

2020/6/22 北大SPH・統計的因果推論と臨床疫学④

傾向スコア解析



北海道大学 医学統計学
横田 勲

1

今回の内容

- ▶ 交絡と傾向スコア
- ▶ マッチング
- ▶ 重み付け解析

2

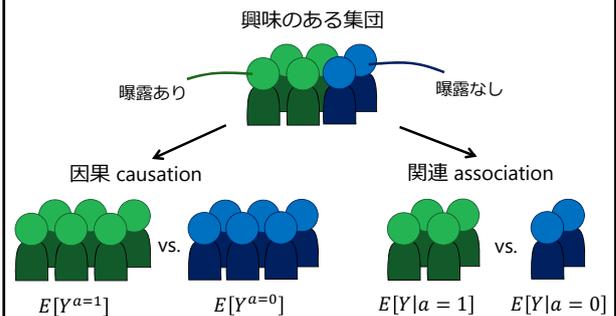
2

平均因果効果 Average Causal Effects 3

- ▶ $E[Y^{a=1}] - E[Y^{a=0}]$
 - ▶ 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

3

Association is not causation 4



Hernán MA, Robins JM (2019). Causal Inference. Chapman & Hall/CRC, forthcoming. を基に作成

4

交絡 confounding 5

- ▶ 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
 - ▶ $E[Y^{a=0}|A=0] \neq E[Y^{a=0}]$

5

ランダム化による交換可能性の成立 6

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

6

交換可能性の意味

7

- ▶ 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

7

条件付き交換可能性

8

- ▶ 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

8

バックドア基準 back-door criterion

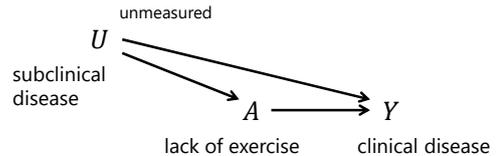
9

- ▶ A は Y の非子孫
- ▶ 次の2条件を満たす頂点集合 S は $A \rightarrow Y$ 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

9

交絡の例①

10

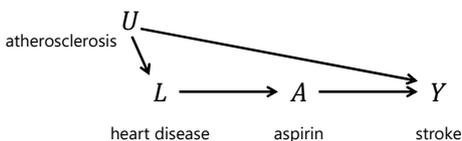


- ▶ 逆因果 reverse causation
- ▶ $A \leftarrow U \rightarrow Y$ のバックドアパスが存在
 - ▶ U をもし観察できれば、条件付けることでバックドアパスをブロック

10

交絡の例②

11



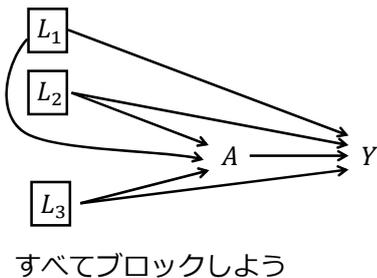
- ▶ 適応による交絡 confounding by indication
- ▶ U か L を条件付ければ $A \leftarrow L \leftarrow U \rightarrow Y$ をブロック
 - ▶ L は U を介して Y に影響するため、経験的同定基準でも交絡因子として定義される

11

交絡因子はたいてい複数

12

- ▶ ただし、すべて観察できている



12

交絡因子の候補が10個

13

- ▶ 層別解析、マッチング
 - ▶ すべて2値変数だったとしても1024層
 - ▶ マッチングでも10個の変数が完全に一致する相手はまれ
 - ▶ 連続変数だったら絶望的
- ▶ 回帰モデルの利用
 - ▶ 10パラメータで済む単純なモデル
 - ▶ 交絡因子間の関係は無視
 - ▶ 曝露効果の大きさは交絡因子によらない

13

回帰モデル

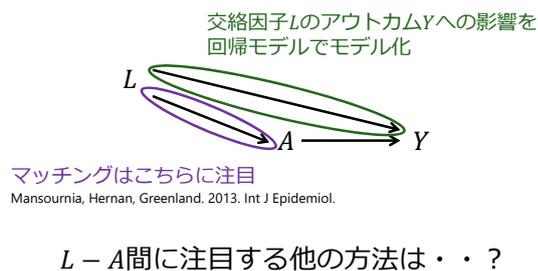
14

- ▶ 単純なモデル
 - ▶ 交絡変数による効果は全員で共通
- ▶ 交互作用を含めたモデル
 - ▶ 曝露や交絡因子の効果の大きさが集団間で異なること考慮
 - ▶ パラメータ数が増えてしまう
 - ▶ 精度が落ちる（信頼区間が広がる）おそれ

14

交絡因子の調整

15



15

$L - A$ 間に注目

16

- ▶ 曝露ありグループとなしグループで交絡因子の分布が異なる
- ▶ 交絡因子によって曝露（治療）の受けやすさが異なる

16

傾向スコア $\pi(L)$

17

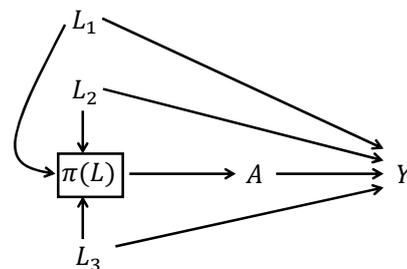
Rosenbaum, Rubin. Biometrika. 1983.

- ▶ 対象者それぞれの曝露（治療）の受けやすさ（確率）
- ▶ 1:1割付のランダム化研究
 - ▶ $\pi(L)$ は全員0.5で共通
- ▶ 観察研究
 - ▶ $\pi(L)$ を対象者ごとに予測
 - ▶ $Y^a \perp\!\!\!\perp A|L \Rightarrow Y^a \perp\!\!\!\perp A|\pi(L)$
 - ▶ 交絡因子をすべて特定している
 - ▶ 交絡因子から傾向スコアを正しく求められている
 - ▶ $\pi(L) = 0, 1$ となる対象者がいない

17

傾向スコアを因果DAGで表現

18



18

20人の仮想データ

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

L = 1である12人

20

- ▶ 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

20

擬似集団 pseudo-population

21

- ▶ 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

21

逆確率重み付け法

22

- ▶ 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

22

(再) 擬似集団を作ってみよう

23

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

23

擬似集団

24

	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)	$\frac{4}{8}$	2倍	8	2
L=0, A=1	5~8 (4人)	$\frac{4}{8}$	2倍	8	2
L=1, A=0	9~11 (3人)	$\frac{3}{12}$	4倍	12	8
L=1, A=1	12~20 (9人)	$\frac{9}{12}$	1.33倍	12	8

24

解析の流れ

25

- ▶ 対象者ごとの傾向スコアを推定
- ▶ 傾向スコアで調整した解析
 - ▶ 回帰分析で共変量に
 - ▶ 限定、層別解析
 - ▶ マッチング
 - ▶ 重み付け解析
- ▶ 交絡因子のバランスがとれたかを確認
- ▶ 平均因果効果の推定
+ 感度解析

25

傾向スコアの作り方

26

- ▶ 対象者ごとの曝露確率の予測値を得たい
 - ▶ 交絡因子のバランスがとれるような予測値
- ▶ どんな統計手法を使ってもよい
 - ▶ ロジスティック回帰
 - ▶ 層別解析
 - ▶ プロビット回帰やノンパラメトリック回帰
 - ▶ 分類木
 - ▶ 機械学習の手法
 - ▶ . . .

26

NHEFSデータの一部

27

National Health and Nutrition Examination Survey Data I Epidemiologic Follow-up Study

- ▶ 25歳から74歳の喫煙者1,566名の11年追跡
- ▶ 曝露 A
 - ▶ ベースライン時に喫煙習慣をやめた($A = 1$)
 - ▶ 喫煙習慣はやめなかった($A = 0$)
- ▶ アウトカム Y : 体重変化量(kg)
- ▶ 交絡 L
 - ▶ 年齢、性別、人種、学歴、ベースライン体重、喫煙本数、喫煙期間、運動習慣、活動性など

27

傾向スコアモデル

28

- ▶ 対象者ごとに喫煙習慣をやめる確率を予測するモデルを作成
 - ▶ 喫煙習慣をアウトカムとして交絡を説明変数に加えたロジスティック回帰
- ▶ 交絡因子 (ベースライン時)
 - ▶ 性別、人種、年齢(2次項まで)、教育歴、喫煙本数(2次項まで)、喫煙期間(2次項まで)、運動習慣、活動性、体重(2次項まで)

28

ベースライン情報①

29

平均, 頻度(割合)	喫煙グループ $A = 0$ ($N = 1,163$)	禁煙グループ $A = 1$ ($N = 403$)
性別 (男性)	542 (47%)	220 (55%)
(女性)	621 (53%)	183 (45%)
人種 (白人)	993 (85%)	367 (91%)
(その他)	170 (15%)	36 (9%)
年齢	42.8	46.2
教育歴 (大卒以上)	115 (10%)	62 (15%)
(大学中退)	92 (8%)	29 (7%)
(高卒)	480 (41%)	157 (39%)
(高校中退)	266 (23%)	74 (18%)
(中卒以下)	210 (18%)	81 (20%)

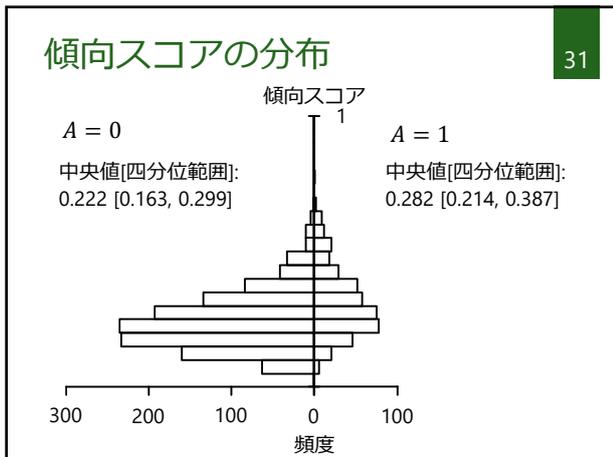
29

ベースライン情報②

30

平均, 頻度(割合)	喫煙グループ $A = 0$ ($N = 1,163$)	禁煙グループ $A = 1$ ($N = 403$)
喫煙本数/日	21.2	18.6
喫煙年数	24.0	26.0
運動 (とてもする)	237 (20%)	63 (16%)
(そこそこ)	485 (42%)	176 (44%)
(ほぼしない)	441 (38%)	164 (41%)
活動 (とても活動的)	532 (46%)	170 (42%)
(そこそこ)	527 (45%)	188 (47%)
(非活動的)	104 (9%)	45 (11%)
体重	70.3	72.4

30



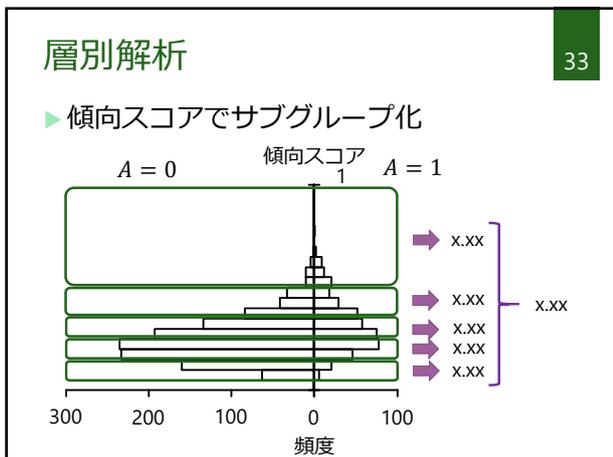
31

解析手段

- ▶ 回帰分析の説明変数に加える
- ▶ 限定、層別解析
- ▶ マッチング
- ▶ 重み付き解析

治療を受ける確率は40%

32



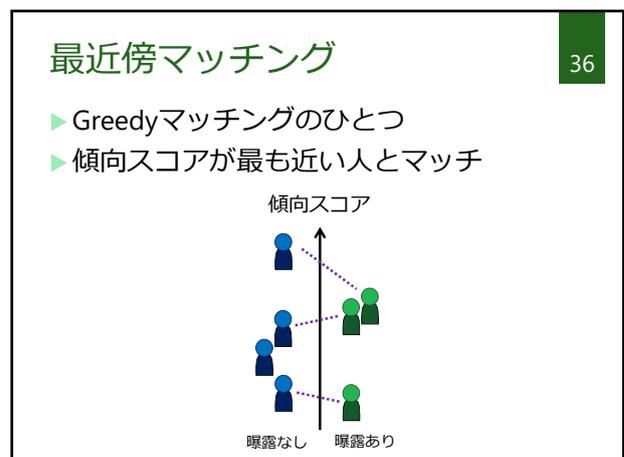
33

- ### 層別解析の前提
- ▶ 残差交絡がない
 - ▶ 層内では交絡が起こっていない
 - ▶ 層別で十分に交絡を調整できる
 - ▶ 異なる傾向スコアをもつ対象者が同じ層にいるのに残差交絡がない??
 - ▶ 離散的な傾向スコアならいいかも
 - ▶ 重み付け解析に比べて利点がほぼない

34

- ### マッチング
- ▶ 同じ傾向スコアを持つ対象者同士を組に
 - ▶ マッチングの種類
 - ▶ マッチングアルゴリズム
 - ▶ 最近傍マッチング、最適マッチング
 - ▶ キャリバーの設定
 - ▶ 復元か非復元か
 - ▶ マッチする人数比

35

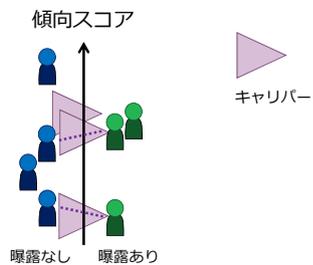


36

キャリア付き最近傍マッチング

37

- ▶ 許容されるマッチ候補との傾向スコアの差
- ▶ キャリパー内に候補複数いたらランダムに選ぶことも

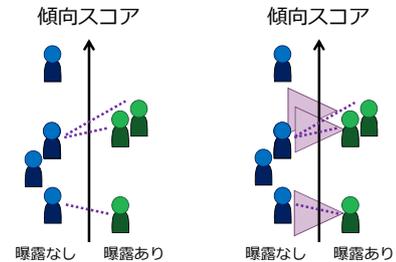


37

復元マッチング

38

- ▶ マッチされる相手として繰り返し登場可
- ▶ 非復元：重婚を認めない



38

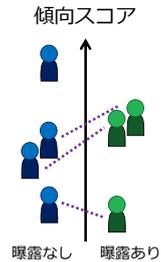
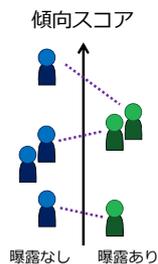
初期値による影響

39

- ▶ マッチさせる順番

- ▶ 小さいほうから

- ▶ 大きいほうから



39

マッチする人数比

40

- ▶ 1:1 マッチ
 - ▶ 群間で人数がそろろう
- ▶ 1:k マッチ
- ▶ 変動比マッチ
 - ▶ 対象者によってマッチする人数が異なる
 - ▶ 曝露グループの人数がとても少ない場合

40

最適マッチング

41

- ▶ 傾向スコアの差の合計が最小になるようあらゆるペアを探索

41

マッチングの利点・弱点

42

- ▶ 利点
 - ▶ ペアになるよう人を選ぶだけでわかりやすい
 - ▶ あたかもランダム化したような気分になれる
- ▶ 弱点
 - ▶ 自分で作った傾向スコアが同じになるよう人を選んだので別にランダム化ではない
 - ▶ 選ばれなかった人の情報を捨てる効率ロス
 - ▶ マッチングの手順を考慮した解析が必要?

42

練習② マッチング

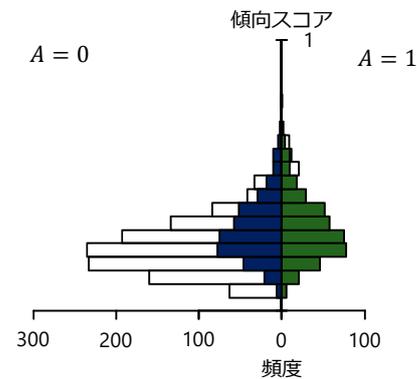
43

- ▶ 以下の傾向スコア
 - ▶ 曝露グループ
0.10, 0.20, 0.40, 0.50, 0.80
 - ▶ 非曝露グループ
0.03, 0.15, 0.30, 0.35, 0.60
- ▶ マッチング方法
 - ▶ 最近傍1:1
 - ▶ (小さいほうから・大きいほうから)
 - ▶ キャリバーなし、キャリバー：±0.05
 - ▶ 最適マッチング

43

マッチされた人・されない人

44



44

1:1最近傍マッチングでの背景①

45

平均, 頻度(割合)	喫煙グループ A = 0 (N=396)	禁煙グループ A = 1 (N=396)
性別 (男性)	227 (57%)	216 (55%)
(女性)	169 (43%)	180 (45%)
人種 (白人)	364 (92%)	360 (91%)
(その他)	32 (8%)	36 (9%)
年齢	46.2	45.9
教育歴 (大卒以上)	57 (14%)	62 (16%)
(大学中退)	31 (8%)	27 (7%)
(高卒)	152 (38%)	154 (39%)
(高校中退)	78 (20%)	74 (19%)
(中卒以下)	78 (20%)	79 (20%)

45

1:1最近傍マッチングでの背景②

46

平均, 頻度(割合)	喫煙グループ A = 0 (N=396)	禁煙グループ A = 1 (N=396)
喫煙本数/日	19.3	18.7
喫煙年数	26.7	26.1
運動 (とてもする)	70 (18%)	62 (16%)
(そこそこ)	176 (44%)	174 (44%)
(ほぼしない)	150 (38%)	160 (40%)
活動 (とても活動的)	163 (41%)	167 (42%)
(そこそこ)	184 (46%)	186 (47%)
(非活動的)	49 (12%)	43 (11%)
体重	71.8	72.2

46

標準化差で評価

47

- ▶ 曝露グループ間での背景因子のバランス
 - ▶ 検定では確認しない
 - ▶ 標準化差が±0.1、±0.25以内に入ればよい

$$\frac{\bar{X}_{a=1} - \bar{X}_{a=0}}{s.d.}$$

- ▶ ただし、 $s.d. = \sqrt{\frac{\text{Var}(X_{a=1}) + \text{Var}(X_{a=0})}{2}}$

47

1:1最近傍マッチングでの標準化差

48

	マッチング前	マッチング後
性別	-0.16	0.06
人種	-0.18	0.03
年齢	0.28	0.02
喫煙本数/日	-0.22	-0.05
喫煙年数	0.16	-0.05
体重	0.13	0.03

- ▶ 交絡因子のバランスがうまくとれたよう

48

体重変化量Yの解析

49

- ▶ 通常のt検定
 - ▶ マッチング前 : 2.54 (95%CI: 1.66, 3.43)
 - ▶ マッチング後 : 3.39 (95%CI: 2.29, 4.50)

49

重み付け解析

50

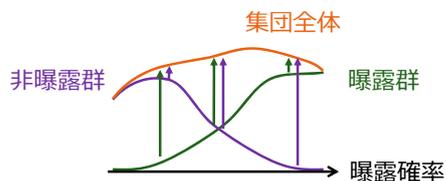
- ▶ 逆確率重み付け法, IPW法
Inverse Probability of Weighting

50

IPW解析での擬似集団①

51

- ▶ 重みの作り方で、いろいろな集団に
 - ▶ Average treatment effect
 - ▶ 曝露群では曝露確率の逆数
 - ▶ 非曝露群では非曝露確率の逆数

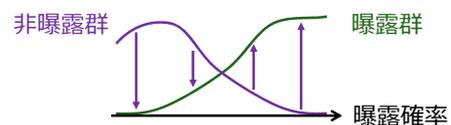


51

IPW解析での擬似集団②

52

- ▶ 重みの作り方で、いろいろな集団に
 - ▶ Average treatment effect as treated (exposed)
 - ▶ 曝露群はそのまま
 - ▶ 非曝露群では曝露確率のオッズ



52

重み付け解析の結果

53

- ▶ ATEを推定
 - ▶ 3.38 (95%CI: 2.42, 4.35)
- ▶ ATTを推定
 - ▶ 3.36 (95%CI: 2.40, 4.33)

53

重み付け解析の利点・弱点

54

- ▶ 利点
 - ▶ 狙った集団での平均因果効果を自在に推定
- ▶ 弱点
 - ▶ アウトカムの解析で重みを考慮した信頼区間
 - ▶ SAS proc causaltrtで簡便に出力できるように
 - ▶ 傾向スコアが0や1に近い場合、大きな重みとなりすぎる
 - ▶ 結果が不安定に
 - ▶ 安定化重み stabilized weightの利用が推奨

54

まとめ

55

- ▶ 傾向スコア
 - ▶ 平均因果効果を推定する強力な手段
 - ▶ 交絡因子がたくさんある場合に有用
- ▶ マッチング
- ▶ 重み付け解析
 - ▶ それぞれの利点・弱点

55