

反事実アウトカムと効果の修飾



北海道大学 医学統計学
横田 勲

1

今回の内容

2

- ▶ 反事実アウトカム
 - ▶ 平均因果効果
 - ▶ 交絡の定義
 - ▶ 標準化と逆確率重み付け法
- ▶ 効果の修飾

2

コントロール control

3

- ▶ 比較する相手
 - ▶ 対照
- ▶ 何かの効果調べる場合には、比較が基本

3

反事実アウトカム

4

- ▶ 事実が観察されたら、観察されないアウトカム
- ▶ 事実データ factual data
 - ▶ 治療を受けて ($a = 1$)、再発しなかった ($Y = 0$)
- ▶ 反事実データ counterfactual data
 - ▶ 治療を受けなかったら ($a = 0$)、どうなった? ($Y = ?$)

4

潜在アウトカム potential outcome

5

- ▶ $\gamma^{a=0}$
 - ▶ 曝露 $a = 0$ を受けた場合のアウトカム
- ▶ $\gamma^{a=1}$
 - ▶ 曝露 $a = 1$ を受けた場合のアウトカム
- ▶ アウトカムも2値(0,1)の場合

	$\gamma^{a=0}$	$\gamma^{a=1}$
Doomed	1	1
Helped	1	0
Hurt	0	1
Immune	0	0

5

潜在アウトカムと観察アウトカム

6

- ▶ 受けた曝露に応じて、潜在アウトカムのいずれかが観察される

	A	$\gamma^{a=0}$	$\gamma^{a=1}$	Y
Doomed	1	1	1	1
Helped	1	1	0	1
Hurt	1	0	1	0
Immune	1	0	0	0
Doomed	0	1	1	1
Helped	0	1	0	0
Hurt	0	0	1	1
Immune	0	0	0	0

6

個人での因果効果

7

	$Y^{a=0}$	$Y^{a=1}$	Causal effect $Y^{a=1} - Y^{a=0}$
Doomed	1	1	$1 - 1 = 0$
Helped	1	0	$1 - 0 = 1$
Hurt	0	1	$0 - 1 = -1$
Immune	0	0	$0 - 0 = 0$

- ▶ データとして観察はできない
 - ▶ 反事実アウトカムとの比較で定義可能
- ▶ Sharp causal null hypothesis
 - ▶ Doomed, Immuneな人しかいない

7

平均因果効果 Average Causal Effects

8

- ▶ $E[Y^{a=1}] - E[Y^{a=0}]$
 - ▶ 集団全員が曝露を受けた場合と
集団全員が曝露を受けなかった場合の差
- ▶ Null hypothesis of no average causal effect
 - ▶ $E[Y^{a=1}] = E[Y^{a=0}]$
 - ▶ Sharp causal null hypothesisに加え、
Helpedな人とHurtな人が同数いる場合も成立

8

因果効果の指標

9

- ▶ 因果リスク差
 - ▶ $\Pr[Y^{a=1} = 1] - \Pr[Y^{a=0} = 1]$
- ▶ 因果リスク比
 - ▶ $\frac{\Pr[Y^{a=1}=1]}{\Pr[Y^{a=0}=1]}$
- ▶ 因果オッズ比
 - ▶ $\frac{\Pr[Y^{a=1}=1]/\Pr[Y^{a=1}=0]}{\Pr[Y^{a=0}=1]/\Pr[Y^{a=0}=0]}$

9

練習① 以下のACEは？

10

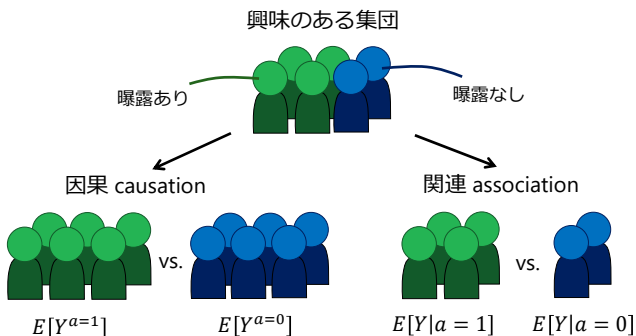
ID	$Y^{a=0}$	$Y^{a=1}$
1	0	1
2	1	0
3	0	0
4	0	0
5	0	0
6	1	0
7	0	0
8	0	1
9	1	1
10	1	0

ID	$Y^{a=0}$	$Y^{a=1}$
11	0	1
12	1	1
13	1	1
14	0	1
15	0	1
16	0	1
17	1	1
18	1	0
19	1	0
20	1	0

10

Association is not causation

11



Hernán MA, Robins JM (2020). Causal Inference. What If. Chapman & Hall/CRC. を基に作成

11

関連効果の指標

12

- ▶ 関連リスク差
 - ▶ $\Pr[Y = 1|A = 1] - \Pr[Y = 1|A = 0]$
- ▶ 関連リスク比
 - ▶ $\frac{\Pr[Y=1|A=1]}{\Pr[Y=1|A=0]}$
- ▶ 関連オッズ比
 - ▶ $\frac{\Pr[Y=1|A=1]/\Pr[Y=0|A=1]}{\Pr[Y=1|A=0]/\Pr[Y=0|A=0]}$

12

因果効果指標と関連効果指標

13

- ▶ 因果効果指標は定義、概念的なもの
 - ▶ 反事実アウトカムを用いて定義されるため
- ▶ 関連効果指標は観察データから求まる

- ▶ 関連効果指標をもって因果効果指標を求めるには？
 - ▶ どのような条件が成立すれば？
 - ▶ どのような解析を行えば？

13

交絡 confounding

14

- ▶ 実際の曝露群での結果と集団全体が曝露した場合が違
 - ▶ $E[Y^{a=1}|A=1] \neq E[Y^{a=1}]$

and / or

- ▶ 実際の非曝露群での結果と集団全体が曝露しなかった場合が違
 - ▶ $E[Y^{a=0}|A=0] \neq E[Y^{a=0}]$

14

ランダム化による交換可能性の成立

15

exchangeability

- ▶ 曝露群での結果と非曝露群が、仮に曝露を受けた場合の結果が一致（その逆も）
 - ▶ $\Pr[Y^{a=1}|A=1] = \Pr[Y^{a=1}|A=0]$
 $\Pr[Y^{a=0}|A=0] = \Pr[Y^{a=0}|A=1]$
 - ▶ $Y^a \perp\!\!\!\perp A$ for all a
- ▶ 片方の集団と全体集団での結果と一致
 - ▶ $\Pr[Y^{a=1}|A=1] = \Pr[Y^{a=1}|A=0] = \Pr[Y^{a=1}]$
 $\Pr[Y^{a=0}|A=0] = \Pr[Y^{a=0}|A=1] = \Pr[Y^{a=0}]$

15

交換可能性の意味

16

- ▶ 実際の曝露と反事実アウトカムが独立
 - ▶ 曝露とアウトカムの関連ナシではない！
- ▶ 交絡が生じる状況では交換可能性が不成立
 - ▶ 曝露群には実はdoomedな人だらけ
非曝露群には実はimmuneな人だらけ

16

条件付き交換可能性

17

- ▶ 予後因子 L が同じ値を持つ集団（層内）では交換可能性が成立
 - ▶ $\Pr[Y^{a=1}|A=1, L=1] = \Pr[Y^{a=1}|A=0, L=1]$
 $\Pr[Y^{a=0}|A=0, L=1] = \Pr[Y^{a=0}|A=1, L=1]$
 - ▶ $\Pr[Y^{a=1}|A=1, L=0] = \Pr[Y^{a=1}|A=0, L=0]$
 $\Pr[Y^{a=0}|A=0, L=0] = \Pr[Y^{a=0}|A=1, L=0]$
 - ▶ $Y^a \perp\!\!\!\perp A|L$ for all a
- ▶ No unmeasured confounding
 - ▶ 残差交絡 residual confounding がない

17

標準化 standardization

18

- ▶ 層ごとの結果の重み付き平均
$$\Pr[Y^a = 1] = \sum_l \Pr[Y^a = 1|L=l] \Pr[L=l]$$
 - ▶ 層別解析 stratified analysisのひとつ

18

練習② 標準化リスク差/比は？

19

ID	L	A	Y	ID	L	A	Y
1	0	0	0	11	1	0	0
2	0	0	1	12	1	1	1
3	0	0	0	13	1	1	1
4	0	0	0	14	1	1	1
5	0	1	0	15	1	1	1
6	0	1	0	16	1	1	1
7	0	1	0	17	1	1	1
8	0	1	1	18	1	1	0
9	1	0	1	19	1	1	0
10	1	0	1	20	1	1	0

19

回帰モデルによる標準化

20

- ▶ 層別しきれないほどの予後因子 L_1, L_2, \dots
- ▶ $E(Y|A, L_1, L_2, \dots)$ を回帰モデルで表現
 - ▶ 例えば、一般化線形モデルの利用
- ▶ 個人*i*ごとに反事実リスクを予測
 - ▶ $A = 1$ だった場合：

$$\hat{R}_i^{a=1} = E(Y|A = 1, L_1 = l_{1i}, L_2 = l_{2i}, \dots)$$
 - ▶ $A = 0$ だった場合：

$$\hat{R}_i^{a=0} = E(Y|A = 0, L_1 = l_{1i}, L_2 = l_{2i}, \dots)$$
- ▶ $\hat{R}_i^{a=1} - \hat{R}_i^{a=0}$ を全員で平均

20

回帰モデルによる効果推定

21

- ▶ 例：ロジスティック回帰モデル

$$\log(\text{オッズ}) = \beta_0 + \beta_1 \times A + \beta_2 \times L$$

オッズ	L = 0	L = 1
A = 1	$\exp(\beta_0 + \beta_1)$	$\exp(\beta_0 + \beta_1 + \beta_2)$
A = 0	$\exp(\beta_0)$	$\exp(\beta_0 + \beta_2)$

L = 0でのオッズ比

$$\frac{\exp(\beta_0 + \beta_1)}{\exp(\beta_0)} = \exp(\beta_1)$$

L = 1でのオッズ比

$$\frac{\exp(\beta_0 + \beta_1 + \beta_2)}{\exp(\beta_0 + \beta_2)} = \exp(\beta_1)$$

21

モデルによる標準化との違い

22

- ▶ 回帰係数；条件付き因果効果
 - ▶ 交絡変数の水準が同じ部分集団での効果
 - ▶ 部分集団間で効果が等しいと仮定
 - ▶ 交互作用項を含めない場合
- ▶ 標準化；周辺因果効果
 - ▶ 集団全体がA = 1だった場合と
集団全体がA = 0だった場合の差
 - ▶ 交絡変数の分布について周辺をとった

22

練習②スライドデータの要約

23

- ▶ L = 0である8人の曝露状況
 - ▶ 曝露なし (A = 0) が4人、イベント率25%
 - ▶ 曝露あり (A = 1) が4人、イベント率25%
- ▶ L = 1である12人の曝露状況
 - ▶ 曝露なし (A = 0) が3人、イベント率67%
 - ▶ 曝露あり (A = 1) が9人、イベント率67%

23

L = 1である12人

24

- ▶ 12人全員が非曝露であったら、
何名がイベントを起こすだろうか？
 - ▶ 実際の非曝露群では3人中2人
 - ▶ 交換可能性成立より、同じ割合でイベント発生
 - ▶ ゆえ、12人中8人がイベントを起こすだろう
- ▶ 12人全員が曝露であったら
何名がイベントを起こすだろうか？
 - ▶ 実際の曝露群では9人中6人
 - ▶ 交換可能性より、12人中8人だろう

24

擬似集団 pseudo-population

25

- ▶ 因果リスクを知る上で必要な、全員が曝露／非曝露である場合の仮想集団
- ▶ $L = 1$ では、実際に曝露を受けた人は9/12
 - ▶ 曝露を受けた割合の逆数をかけてみよう
 - ▶ $\frac{1}{\frac{9}{12}} = \frac{12}{9}$ 倍
- ▶ 曝露を受けなかった人は3/12
 - ▶ 逆数をかけよう； $\frac{1}{\frac{3}{12}} = 4$ 倍

25

逆確率重み付け法

26

- ▶ Inverse probability weighting method
 - ▶ IPW法
- ▶ 生成した擬似集団での関連効果の指標は
 - ▶ 擬似集団での因果効果の指標
 - ▶ 元の集団での因果効果の指標 に同じ
 - ▶ $Y^a \perp\!\!\!\perp A|L$ for all a ゆえ

26

練習③ 擬似集団を作ってみよう

27

ID	L	A	Y	ID	L	A	Y
1	0	0	0	11	1	0	0
2	0	0	1	12	1	1	1
3	0	0	0	13	1	1	1
4	0	0	0	14	1	1	1
5	0	1	0	15	1	1	1
6	0	1	0	16	1	1	1
7	0	1	0	17	1	1	1
8	0	1	1	18	1	1	0
9	1	0	1	19	1	1	0
10	1	0	1	20	1	1	0

27

擬似集団

28

	ID (人数)	曝露／非曝露確率	IP	擬似集団人数	擬似集団イベント数
L=0, A=0	1~4 (4人)				
L=0, A=1	5~8 (4人)				
L=1, A=0	9~11 (3人)				
L=1, A=1	12~20 (9人)				

28

練習④ 新たな以下の40例

29

L	A	$\gamma^{a=0}$	$\gamma^{a=1}$	人数	L	A	$\gamma^{a=0}$	$\gamma^{a=1}$	人数
0	0	1	1	3	1	0	1	1	2
0	0	1	0	3	1	0	1	0	2
0	0	0	0	9	1	0	0	0	1
0	1	1	1	1	1	1	1	1	6
0	1	1	0	1	1	1	1	0	6
0	1	0	0	3	1	1	0	0	3

- ▶ 集団全体、Lごとに観察データを分割表で要約してみよう

29

練習④ 分割表での要約

30

▶ $L = 0$

$L = 1$

	Y = 1	Y = 0		Y = 1	Y = 0
A = 1			A = 1		
A = 0			A = 0		

▶ 集団全体

	Y = 1	Y = 0
A = 1		
A = 0		

30

練習④ 議論しよう

31

- ▶ L による交絡が生じているか
- ▶ L を無視した場合のACEは？
 - ▶ 効果の指標：リスク差、リスク比
- ▶ L の水準ごとのACEは？
 - ▶ 効果の指標：リスク差、リスク比
 - ▶ この例では条件付き因果効果と同じ

31

効果修飾因子 effect modifier

32

- ▶ 効果修飾因子 V の水準によって A の Y に関する平均因果効果ACEが異なる
 - ▶ 特に、 V の水準によってACEの方向が逆転する場合を“質的な効果修飾”
- ▶ 曝露前に得られる変数
 - ▶ 曝露の影響を受ける変数は、中間変数かもしれない
- ▶ 効果指標に応じて修飾の有無が決まる

32

効果修飾がある ≠ 交絡がある

33

- ▶ 効果修飾因子はリスク因子でないかも
 - ▶ 例：ローマとギリシャで手術の質が違う
 - ▶ 国籍は必ずしも効果を修飾しない

33

ランダム化研究でも効果修飾

34

- ▶ 同じ疾患でも一部の集団にのみ治療効果がみられる薬剤が登場
 - ▶ 効果修飾がある
- ▶ 治療が効く人の分布によって、集団全体（周辺）の効果の大きさが変化

34

練習⑤ ランダム化試験

35

V	A	$\gamma^{a=0}$	$\gamma^{a=1}$	人数	V	A	$\gamma^{a=0}$	$\gamma^{a=1}$	人数
0	0	1	1	3	1	0	1	1	2
0	0	1	0	1	1	0	1	0	1
0	0	0	0	2	1	0	0	0	3
0	1	1	1	3	1	1	1	1	2
0	1	1	0	1	1	1	1	0	1
0	1	0	0	2	1	1	0	0	3

- ▶ リスク差、リスク比、オッズ比を効果の指標とした場合、効果修飾は？

35

併合可能性 collapsibility

36

- ▶ 集団全体の効果指標が層別の効果指標の重み付き平均であるか
 - ▶ リスク差、リスク比は collapsible
 - ▶ オッズ比は non-collapsible
- ▶ 例えば、層別オッズ比はすべて1でも、全体のオッズ比は1でない場合がある
 - ▶ 効果の修飾や交絡とは関係なく、効果指標の特徴

36

まとめ

37

- ▶ 反事実アウトカムと因果効果
 - ▶ 交換可能性と交絡
 - ▶ 標準化（層別解析）
 - ▶ 回帰モデルによる調整
 - ▶ 逆確率重み付け法
- ▶ 効果の修飾と交絡
 - ▶ 似て非なるもの
 - ▶ 効果修飾は効果指標に依存

37

教科書など

38

- ▶ Hernán MA, Robins JM. *Causal Inference*. Boca Raton: Chapman & Hall/CRC, forthcoming.
 - ▶ Hernanのwebより草稿を閲覧可能
- ▶ 黒木学. *構造的因果モデルの基礎*. 共立出版. 2017.
- ▶ Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology 3rd ed*. LWW. 2008.
- ▶ 田栗正隆. SASによる因果推論：CAUSALTRT プロシジャの紹介. 第36回SASユーザー総会. 2017/8/6.
 - ▶ Webより資料をDL可能

38