

2021/1/21 医療統計学（3年）⑥

# 反事実アウトカムと因果DAG



北海道大学 医学統計学  
横田 勲

1

## 今回の内容のポイント

2

- ▶ 因果推論の入門的話題をとして、以下の2つの因果モデルを扱う
  - ▶ 反事実アウトカムモデル、潜在アウトカム
  - ▶ 因果ダイアグラム（次回）
- ▶ 標的集団
  - ▶ どのような集団における効果なのか？

2

## Hさんがジムに！

4

- ▶ Hさんはなんと7kgもやせた！（事実）
  - ▶ さすがはプライベートジム！
- ▶ ジムは効果がある
  - ▶ 他の人にも同様・・・？
- ▶ ジムはHさんには効果がある
  - ▶ ただの正月太りが戻っただけかも・・・？

どちらも結論できない

4

## コントロール control

5

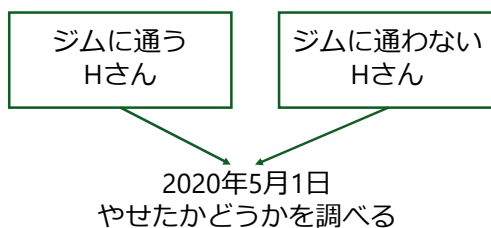
- ▶ 比較する相手
  - ▶ 対照
- ▶ 何かの効果を調べる場合には、比較が基本

5

## Hさん個人について

6

- ▶ ジムでやせるか？
- ▶ 2020年1月1日



6

## 反事実アウトカム

7

- ▶ 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
  - ▶ Hさんがジムに通って ( $a_H = 1$ )、4ヶ月後にやせた ( $Y_H = 1$ )
- ▶ 反事実データ counterfactual data
  - ▶ Hさんがジムに通わなかったら ( $a_H = 0$ )、4ヶ月後には？ ( $Y_H = ?$ )

7

### 潜在アウトカム potential outcome

8

- ▶  $Y^{a=1}$ 
  - ▶ 曝露  $a = 1$  を受けた場合のアウトカム
- ▶  $Y^{a=0}$ 
  - ▶ 曝露  $a = 0$  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

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

8

### 潜在アウトカムと観察アウトカム

9

- ▶ 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	$Y^{a=1}$	$Y^{a=0}$	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

9

### 個人での因果効果

10

	$Y^{a=1}$	$Y^{a=0}$	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

10

### 平均因果効果 Average Causal Effects

11

- ▶  $E[Y^{a=1}] - E[Y^{a=0}]$ 
  - ▶ 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

11

### 因果効果の指標

12

- ▶ 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
  - ▶  $\frac{\Pr[Y^{a=1} = 1] / \Pr[Y^{a=1} = 0]}{\Pr[Y^{a=0} = 1] / \Pr[Y^{a=0} = 0]}$

12

### 練習① 以下のACEは?

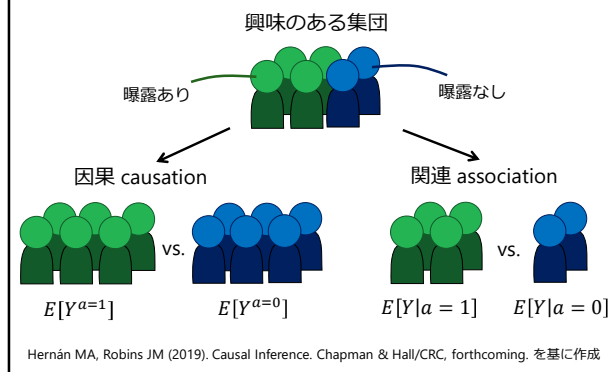
13

ID	$Y^{a=1}$	$Y^{a=0}$	ID	$Y^{a=1}$	$Y^{a=0}$
1	0	1	11	0	1
2	1	0	12	1	1
3	0	0	13	1	1
4	0	0	14	0	1
5	0	0	15	0	1
6	1	0	16	0	1
7	0	0	17	1	1
8	0	1	18	1	0
9	1	1	19	1	0
10	1	0	20	1	0

13

## Association is not causation

14



14

## 関連効果の指標

15

- ▶ 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
  - ▶  $\frac{\Pr[Y=1|A=1]/\Pr[Y=0|A=1]}{\Pr[Y=1|A=0]/\Pr[Y=0|A=0]}$

15

## 因果効果指標と関連効果指標

16

- ▶ 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？

16

## 交絡 confounding

17

- ▶ 実際の曝露群での結果と集団全体が曝露した場合が違う
  - ▶  $E[Y^{a=1}|A = 1] \neq E[Y^{a=1}]$
- 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う
  - ▶  $E[Y^{a=0}|A = 0] \neq E[Y^{a=0}]$

17

## ランダム化による交換可能性の成立

18

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

18

## 交換可能性の意味

19

- ▶ 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

19

### 条件付き交換可能性

20

- ▶ 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

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

$$L=0 \text{でのオッズ比} \quad \frac{\exp(\beta_0 + \beta_1)}{\exp(\beta_0)} = \mathbf{\exp(\beta_1)}$$

$$L=1 \text{でのオッズ比} \quad \frac{\exp(\beta_0 + \beta_1 + \beta_2)}{\exp(\beta_0 + \beta_2)} = \mathbf{\exp(\beta_1)}$$

21

### 練習② - 1 練習①データの続き

22

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

22

### 練習② - 2

23

- ▶ 粗リスク差を求めてみよう
  - ▶  $L$ を無視して、20名全体での曝露あり ( $A=1$ ) での発生リスクから曝露なし ( $A=0$ ) での発生リスクを引く
- ▶ 条件付きリスク差を求めてみよう
  - ▶  $L=0$ であったグループと  $L=1$ であったグループでそれぞれリスク差を計算

23

### 因果推論に必要なもの

24

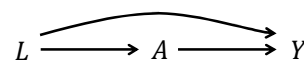
- ▶ 因果ネットワークに関する専門家の意見と検証不能な仮定
- ▶ 因果ダイアグラム causal diagram
  - ▶ 因果関係を仮定、図示化
  - ▶ 生じうるバイアスを整理
    - ▶ 交絡バイアス、選択バイアス、情報バイアス
  - ▶ 因果効果の分離
    - ▶ 直接効果・間接効果

24

### 有向非循環グラフ

25

- ▶ Directed Acyclic Graphs; DAGs
- ▶ Directed 有向
  - ▶ ノード node 間の矢線 arrow で順序性をいう
  - ▶  $L$ が $A$ の原因
- ▶ Acyclic 非循環
  - ▶ 自分自身の原因となることがない



25

## DAGで出てくる用語

26

- ▶ ノード、節点 node、点 vertex
  - ▶ 各変数をノードにおく
- ▶ 矢線 arrow、辺 edge
  - ▶ 一般に、辺は方向によらず使える言葉
- ▶ パス、経路、道 path
  - ▶ あるノードから異なるノードまでの行き方
- ▶ 親 parent
  - ▶ 祖先 ancestor : 親の親、その親・・・を含める
- ▶ 子 child
  - ▶ 子孫 descendant : 子の子、その子・・・を含める

26

## 因果DAG

27

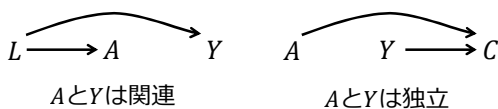
- ▶ 以下のようなDAG
  - ▶ ノード間を直接結ぶ矢線がない場合、直接(因果)効果がない
    - ▶ あるかもしれない、なら矢線を示しておく
  - ▶ ある変数達に共通する原因は、観察できないとしても、同じグラフ上に示す
  - ▶ いかなる変数もその子孫に対し原因となる
- ▶ 因果DAGは背景にある反事実モデルを表現

27

## 周辺独立 marginally independent

28

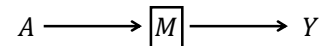
- ▶ 因果DAGにおける2変数間の特徴
- ▶ 以下のいずれかを満たせば“(周辺)関連”
  - ▶ 一方がもう一方の原因
  - ▶ 共通の原因(親)をもつ
- ▶ 関連しない場合、(周辺)独立



28

## 条件付き独立 conditional independence

29

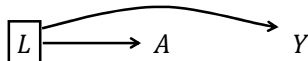


- ▶ AとYに周辺関連がある
  - ▶ Mは中間変数、媒介変数 mediator
- ▶ Mの水準を限定したら？
  - ▶ 条件付ける conditional on
  - ▶ □で囲う
- ▶ Mで条件付けることで、関連のあったパスA → M → Yをブロック
  - ▶ 条件付き独立にした

29

## 共通原因をBlocked

30

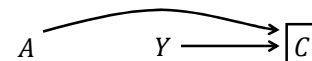


- ▶ AとYに周辺関連がある
  - ▶ Lが共通原因
- ▶ Lを条件付け
- ▶ 関連のあったパスA ← L → Yをブロック
  - ▶ 条件付き独立にした

30

## 合流点 collider をブロック

31

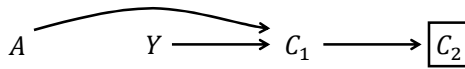


- ▶ AとYは周辺独立
  - ▶ A → C ← Yというパスは関連を生まない
  - ▶ Cが合流点 collider
- ▶ Cを条件付け
- ▶ A → C ← Yというブロックされていたパスをオープンに
  - ▶ 関連が生じる

31

### 合流点の子孫をブロック

32



- ▶ 合流点のみならず、その子孫についても AとYは原因となっていた
- ▶ C<sub>2</sub>で条件つけても、A → C<sub>1</sub> ← Yをオープンに
  - ▶ 直接の合流点C<sub>1</sub>で条件付けることと同様

32

### blockedかopenか

33

- ▶ パスがblockedな状況は以下のいずれか
  - ▶ 非合流点で条件付け
    - ▶ 中間変数や共通原因で条件付け
  - ▶ 合流点とその子孫は条件付けない
- ▶ blockedでないパスがopen path

33

### 有向分離 d-separation

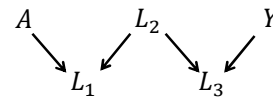
34

- ▶ 次の条件のいずれかを満たすとき、{A, Y}と排反な変数集合Sが A - Y間を有向分離する
  - ▶ A - Y間のすべてのパスにおける合流点で、その合流点と子孫がSに含まれないものがある
  - ▶ A - Y間のすべてのパスに非合流点で、Sに含まれるものがある
- ▶ Sで条件付ければ、A - Y間をつなぐパスをすべてblocked
  - ▶ open pathが含まれる場合をd-connected

34

### 練習③ 有向分離するSは？

35



- |                                       |       |  |
|---------------------------------------|-------|--|
| 1) { $\phi$ }                         | (空集合) | 5) {L <sub>2</sub> }                                   |
| 2) {L <sub>1</sub> }                  |       | 6) {L <sub>2</sub> , L <sub>3</sub> }                  |
| 3) {L <sub>1</sub> , L <sub>2</sub> } |       | 7) {L <sub>3</sub> }                                   |
| 4) {L <sub>1</sub> , L <sub>3</sub> } |       | 8) {L <sub>1</sub> , L <sub>2</sub> , L <sub>3</sub> } |

35

### バックドア基準 back-door criterion

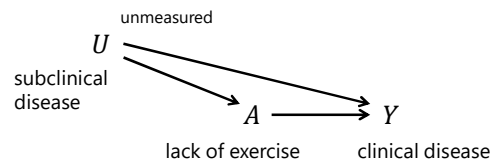
36

- ▶ AはYの非子孫
- ▶ 次の2条件を満たす頂点集合Sは A - Yについてバックドア基準を満たす
  - ▶ AからSの任意の要素へ有向道がない
  - ▶ Aから出る矢線をすべて除いたグラフにおいて、SがAとYを有向分離する
- ▶ S, A, Yが観察されていれば、AからYへの因果効果は識別可能

36

### 交絡の例①

37



- ▶ 逆因果 reverse causation
- ▶ A ← U → Yのバックドアパスが存在
  - ▶ Uをもし観察できれば、条件付けることでバックドアパスをブロック

37

### 交絡の例②

```

    graph LR
      U[atherosclerosis] --> L[heart disease]
      U --> Y[stroke]
      L --> A[aspirin]
      A --> Y
    
```

- ▶ 適応による交絡 confounding by indication
- ▶  $U$ か $L$ を条件付ければ $A \leftarrow L \leftarrow U \rightarrow Y$ をブロック
  - ▶  $L$ は $U$ を介して $Y$ に影響するため、経験的同定基準でも交絡因子として定義される

38

### 交絡の例③

```

    graph LR
      U1[pre-cancer lesion] --> L[diagnostic test]
      U1 --> Y[cancer]
      U2[health-conscious personality] --> L
      U2 --> A[physical activity]
      L --> A
      A --> Y
    
```

- ▶  $A \leftarrow U_2 \rightarrow L \leftarrow U_1 \rightarrow Y$ は既にブロック
  - ▶  $L$ が合流点ゆえ
  - ▶ しかし $L$ は経験的同定基準では交絡因子
- ▶  $L$ で条件付けるとバックドアパスが開く

39

### 選択バイアスの例①

```

    graph LR
      A[folic acid] --> Y[cardiac malformation]
      A --> C[alive]
      Y --> C
    
```

- ▶ 周産期の疫学研究でよくある例
  - ▶ 流産、死産例を無視することで生じる

40

### 選択バイアスの例②

```

    graph LR
      L[symptoms] --> A[HIV treatment]
      L --> Y[death]
      A --> C[uncensored]
      C --> Y
      U[immunosuppression] --> L
      U --> Y
    
```

- ▶  $C$ を条件付けると  $A \rightarrow C \leftarrow L \leftarrow U \rightarrow Y$ のパスが開く

41

### まとめ

- ▶ 反事実アウトカムと因果効果
  - ▶ 交換可能性と交絡
  - ▶ 条件付け解析
- ▶ 因果DAG
  - ▶ 有向分離 d-separation
  - ▶ 交絡バイアス、選択バイアス

42