monotonically increasing in both a and e.
- equivalently, if there are no individuals with $(Y^{a=1,e=1} = 0), Y^{a=0,e=1} = 1)), (Y^{a=1,e=0} = 0), Y^{a=0,e=0} = 1)),$ $(Y^{a=1,e=1} = 0), Y^{a=1,e=0} = 1)), (Y^{a=0,e=1} = 0), Y^{a=0,e=0} = 1))$

Counterfactual and interaction

If we are interested in learning whether some individuals will develop the outcome when receiving both treatments $E = 1$ and $A = 1$, but not when receiving only one of the two

- i.e., whether there are individuals with $Y^{a=1e=1} = 1$, $Y^{a=0e=1} = Y^{a=1e=0} = 0$ (type 7 and 8)

A sufficient condition for these individuals to exist:

$$Pr(Y^{a=1,e=1} = 1) - \{Pr(Y^{a=0,e=1} = 1) + Pr(Y^{a=1,e=0} = 1)\} > 0$$

- equivalently, $Pr(Y^{a=1,e=1} = 1) > Pr(Y^{a=0,e=1} = 1) + Pr(Y^{a=1,e=0} = 1)$
- proof in VanderWeele and Robins (2007a, 2008)
- called synergism between A and E

Counterfactual and interaction

If one is willing to assume that intervention A AND E are both monotonic, then a sufficient condition for synergism

$$Pr(Y^{a=1,e=1} = 1) - Pr(Y^{a=0,e=1} = 1) > Pr(Y^{a=1,e=0} = 1) - Pr(Y^{a=0,e=0}$$

- When the effects of A and E are monotonic, the presence of superadditive interaction implies the presence of individuals in type 8
- monotonicity rules out individuals with type 7
- proof in Greenland, Lash, and Rothman (2008).
- Individuals in type 8 are of particularly interests in genetic study (compositional epistasis)

Sufficient causes

- Some individuals die when they are treated, others when they are not treated, others die no matter what, and others do not die no matter what.
- Therefore, treatment A is not the only variable that determines whether or not the outcome Y occurs.
- Suppose that heart transplant $A = 1$ only results in death in individuals allergic to anesthesia
- We refer to the smallest set of background factors that, together with $A = 1$, are sufficient to inevitably produce the outcome as U_1.
- The simultaneous presence of treatment $(A = 1)$ and allergy to anesthesia $(U_1 = 1)$ is a minimal sufficient cause of the outcome Y

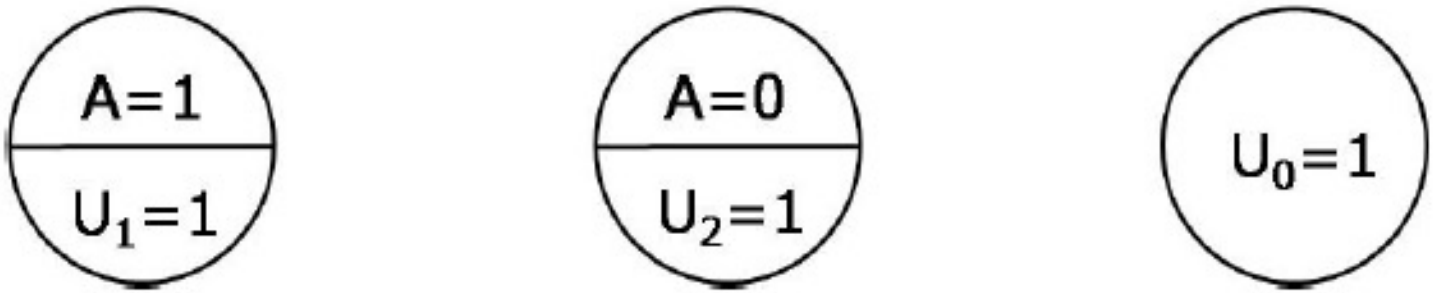
Sufficient causes

- Suppose that no heart transplant $A = 0$ only results in death if individuals have an ejection fraction less than 20%.
- We refer to the smallest set of background factors that, together with $A = 0$, are sufficient to produce the outcome as U_2.
- The simultaneous absence of treatment $(A = 0)$ and presence of low ejection fraction $(U_2 = 1)$ is another sufficient cause of the outcome Y

Sufficient causes

- Suppose there are some individuals who have neither U_1 nor U_2 and that would have developed the outcome whether they had been treated or untreated.
- The existence of these "doomed" individuals implies that there are some other background factors that are themselves sufficient to bring about the outcome.
- Suppose that all individuals with pancreatic cancer at the start of the study will die.
- We refer to the smallest set of background factors that are sufficient to produce the outcome regardless of treatment status as U_0.
- The presence of pancreatic cancer ($U_0 = 1$) is another sufficient cause of the outcome Y .

Sufficient causes



- Graphical representation of sufficient-component causes helps visualize a key consequence of effect modification
- The magnitude of the causal effect of treatment A depends on the distribution of effect modifiers
  - population with 1% vs 10% of individuals with $U_1 = 1$ (i.e., allergy to anesthesia).

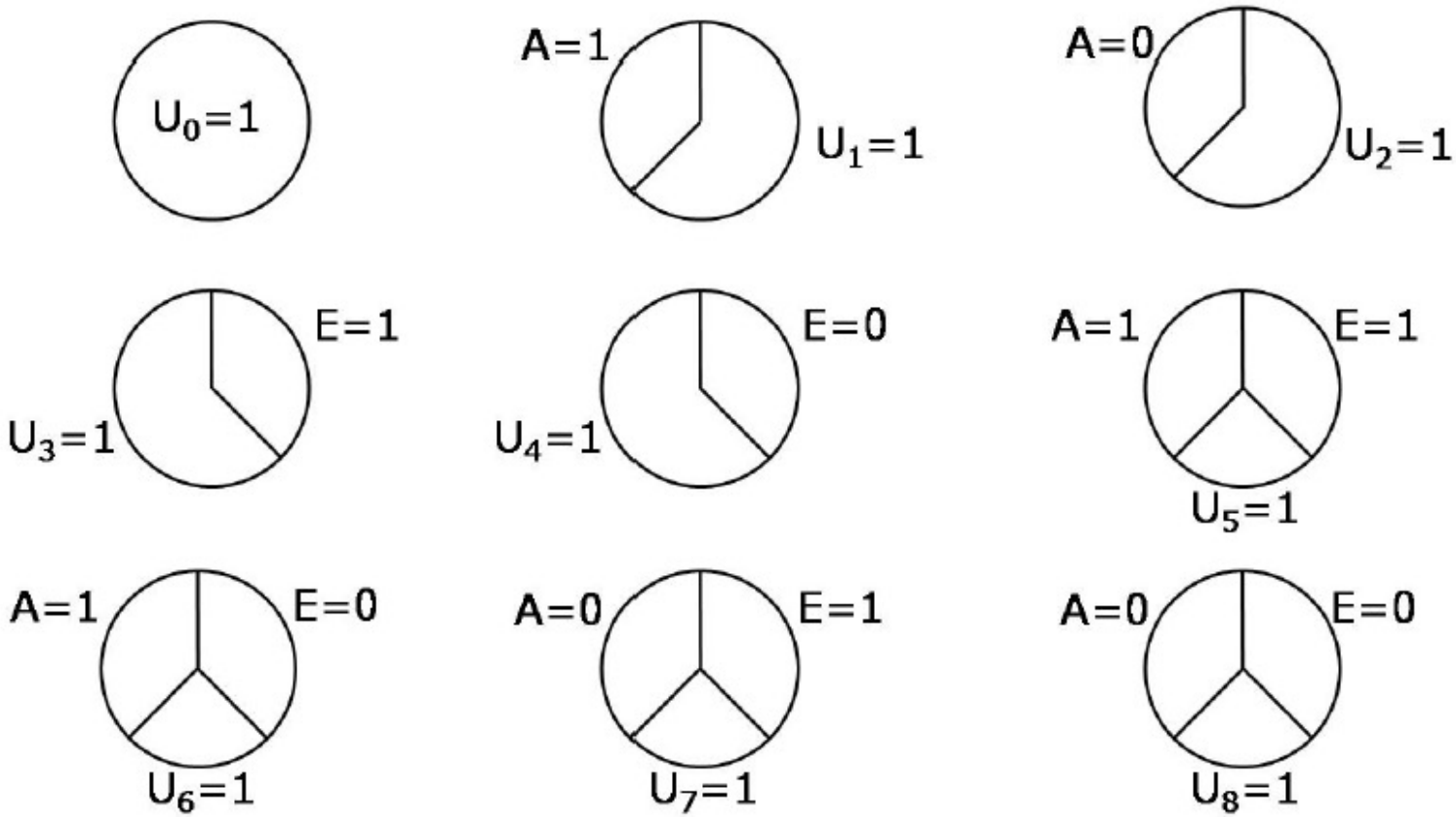
Sufficient causes

When joint intervention of A and E, ways to die

1. by treatment A (treatment E is irrelevant)
2. by the absence of treatment A (treatment E is irrelevant)
3. by treatment E (treatment A is irrelevant)
4. by the absence of treatment E (treatment A is irrelevant)
5. by both treatments A and E
6. by treatment A and the absence of E
7. by treatment E and the absence of A
8. by the absence of both A and E
9. by other mechanisms (both treatments A and E are irrelevant)

Each of these sufficient causes includes a set of background factors from U_1,..., U_8 and U_0.

Sufficient causes



Not all 9 sufficient-component causes for a dichotomous outcome and two treatments exist in all settings

Sufficient cause interaction

A sufficient cause interaction between A and E exists in the population if A and E occur together in a sufficient cause.

- Suppose individuals with background factors $U_5 = 1$ will develop the outcome when jointly receiving vitamins ($E = 1$) and heart transplant ($A = 1$), but not when receiving only one of the two treatments.
- Then a sufficient cause interaction between A and E exists if there exists an individual with $U_5 = 1$.
- Sufficient cause interactions can be synergistic or antagonistic
  - There is synergism between treatment A and treatment E when $A = 1$ and $E = 1$ are present in the same sufficient cause
  - There is antagonism between treatment A and treatment E when $A = 1$ and $E = 0$ (or $A = 0$ and $E = 1$) are present in the same sufficient cause