

東京大学SPH 特別講義シリーズ2023
因果推論の統計学的アプローチ

第4回

時間変動する曝露の効果を推定する

*Associate Professor
Kosuke Inoue, MD, PhD*

*Hakubi Center, Department of Social Epidemiology,
Department of Diabetes, Endocrinology, and Nutrition,
Kyoto University*

**Special thanks to Yusuke Okubo & Tomohiro Shinozaki*

本日のポイント

時間変動する曝露の効果を推定する際は

- DAGを用いることで変数の複雑なかわりを整理できる。
- 時間変動する交絡因子について考える必要がある。
- Inverse probability of treatment weight、G-computationといった手法が使える。
- 問いを明確にすることが重要である。

本日のアジェンダ

- Recap (DAG, Probability)
- 時間変動する曝露の効果を推定する
 - Inverse probability of treatment weight
 - G-computation
- 実例紹介 (Sustained Treatment, Different time point, Subsequent symptoms)
- $+\alpha$: 時間変動するリスクを推定する

本日のアジェンダ

- Recap (DAG, Probability)
- 時間変動する曝露の効果を推定する
 - Inverse probability of treatment weight
 - G-computation
- 実例紹介 (Sustained Treatment, Different time point, Subsequent symptoms)
- + α : 時間変動するリスクを推定する

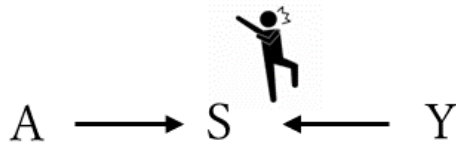
DAGを扱う際の基本事項



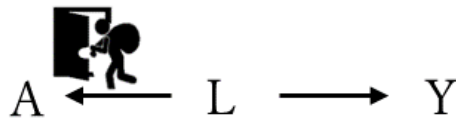
前向きに進み続けられる矢線は、スタート地点の変数 (A) がゴール地点の変数 (Y) に与える**影響**を示す。



通過地点がある場合はそれを**中間因子**と呼び、前向きに進め続けられるため、経路が“**開いている**”、と判断する。



途中で矢線が向き合っている場合はそこを**合流点**と呼び、それ以上前に進めなくなってしまうため、経路が“**閉じている**”、と判断する。



Aに向かう矢線から始まる経路は**バックドア経路**と呼び、その後の経路上の変数で矢線が向き合っていない場合は、経路が“**開いている**”、と判断する。

Disjunctive Cause Criterion

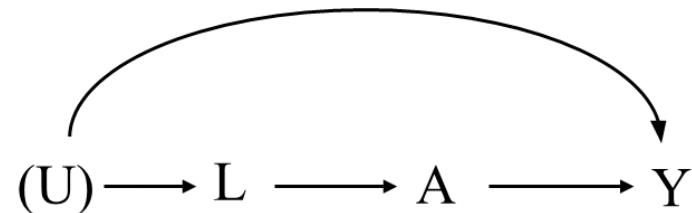
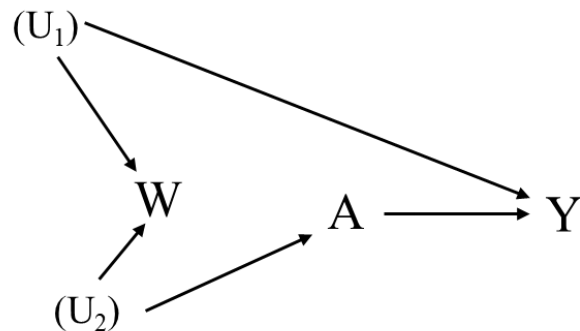
Pretreatment Criterion: 変数を必要以上に選んでしまう。

Common Cause Criterion: 有用な変数を取りこぼしてしまう。

⇒ **Disjunctive Cause Criterion**: 両者を掛け合わせたアプローチ

★曝露より先に測定された変数で、曝露・アウトカムいずれかの原因であれば調整する

このアプローチを用いると、先述の例でも適切に調整変数を選択し、バイアスを回避することができる。

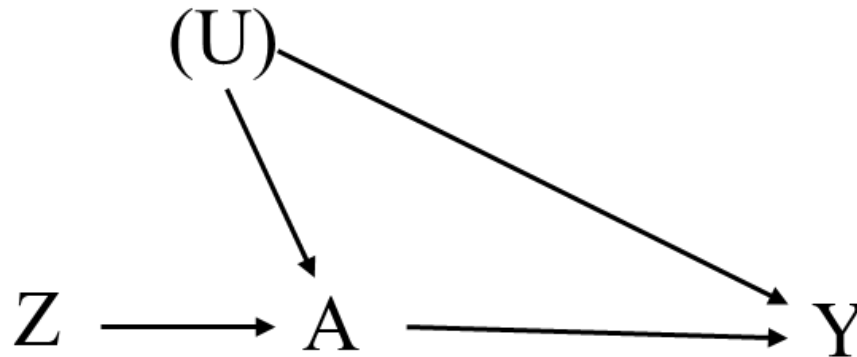


Modified Disjunctive Cause Criterion

Modified Disjunctive Cause Criterion

★曝露より先に測定された変数で、曝露・アウトカムいずれかの原因であれば調整する ＋操作変数を含めない

Disjunctive Cause Criterionだと、操作変数にあたるZが曝露の原因であるため、調整する変数として選択されてしまう。一方で、操作変数Zを調整すると、未測定変数Uによる交絡を増幅することが知られているため、含めるべきではない。

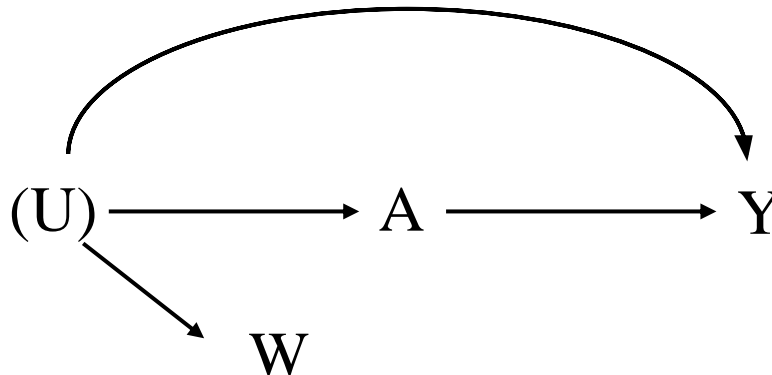


Modified Disjunctive Cause Criterion

Modified Disjunctive Cause Criterion

★曝露より先に測定された変数で、曝露・アウトカムいずれかの原因であれば調整する + 曝露とアウトカムの共通の原因の proxyを含める

下記のDAGにおいて、未測定交絡因子UのproxyであるWは Disjunctive Cause Criterionでは選択されない。一方でWを調整することで、Uによるバイアスを軽減する可能性がある。



Causal Assumptions

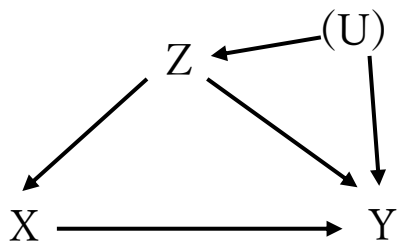
仮定の名称	意味*
条件付き交換可能性 (Conditional Exchangeability)	Zで条件付けた際に、X-Y間に未調整交絡因子がなく、全てのバックドア経路が閉じられている。
正值性 (Positivity)	各Zの組み合わせにおいて $P(X) > 0$ でない。
一貫性 (Consistency)	Xがある値のときに実際に観測されたYと、Xがその値を取るように仮想的介入がなされたときに観察されるであろうYとが一致する。
個体間の相互作用なし (No interference)	ある個人におけるXの影響は、他の個人のXによって影響を受けない。
正しいモデルの設定 (No model misspecification)	アウトカムとの関連に必要なモデルが正しく設定されている
測定誤差がない (No measurement error)**	モデルに用いる変数は全て正しく測定されている。
選択バイアスがない (No selection bias)**	対象者の(意図的・偶発的)選択によるバイアスがない。

Homework 1

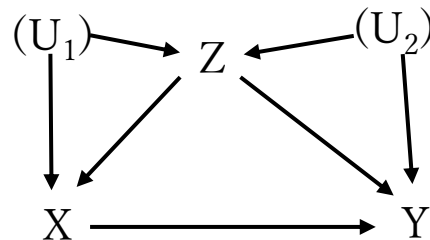
XのYに対する因果効果をバイアスなく求めることができるDAGは次のうちどれ？

※(U)は未測定交絡因子、[S]はSで条件付けされていることを示す

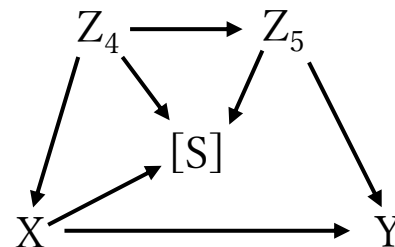
A



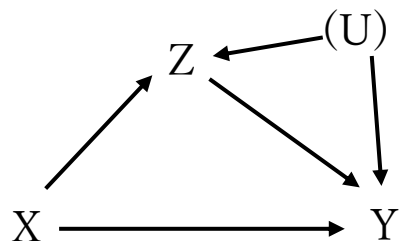
B



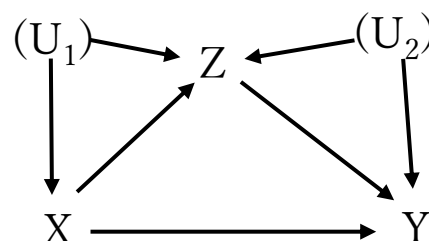
C



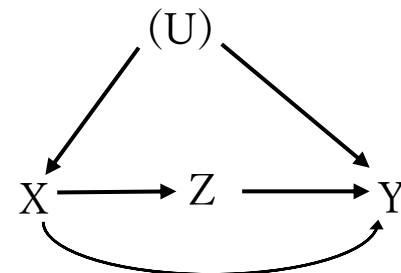
D



E



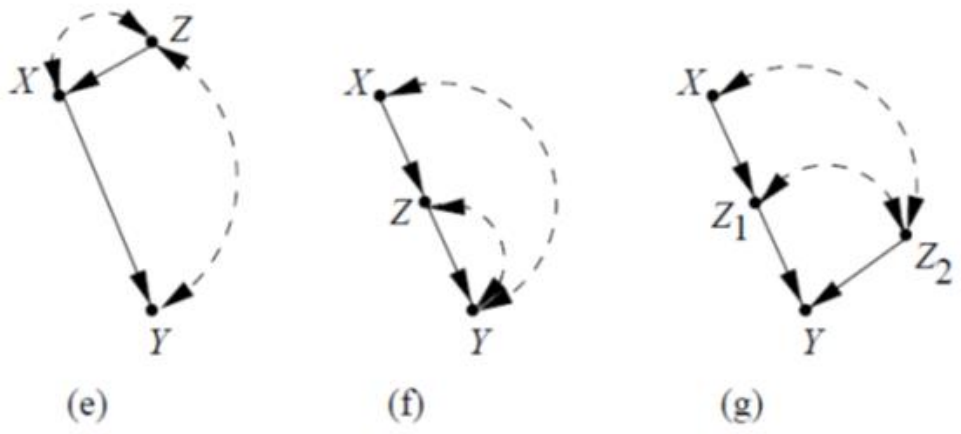
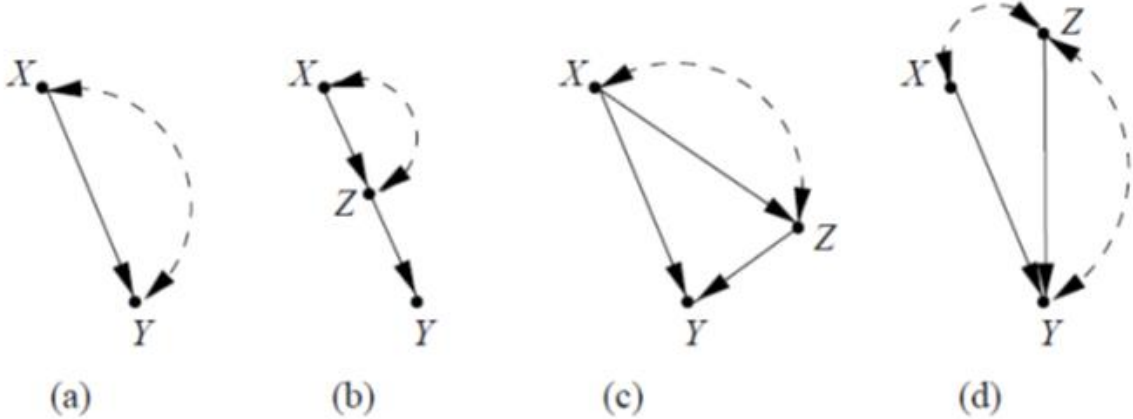
F



Homework 2

XのYに対する因果効果をバイアスなく求めることができるDAGは次のうちどれ？

※<----->は未測定交絡因子の存在をしめす



Homework 3

Option 1. Draw DAG of your own research

Option 2. Draw DAG of the following paper

Abstract

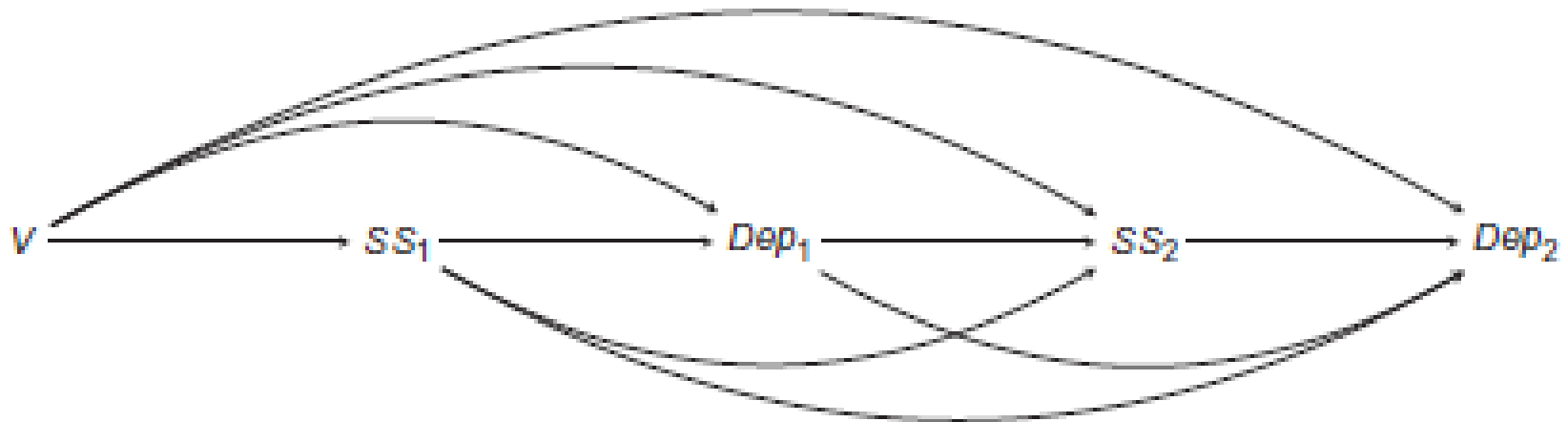
We used marginal structural models to evaluate associations of social support with antepartum depression in late pregnancy, if everyone had had high social support both before pregnancy and during early pregnancy, compared with having low social support at one of the 2 time points or low social support at both time points. In 2012–2014, pregnant Peruvian women ($n = 3,336$) were recruited into a prospective cohort study (at a mean gestational age of 9 weeks). A follow-up interview ($n = 2,279$) was conducted (at 26–28 weeks of gestation). Number of available support providers and satisfaction with social support were measured using Sarason Social Support Questionnaire–6. Depression was measured using the Edinburgh Postnatal Depression Scale. Low number of support providers at both time points was associated with increased risk of depression (odds ratio = 1.62, 95% confidence interval: 1.12, 2.34). The association for low satisfaction at both time points was marginally significant (odds ratio = 1.41, 95% confidence interval: 0.99, 1.99). Depression risk was not significantly higher for women who reported high social support at one of the 2 time points. Our study reinforces the importance of assessing social support before and during pregnancy and underscores the need for future interventions targeted at increasing the number of support providers to prevent antepartum depression.

American journal of epidemiology, 187(9), pp.1871-1879.

Original Contribution

Causal Model of the Association of Social Support With Antepartum Depression: A Marginal Structural Modeling Approach

Qiu-Yue Zhong*, Bizu Gelaye, Tyler J. VanderWeele, Sixto E. Sanchez, and Michelle A. Williams



Recap of Probability function

1. Product decomposition

$$\begin{aligned}P(A,B,C,D) &= P(A|B, C, D) P(B, C, D) \\ &= P(A|B, C, D) P(B|C, D) P(C, D) \\ &= P(A|B, C, D) P(B|C, D) P(C|D)P(D)\end{aligned}$$

2. Conditional Probability

$$\begin{aligned}P(A,B|C,D) &= P(A,B,C,D)/P(C,D) \\ &= P(A|B,C,D)P(B|C,D)P(C,D) /P(C,D) \\ &= P(A|B,C,D)P(B|C,D)\end{aligned}$$

Recap of Probability function

3. Independence

$P(A, B | C) = P(A|B, C)P(B|C) = P(A)P(B|C)$ when $P(A|B, C)=P(A)$

$P(A, B | C) = P(A|B, C)P(B|C) \neq P(A|C)P(B|C)$ when $P(A|B)=P(A)$

4. Marginalization

$$P(A) \sum_b P(B) = P(A)$$

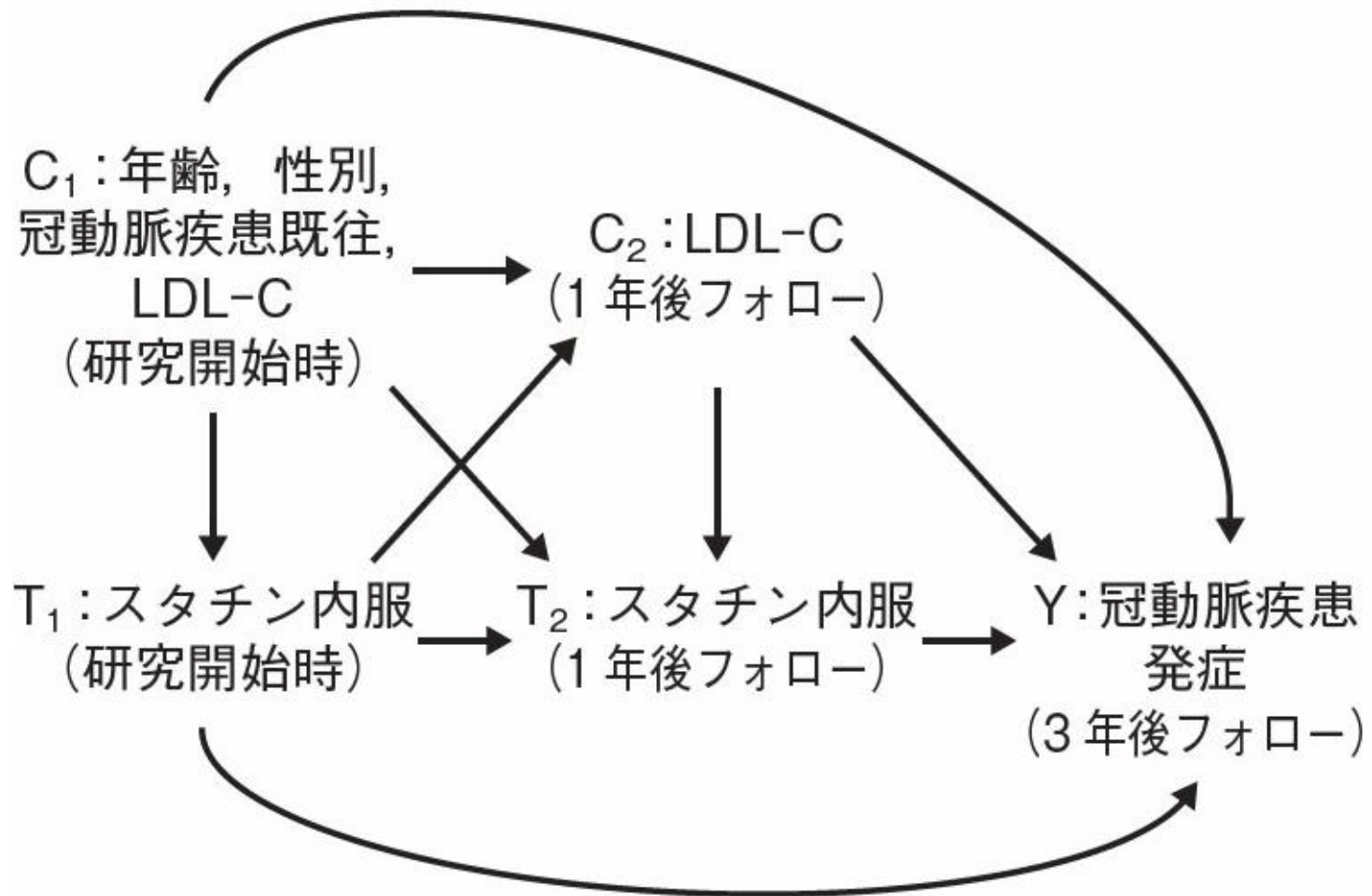
$$\sum_b P(A|B)P(B) = P(A)$$

$$\sum_{b,c} P(A, B, C) = P(A)$$

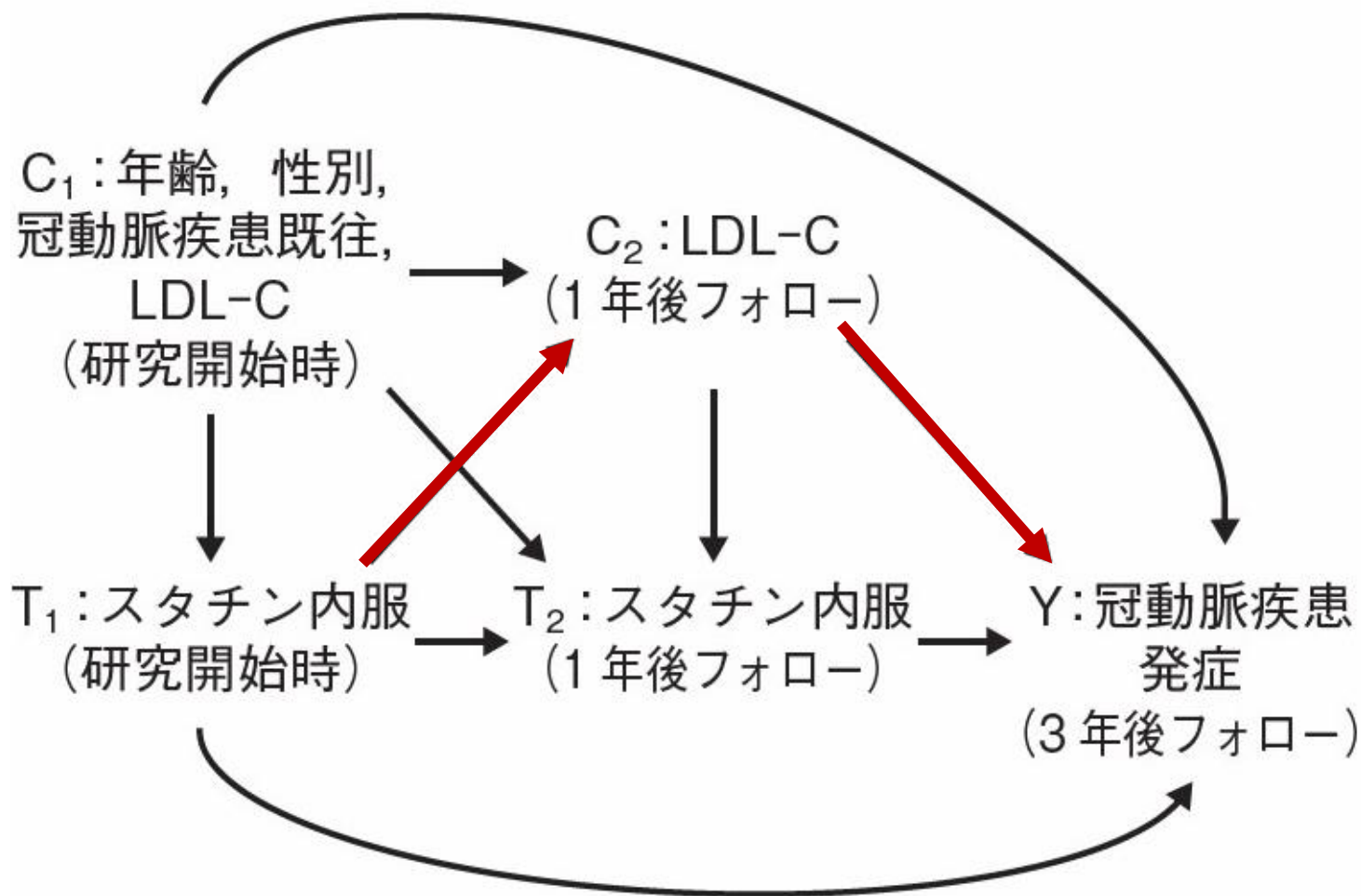
本日のアジェンダ

- Recap (DAG, Probability)
- 時間変動する曝露の効果を推定する
 - **Inverse probability of treatment weight**
 - G-computation
- 実例紹介 (Sustained Treatment, Different time point, Subsequent symptoms)
- + α : 時間変動するリスクを推定する

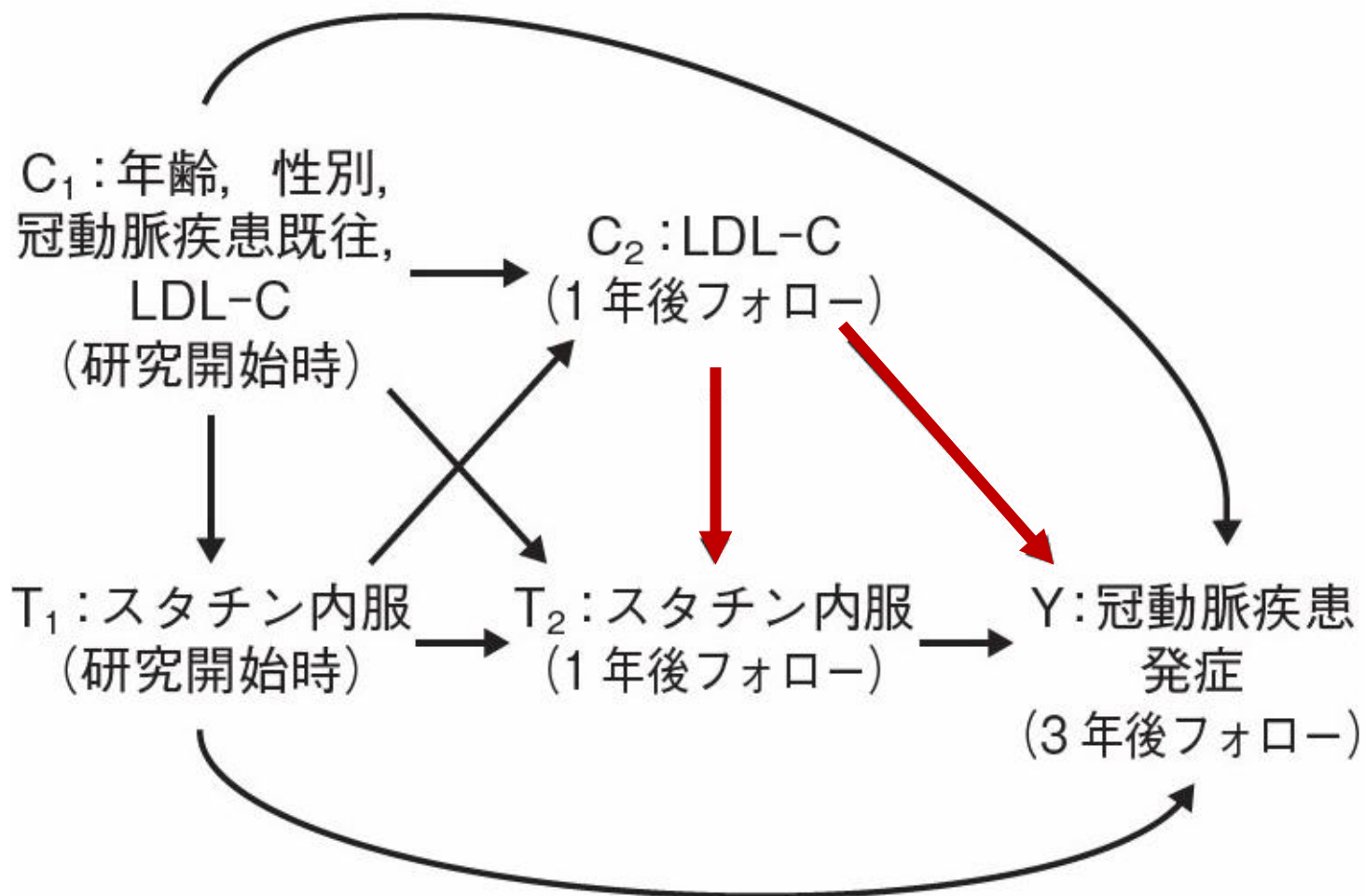
なぜ時間変動性曝露を考える必要があるのか？



時間変動性曝露を考える際の問題点



時間変動性曝露を考える際の問題点



Notation / 定義

T : 曝露/治療

Y : アウトカム

C : 交絡変数

潜在アウトカム変数

$Y^{T=1}$: T=1を受けた場合に観察されるであろうアウトカム

$Y^{T=0}$: T=0を受けた場合に観察されるであろうアウトカム

平均因果効果

$$E(Y^{T=1}) - E(Y^{T=0})$$

$$\sum_c E(Y^{T=1} | C = c) P(C = c)$$

$$= \sum_c E(Y^{T=1} | T = 1, C = c) P(C = c)$$

※条件付き交換可能性

$$= \sum_c E(Y | T = 1, C = c) P(C = c)$$

※一貫性

$$\sum_c E(Y^{T=0} | C = c) P(C = c)$$

$$= \sum_c E(Y^{T=0} | T = 0, C = c) P(C = c)$$

※条件付き交換可能性

$$= \sum_c E(Y | T = 0, C = c) P(C = c)$$

※一貫性

$$E(Y^{T=1})$$

$$= \sum_c E(Y^{T=1} | C = c) P(C = c)$$

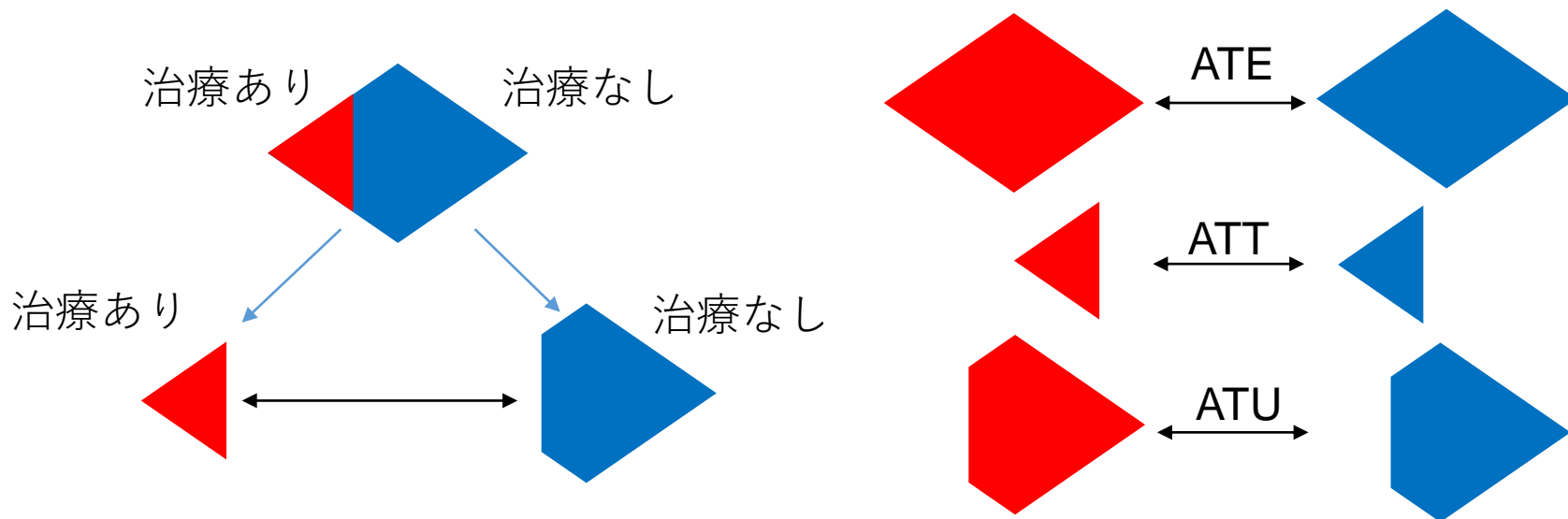
$$= \sum_c E(Y^{T=1} | T = 1, C = c) P(C = c)$$

※条件付き交換可能性

$$= \sum_c E(Y | T = 1, C = c) P(C = c)$$

※一緻性

Inverse probability of treatment weight



治療を受けたグループと受けなかったグループは異なる！

ATE (Average Treatment Effect)

「もし全員が治療を受けたら？ 全員が受けなかったら？」

ATT (Average Treatment Effect **on the Treated**)

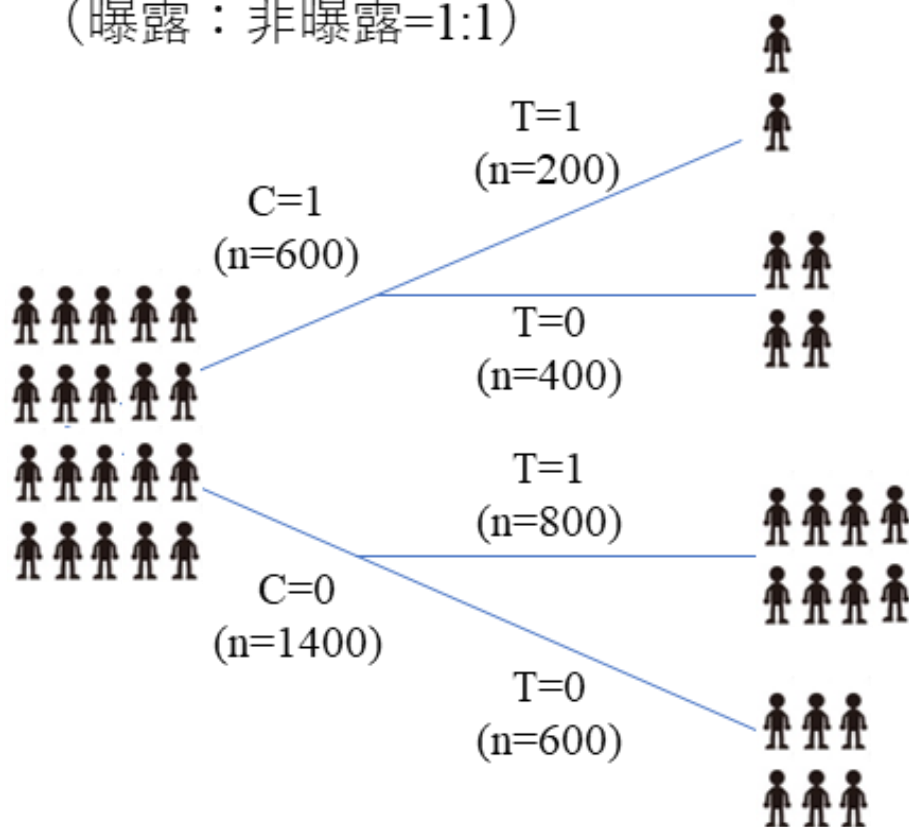
「もし治療したグループが、治療されなかったら？」

ATU (Average Treatment effect **on the Untreated**)

「もし治療しなかったグループが、治療されたら？」

Inverse probability of treatment weight

例：対象集団2000人
(曝露：非曝露=1:1)



重みの計算

$$\frac{P(T=1)}{P(T=1|C=1)} = \frac{(200+800)/(2000)}{200/(200+400)} = 1.5$$

$$\frac{P(T=0)}{P(T=0|C=1)} = \frac{(400+600)/(2000)}{400/(200+400)} = 0.75$$

$$\frac{P(T=1)}{P(T=1|C=0)} = \frac{(200+800)/(2000)}{800/(800+600)} = 0.875$$

$$\frac{P(T=0)}{P(T=0|C=0)} = \frac{(400+600)/(2000)}{600/(800+600)} = 1.17$$

Inverse probability of treatment weight

- それぞれのデータに「逆確率」を重み付けすることで、交絡を対処する
- 逆確率とは？（Non-stabilized IPW）
傾向スコアの逆数： $1/PS = 1/P(T=1|C)$
計測されたデータ " $P(Y, T, C)$ " に逆確率を重み付けすることで、潜在アウトカムが計測できる

Inverse probability of treatment weight

- それぞれのデータに「逆確率」を重み付けすることで、交絡を対処する
- 逆確率とは？ (Non-stabilized IPW)
傾向スコアの逆数： $1/PS = 1/P(T=1|C)$
計測されたデータ "P(Y, T, C)" に逆確率を重み付けすることで、潜在アウトカムが計測できる

$$\begin{aligned} P(Y^T) &= \sum_c P(Y^T | C) P(C) \\ &= \sum_c P(Y|T, C) P(C) && \text{G-formula: Empirical Analogue} \\ &= \sum_c \frac{P(Y, T, C)}{P(T, C)} P(C) \\ &= \sum_c \frac{P(Y, T, C)}{P(T|C)P(C)} P(C) \\ &= \sum_c P(Y, T, C) \times \frac{1}{P(T|C)} && \text{Inverse probability weight} \end{aligned}$$

Inverse probability of treatment weight

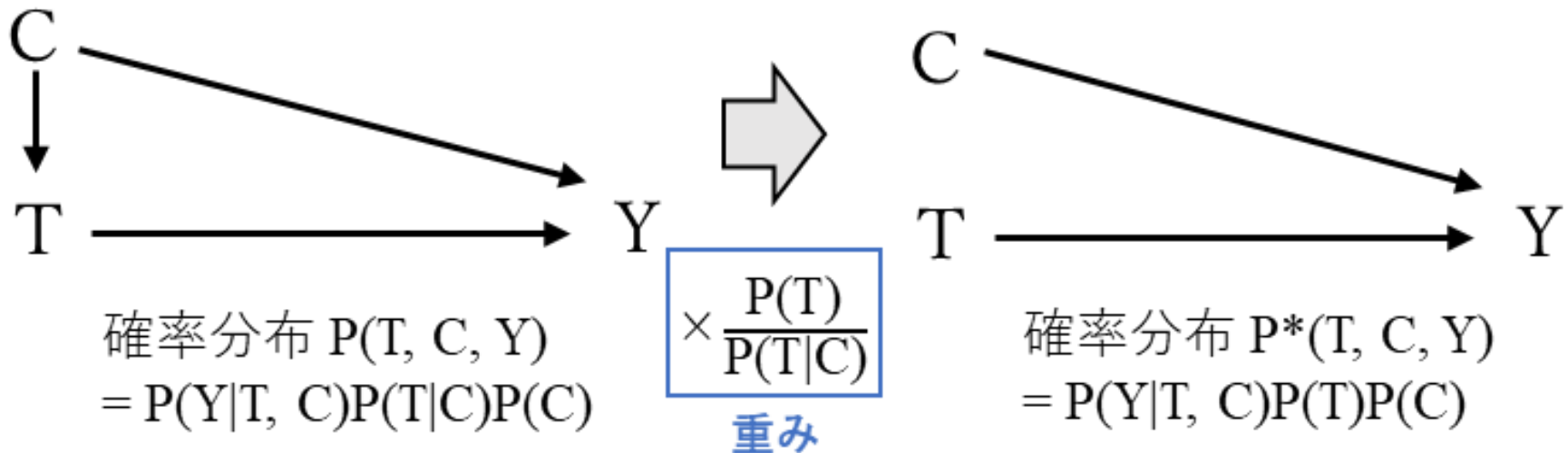
	Non-stabilized IPW	Stabilized IPW
数式	$1/P(T C)$	$P(T)/P(T C)$
Pseudo-population	Originalの2倍	Originalの1倍
IPWの平均	2.0	1.0
PSが小さいデータの 影響	大きい	
分散	大きくなりやすい	適切
Type I error	大きくなりやすい	適切

- Stabilized IPWは
Original sample sizeと同じ
PSが小さいデータの影響が少なめ
分散やType I errorが適切に

Inverse probability of treatment weight

観察されたデータ構造

逆確率重みづけにより変換したいデータ構造



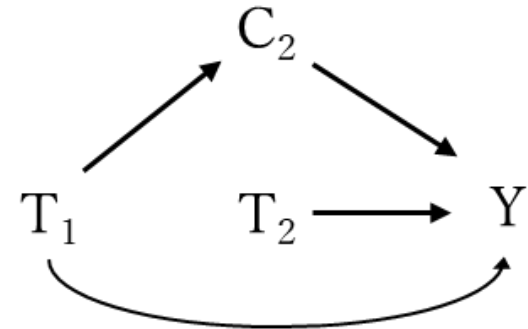
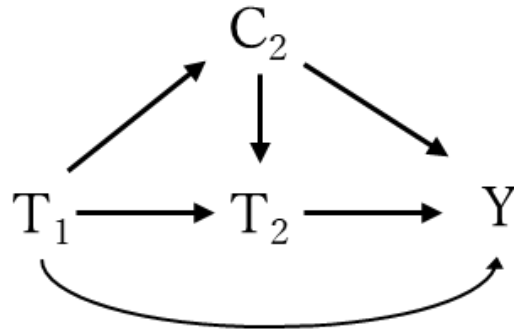
$$\frac{P(Y|T, C)P(T)P(C)}{P(Y|T, C)P(T|C)P(C)} = \frac{P(T)}{P(T|C)}$$

Inverse probability of treatment weight

観察されたデータ構造

逆確率重みづけにより変換したいデータ構造

DAG



確率分布

$$\begin{aligned}
 &P(Y, T_1, T_2, C_2) \\
 &= P(Y|T_1, T_2, C_2) \times P(T_2|T_1, C_2) \\
 &\quad \times P(C_2|T_1) \times P(T_1)
 \end{aligned}$$

$$\begin{aligned}
 &P(Y, T_1, T_2, C_2) \\
 &= P(Y|T_1, T_2, C_2) \times P(T_2) \\
 &\quad \times P(C_2|T_1) \times P(T_1)
 \end{aligned}$$

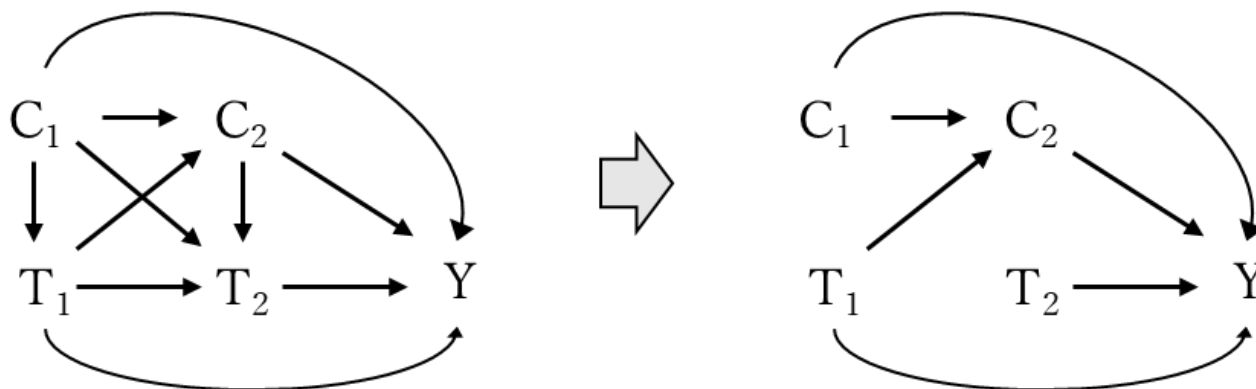
$$\frac{P(Y|T_1, T_2, C_2)P(T_2) P(C_2|T_1)P(T_1)}{P(Y|T_1, T_2, C_2)P(T_2|T_1, C_2) P(C_2|T_1)P(T_1)} = \frac{P(T_2)}{P(T_2|T_1, C_2)}$$

Inverse probability of treatment weight

観察されたデータ構造

逆確率重みづけにより変換したい
データ構造

DAG



$$\frac{P(Y|T_1, T_2, C_1, C_2)P(T_2) P(C_2|T_1, C_1)P(T_1)P(C_1)}{P(Y|T_1, T_2, C_1, C_2)P(T_2|T_1, C_1, C_2) P(C_2|T_1, C_1)P(T_1|C_1)P(C_1)}$$

$$= \frac{P(T_2)}{P(T_2|T_1, C_1, C_2)} \times \frac{P(T_1)}{P(T_1|C_1)}$$

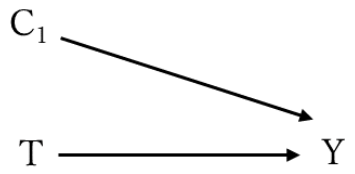
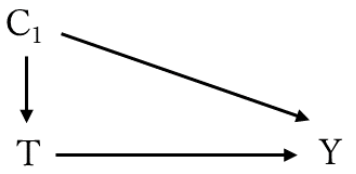
A

観察されたデータ構造

逆確率重みづけにより
変換したいデータ構造

逆確率重みづけ

DAG



$$\frac{P(Y|T, C_1)P(T)P(C_1)}{P(Y|T, C_1)P(T|C_1)P(C_1)}$$

$$= \frac{P(T)}{P(T|C_1)}$$

確率分布

$$P(Y, T, C_1)$$

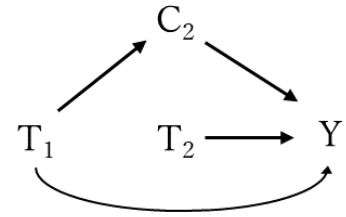
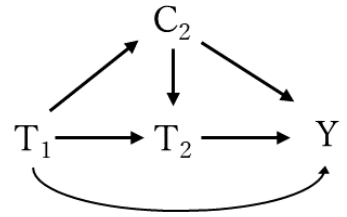
$$= P(Y|T, C_1) \times P(T|C_1) \times P(C_1)$$

$$P(Y, T, C_1)$$

$$= P(Y|T, C_1) \times P(T) \times P(C_1)$$

B

DAG



$$\frac{P(Y|T_1, T_2, C_2)P(T_2) P(C_2|T_1)P(T_1)}{P(Y|T_1, T_2, C_2)P(T_2|T_1, C_2) P(C_2|T_1)P(T_1)}$$

$$= \frac{P(T_2)}{P(T_2|T_1, C_2)}$$

確率分布

$$P(Y, T_1, T_2, C_2)$$

$$= P(Y|T_1, T_2, C_2) \times P(T_2|T_1, C_2)$$

$$\times P(C_2|T_1) \times P(T_1)$$

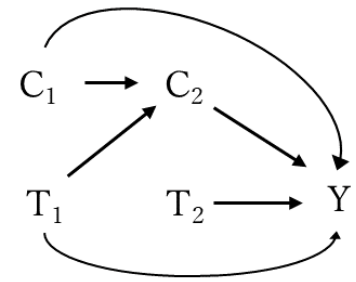
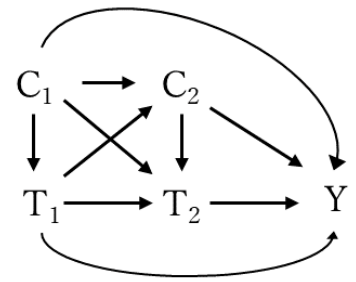
$$P(Y, T_1, T_2, C_2)$$

$$= P(Y|T_1, T_2, C_2) \times P(T_2)$$

$$\times P(C_2|T_1) \times P(T_1)$$

C

DAG



$$\frac{P(Y|T_1, T_2, C_1, C_2)P(T_2) P(C_2|T_1, C_1)P(T_1)P(C_1)}{P(Y|T_1, T_2, C_1, C_2)P(T_2|T_1, C_1, C_2) P(C_2|T_1, C_1)P(T_1|C_1)P(C_1)}$$

$$= \frac{P(T_2)}{P(T_2|T_1, C_1, C_2)} \times \frac{P(T_1)}{P(T_1|C_1)}$$

確率分布

$$P(Y, T_1, T_2, C_1, C_2)$$

$$= P(Y|T_1, T_2, C_1, C_2) \times P(T_2|T_1, C_1, C_2)$$

$$\times P(C_2|T_1, C_1) \times P(T_1|C_1) \times P(C_1)$$

$$P(Y, T_1, T_2, C_1, C_2)$$

$$= P(Y|T_1, T_2, C_1, C_2) \times P(T_2) \times P(C_2|T_1,$$

$$C_1) \times P(T_1) \times P(C_1)$$

Inverse probability of treatment weight

ステップ1. 交絡因子(C)で条件づけた際の曝露(T)の確率（傾向スコア）： $P(T|C)$ を計算する。

ステップ2. 曝露された対象者には $\frac{P(T=1)}{P(T=1|C)}$, 曝露されなかった対象者には $\frac{P(T=0)}{P(T=0|C)}$ を重みとして割り当てる。

ステップ3. アウトカムに対してTを用いた回帰モデル $\alpha + \beta_T T$ に、ステップ2で割り当てられた重みづけを行うことで、 β_T から因果効果が求められる。

Stabilized IPWの平均や分布を確認する

Stabilized IPWの平均や分布について

- 平均が1にならない
モデルの誤設定や欠損値の対処の問題
- 非常に高いIPWがある
 - モデルの誤設定の可能性：組みなおしてみる
 - non-positivityの可能性
 - 重みの打ち切り (weight truncation) を考慮:
例) 1 percentile to 99 percentileのみ使用

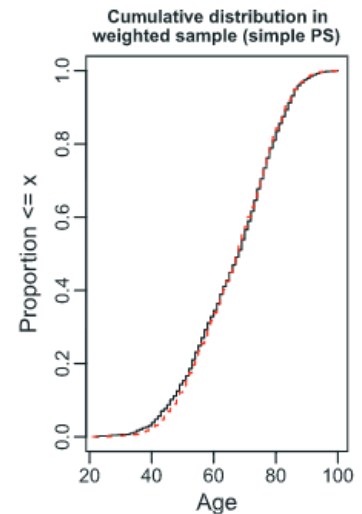
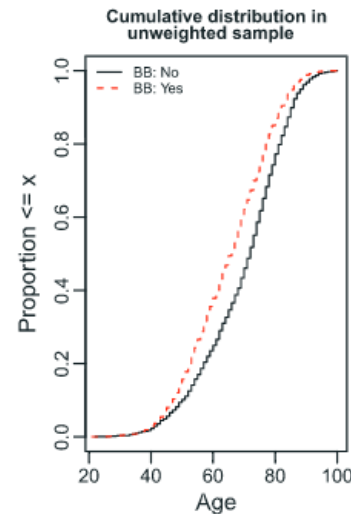
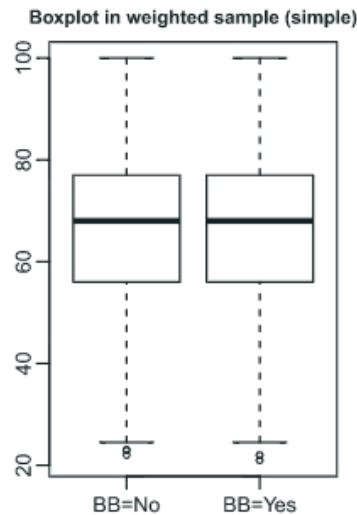
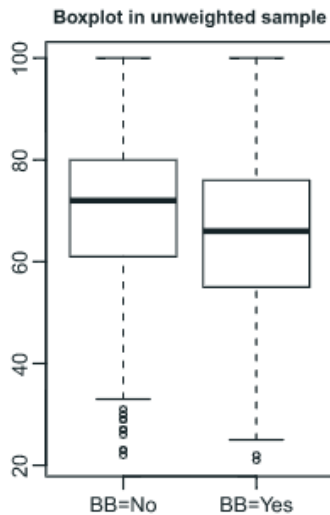
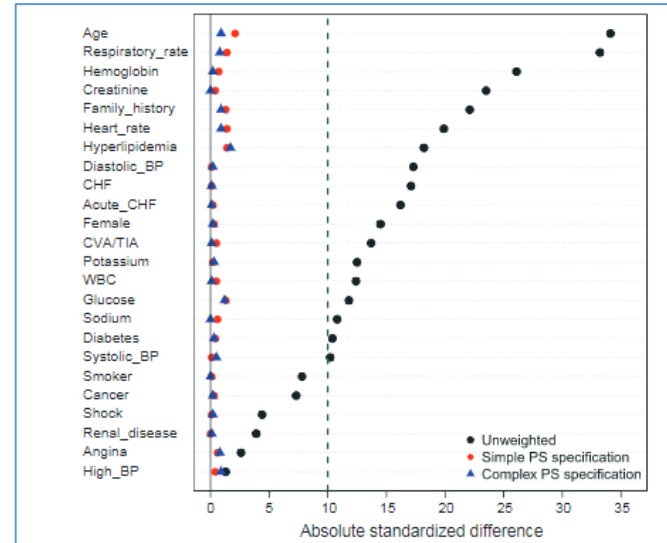
Weight Trimming and Propensity Score Weighting

Brian K. Lee^{1*}, Justin Lessler², Elizabeth A. Stuart^{3,4}

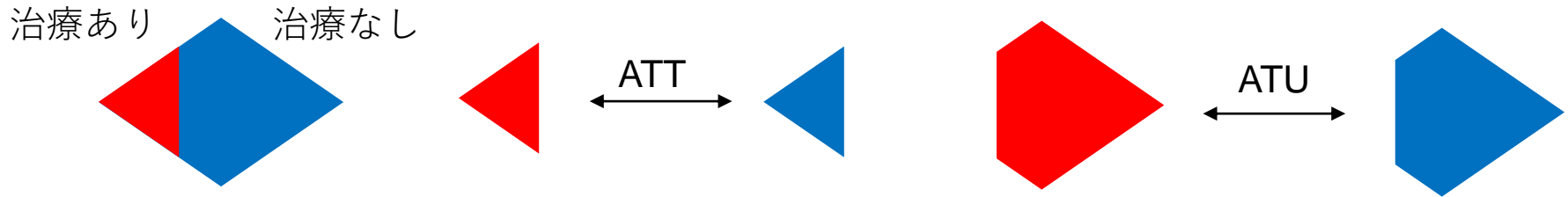
¹Department of Epidemiology and Biostatistics, Drexel University School of Public Health, Philadelphia, Pennsylvania, United States of America, ²Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States of America, ³Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States of America, ⁴Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States of America

グループ間のバランスをチェックする

- **Standardized difference**を使用
± 10%以内ならバランスOKとする
- グラフでチェック
Side-by-side Box plot
Cumulative distribution function



IPTW: ATTやATUの場合の重み付けは？

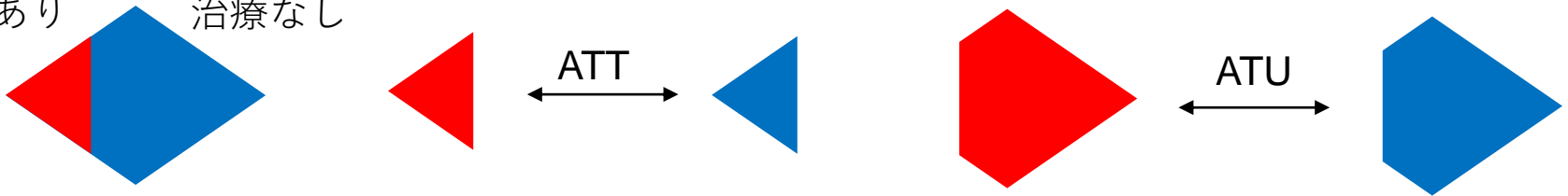


	T = 1 (Treated)	T = 0 (Untreated)
ATT	Weight = 1	$nsIPW = \frac{P(T C)}{1 - P(T C)}$ $sIPW = \frac{P(T C)}{1 - P(T C)} \bigg/ \frac{P(T)}{1 - P(T)}$
ATU	$nsIPW = \frac{1 - P(T C)}{P(T C)}$ $sIPW = \frac{1 - P(T C)}{P(T C)} \bigg/ \frac{1 - P(T)}{P(T)}$	Weight = 1

IPTW: ATTやATUの場合の重み付けは？

治療あり

治療なし



	T = 1 (Treated)	T = 0 (Untreated)
ATT	Weight = 1	$nsIPW = \frac{P(T C)}{1 - P(T C)}$ $sIPW = \frac{1 - P(T)}{1 - P(T C)} \bigg/ \frac{P(T)}{P(T C)}$
ATU	$nsIPW = \frac{1 - P(T C)}{P(T C)}$ $sIPW = \frac{P(T)}{P(T C)} \bigg/ \frac{1 - P(T)}{1 - P(T C)}$	Weight = 1

本日のアジェンダ

- Recap (DAG, Probability)
- 時間変動する曝露の効果を推定する
 - Inverse probability of treatment weight
 - **G-computation**
- 実例紹介 (Sustained Treatment, Different time point, Subsequent symptoms)
- + α : 時間変動するリスクを推定する

G-computation

$$P(Y^T) = \sum_c P(Y^T | C) P(C)$$

$$= \sum_c P(Y|T, C) P(C) \quad \text{G-formula: Empirical Analogue}$$

$$= \sum_c \frac{P(Y, T, C)}{P(T, C)} P(C)$$

$$= \sum_c \frac{P(Y, T, C)}{P(T|C)P(C)} P(C)$$

$$= \sum_c P(Y, T, C) \times \frac{1}{P(T|C)} \quad \text{Inverse probability weight}$$

G-computation

集団全員が曝露した（何らかの介入を受けた）場合としなかった（介入を受けなかった）場合のアウトカムを比較することで平均因果効果を推定する。

①もし集団全員がアスピリン投与を受けていたら

事実	仮想的介入	冠動脈疾患既往あり		冠動脈疾患既往なし	
		冠動脈疾患発生	合計	冠動脈疾患発生	合計
アスピリン投与	アスピリン投与 (実際と同じ)	30	300	2	200
アスピリン非投与	アスピリン投与 (反事実)	? [20]	200	? [23]	2300
合計		30 + [20]	500	2 + [23]	2500

反事実リスク = $\frac{75}{3000} = 0.025$

②もし集団全員がアスピリン投与を受けていなかったら

事実	仮想的介入	冠動脈疾患既往あり		冠動脈疾患既往なし	
		冠動脈疾患発生	合計	冠動脈疾患発生	合計
アスピリン投与	アスピリン非投与 (反事実)	? [90]	300	? [6]	200
アスピリン非投与	アスピリン非投与 (実際と同じ)	60	200	69	2300
合計		[90] + 60	500	[6] + 69	2000

反事実リスク = $\frac{225}{3000} = 0.075$

設定

アスピリン投与による冠動脈疾患発生の減少効果を求める。

(※簡潔な説明のため、交絡は冠動脈疾患既往のみとする)

因果リスク差

$$= 0.025 - 0.075$$

$$= -0.050$$

因果リスク比

$$= 0.025 / 0.075$$

$$= 0.333$$

井上、杉山、後藤
医学界新聞 臨床研究・
疫学研究のための因果推
論レクチャー第4回

G-computation

実際のデータは多くの場合層別しきれない！

時点が増えたら...

交絡変数の次元が高くなったら...

Parametric g-formula

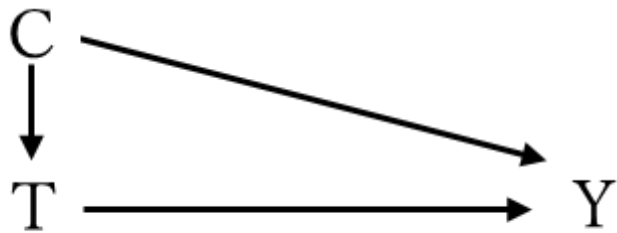
$$\begin{aligned} & E(Y^{T_1 T_2}) \\ &= \sum_{c_1} \sum_{c_2} E(Y|T_1 = t_1, T_2 = t_2, C_1 = c_1, C_2 = c_2) \\ & \quad \times P(C_2 = c_2|T_1 = t_1, C_1 = c_1)P(C_1 = c_1) \end{aligned}$$

C_i ごとに予測値を出して平均すればよい！

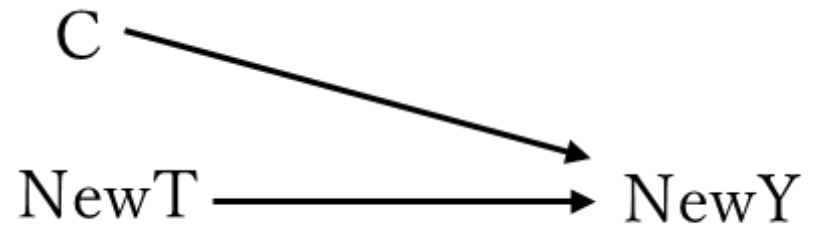
回帰標準化と同じ

G-computation

Yに対する予測モデルの構築



予測モデルを用いて、新しく割り当てたNewTに対するNewYを予測する



G-computation

ステップ1. 曝露因子(T)、交絡因子(C)を用いてアウトカム(Y)の予測モデルを構築する。

ステップ2. データセット(サンプル数=N)のコピーを二つ用意し、他の共変量は変えずに、一つのデータは全員が曝露された状況(NewT=1)、もう一つのデータは全員が曝露されなかった状況(NewT=0)を作る。

ステップ3. 二つのコピーを合わせたデータ(サンプル数=2N)に対して、ステップ2で割り当てられた曝露状況と、対象者それぞれが有している交絡因子の情報を、ステップ1で構築されたアウトカムモデルに当てはめることで、「全員が曝露された場合」、「全員が曝露されなかった場合」に予測されるアウトカムを算出する。

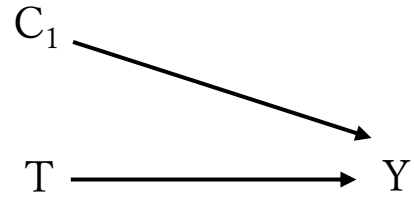
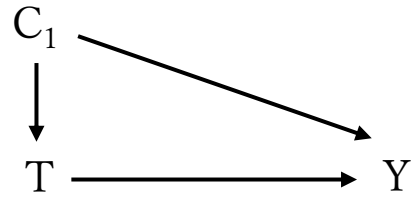
ステップ4. ステップ3で算出されたアウトカムの予測値に対してNewTを用いた回帰モデル $\alpha + \beta_{NewT} NewT$ を構築することで、 β_{NewT} から因果効果が求められる。

A

観察されたデータ構造

G-computationにより変換
したいデータ構造

DAG



ID	T	C ₁	Y
1	0	0	1
2	1	0	1
3	0	1	0
4	1	1	0

予測モデル構築

ID	T	C ₁	Y
1	0	0	1
2	0	0	?
3	0	1	0
4	0	1	?

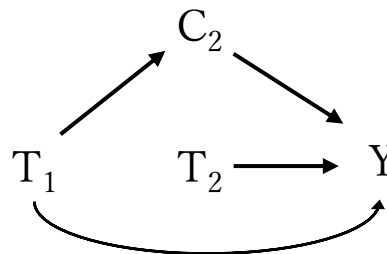
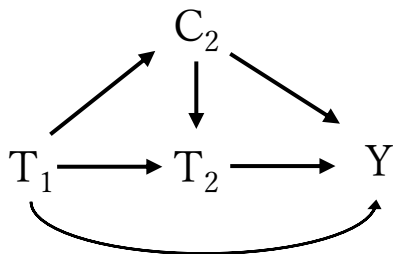
ID	T	C ₁	Y
1	1	0	?
2	1	0	1
3	1	1	?
4	1	1	0

B

観察されたデータ構造

G-computationにより変換
したいデータ構造

DAG



ID	T ₁	T ₂	C ₂	Y
1	0	1	0	1
2	1	1	0	0
3	0	0	1	0
4	1	0	1	0

ID	T ₁	T ₂	C ₂	Y
1	1	0	?	?
2	1	0	?	?
3	1	0	?	?
4	1	0	1	0

ID	T ₁	T ₂	C ₂	Y
1	1	1	?	?
2	1	1	0	0
3	1	1	?	?
4	1	1	?	?

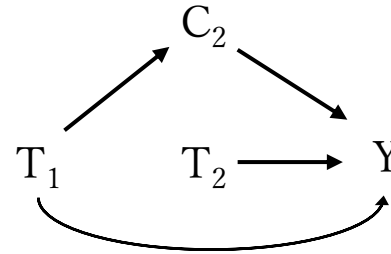
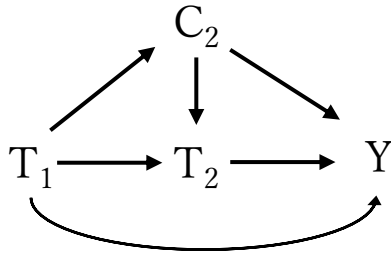
- ①C₂の予測モデル構築
- ②Yの予測モデル構築

B

観察されたデータ構造

G-computationにより変換
したいデータ構造

DAG



ID	T ₁	T ₂	C ₂	Y
1	0	1	0	1
2	1	1	0	0
3	0	0	1	0
4	1	0	1	0

ID	T ₁	T ₂	C ₂	Y
1	0	0	?	?
2	0	0	?	?
3	0	0	1	0
4	0	0	?	?

ID	T ₁	T ₂	C ₂	Y
1	0	1	0	1
2	0	1	?	?
3	0	1	?	?
4	0	1	?	?

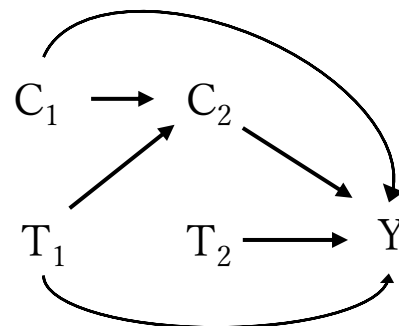
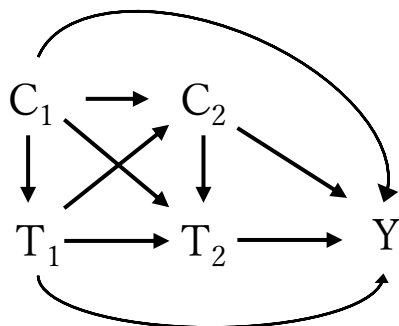
- ①C₂の予測モデル構築
- ②Yの予測モデル構築

C

観察されたデータ構造

G-computationにより変換
したいデータ構造

DAG



ID	T ₁	T ₂	C ₁	C ₂	Y
1	0	1	0	0	1
2	1	1	0	0	0
3	0	0	1	0	0
4	1	0	1	1	0

ID	T ₁	T ₂	C ₁	C ₂	Y
1	1	0	0	?	?
2	1	0	0	?	?
3	1	0	1	?	?
4	1	0	1	1	0

ID	T ₁	T ₂	C ₁	C ₂	Y
1	1	1	0	?	?
2	1	1	0	0	0
3	1	1	1	?	?
4	1	1	1	?	?

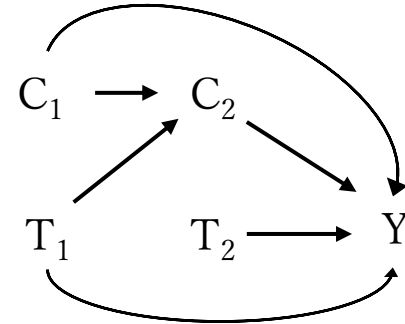
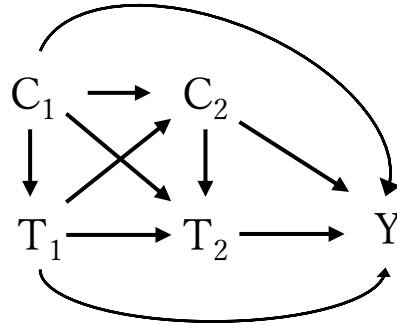
- ① C₂の予測モデル構築
- ② Yの予測モデル構築

C

観察されたデータ構造

G-computationにより変換
したいデータ構造

DAG



ID	T ₁	T ₂	C ₁	C ₂	Y
1	0	1	0	0	1
2	1	1	0	0	0
3	0	0	1	0	0
4	1	0	1	1	0

ID	T ₁	T ₂	C ₁	C ₂	Y
1	0	0	0	?	?
2	0	0	0	?	?
3	0	0	1	0	0
4	0	0	1	?	?

ID	T ₁	T ₂	C ₁	C ₂	Y
1	0	1	0	0	1
2	0	1	0	?	?
3	0	1	1	?	?
4	0	1	1	?	?

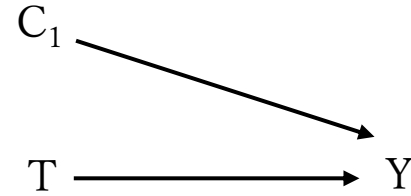
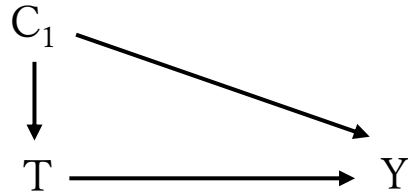
- ① C₂の予測モデル構築
- ② Yの予測モデル構築

A

観察されたデータ構造

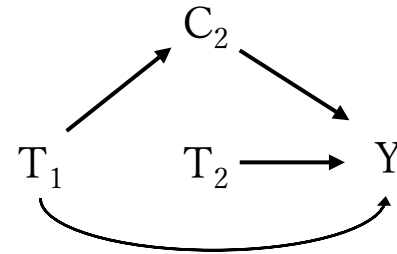
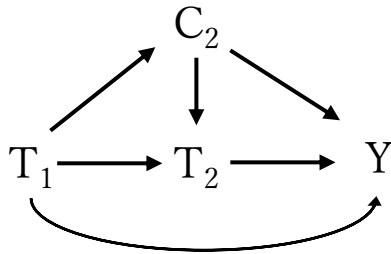
G-computationにより変換
したいデータ構造

DAG



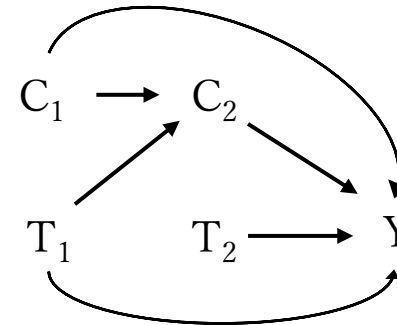
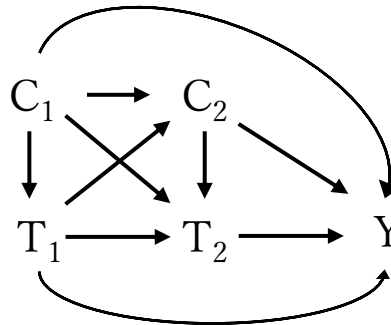
B

DAG



C

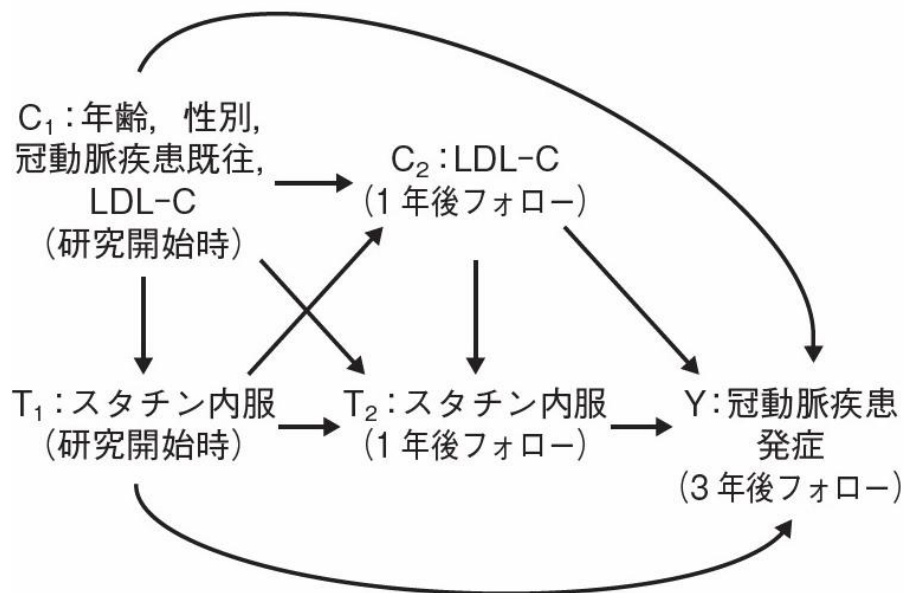
DAG



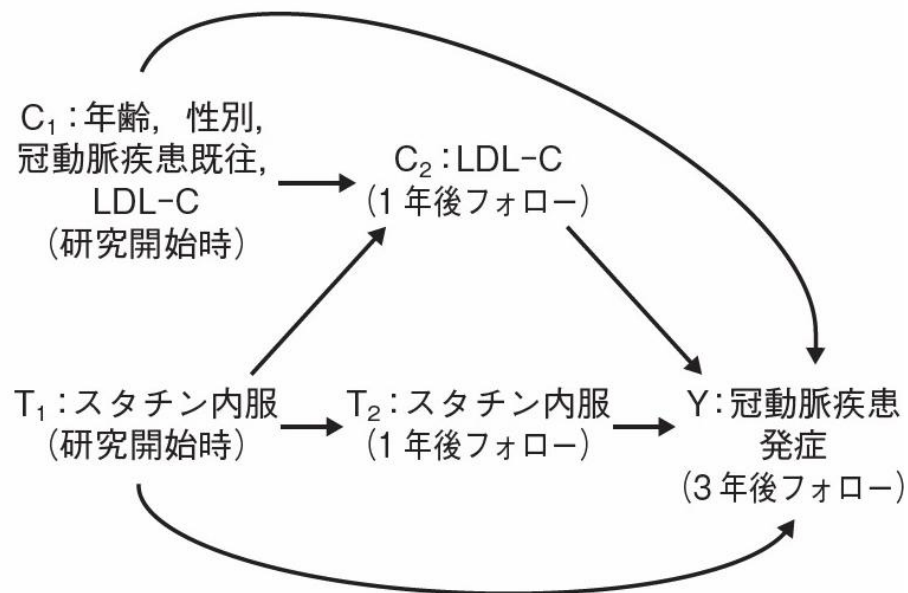
Q1 : IPTWは？

Q2 : G-computationを用いる場合はどの予測モデルを作る？

A 観察されたデータ構造



B T₁, T₂ が C₁, C₂ から影響を受けない場合



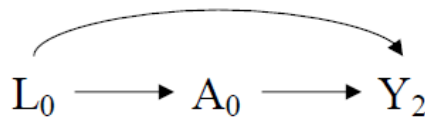
$$\frac{P(T_1)}{P(T_1 | C_1)} \times \frac{P(T_2 | T_1)}{P(T_2 | T_1, C_1, C_2)}$$

本日のアジェンダ

- Recap (DAG, Probability)
- 時間変動する曝露の効果を推定する
 - Inverse probability of treatment weight
 - G-computation
- 実例紹介 (Sustained Treatment, Different time point, Subsequent symptoms)
- $+\alpha$: 時間変動するリスクを推定する

Estimating the impact of sustained social participation on depressive symptoms in older adults

(A) Analysis of social participation at baseline



Causal Contrast of Interest

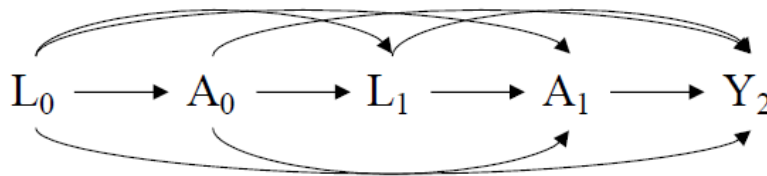
1. Prevalence Difference

$$\Pr[Y_2^{a_0=1} = 1] - \Pr[Y_2^{a_0=0} = 1]$$

2. Prevalence Ratio

$$\frac{\Pr[Y_2^{a_0=1} = 1]}{\Pr[Y_2^{a_0=0} = 1]}$$

(B) Analysis of time-varying social participation from two time points



Causal Contrast of Interest

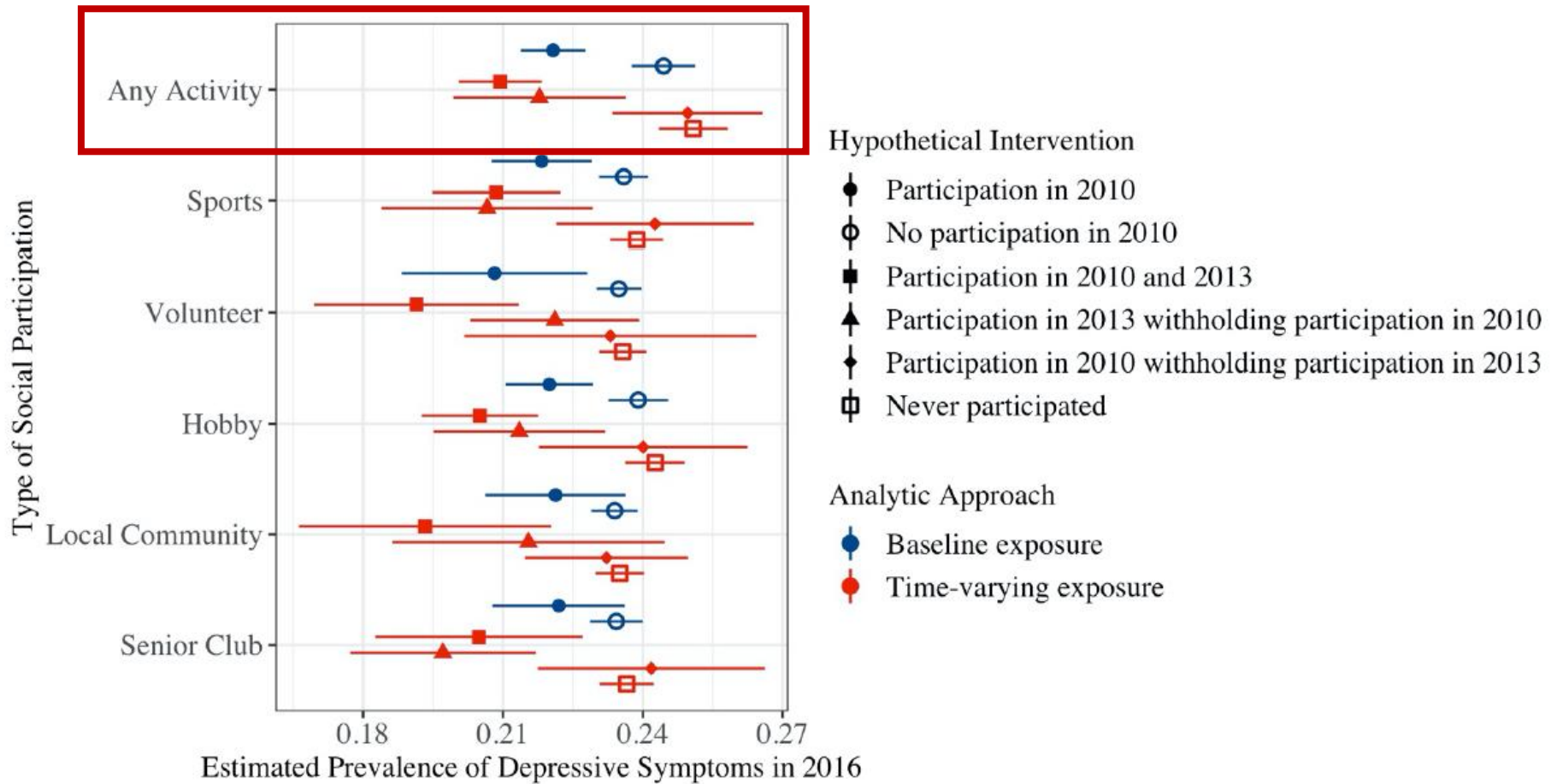
1. Prevalence Difference

$$\Pr[Y_2^{a_0=1, a_1=1} = 1] - \Pr[Y_2^{a_0=0, a_1=0} = 1]$$

2. Prevalence Ratio

$$\frac{\Pr[Y_2^{a_0=1, a_1=1} = 1]}{\Pr[Y_2^{a_0=0, a_1=0} = 1]}$$

Sustained social participation may be a key for long-term prevention of depressive symptoms among older adults.



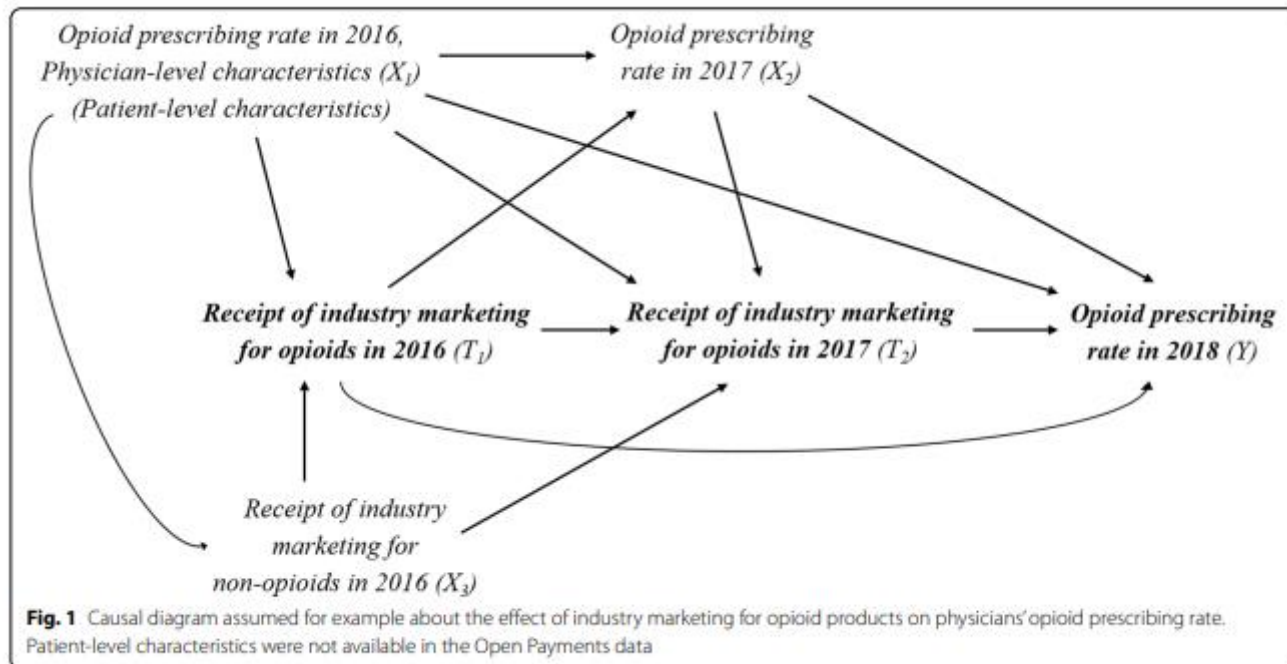
RESEARCH

Open Access

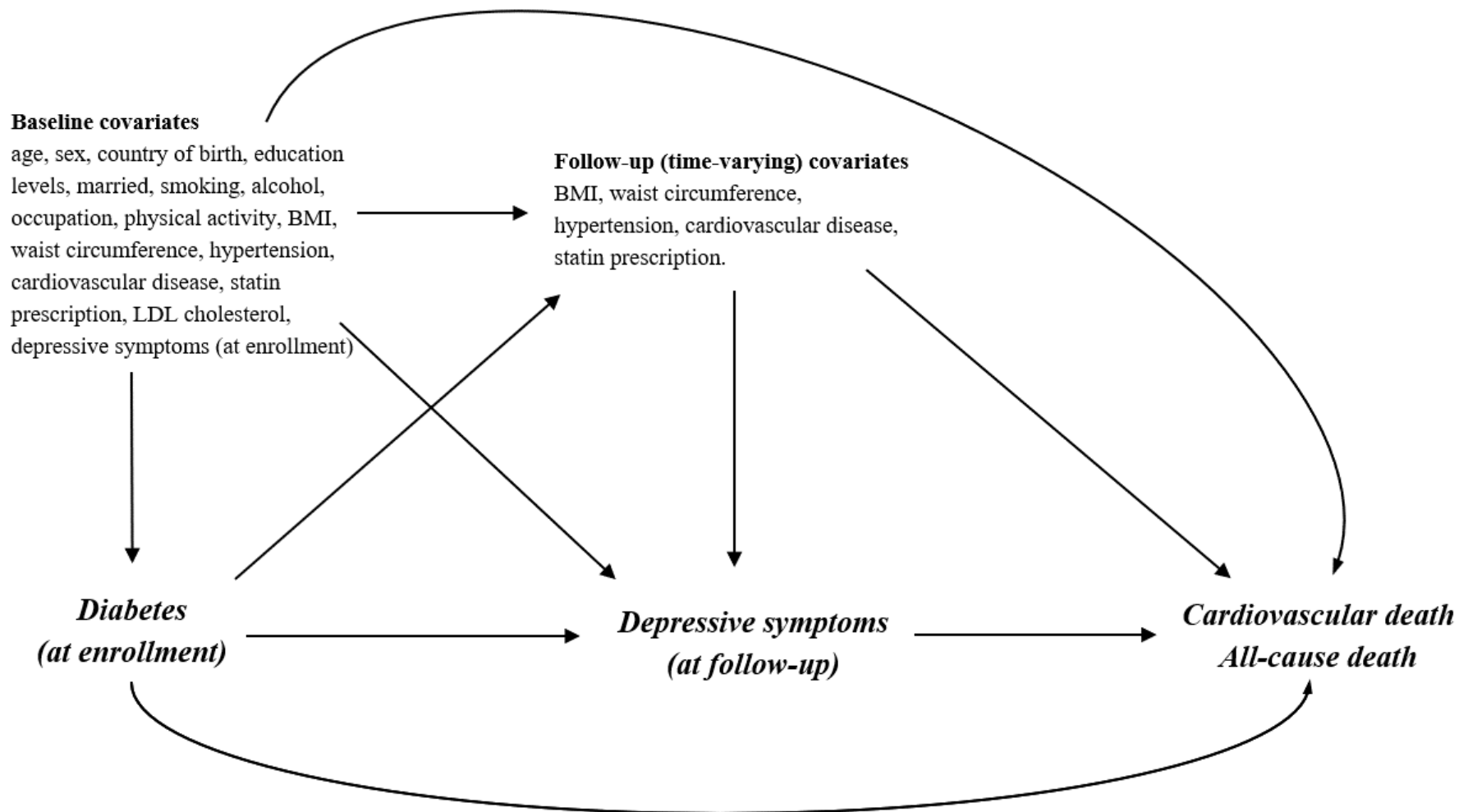


Bias amplification in the g-computation algorithm for time-varying treatments: a case study of industry payments and prescription of opioid products

Kosuke Inoue^{1,2*}, Atsushi Goto³, Naoki Kondo^{1,4} and Tomohiro Shinozaki⁵



Estimating the joint effect of diabetes and subsequent depressive symptoms on mortality among older Latinos



Introduction



- Depression is a well-known factor closely associated with diabetes, CVD, and mortality.

Lancet Diabetes Endocrinol. 2015;3(6):461-471
Nat Rev Cardiol. 2017;14(3):145-155

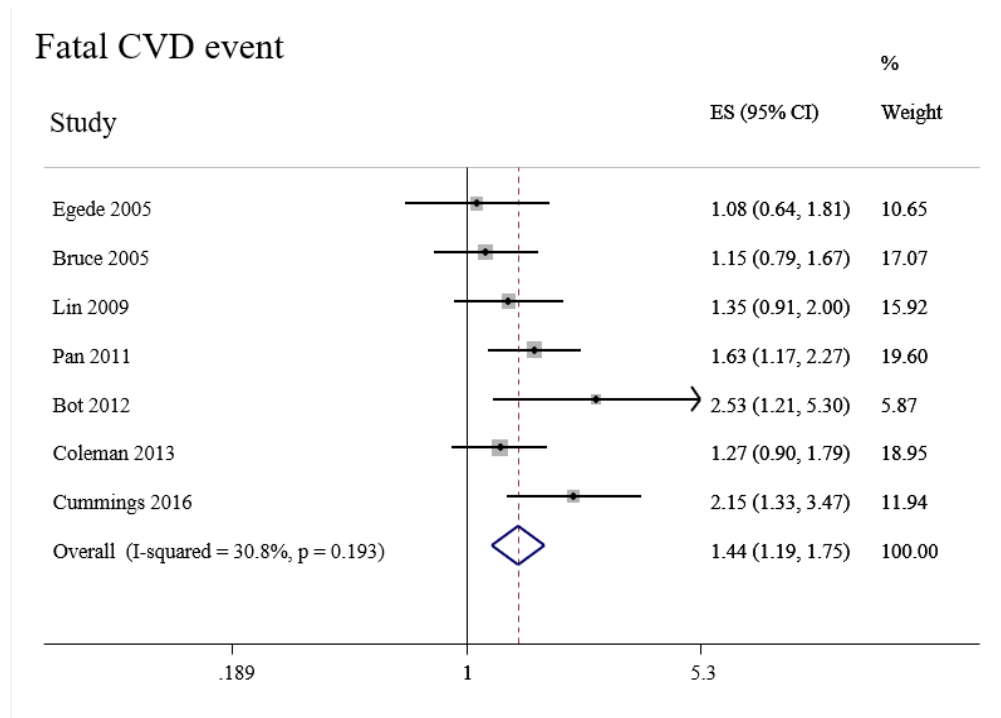
- The prevalence of depression in people with type 2 diabetes is **almost double** compared to people without diabetes.

Diabetes Care. 2001;24(6):1069-1078

- *Psychosocial stress is **the third contributing factor*** to the attributable risk of acute myocardial infarction (approximately 30%) after lipids and smoking.

Lancet. 2004;364(9438):937-952

- Our recent meta-analysis showed the association between depression and cardiovascular mortality [RR=1.44]



J Diabetes Complications . 2020;34(12):107710.

- However, all studies did not clarify the temporal ordering of diabetes and depression

To estimate the joint effect of diabetes and subsequent depressive symptoms on cardiovascular and all-cause mortality using marginal structural models.

Study population & Measurement

- The Sacramento Area Latino Study on Aging (SALSA)
- **1495 participants** had data on diabetes and potential confounders at baseline.
- Exposure: Diabetes
- Mediator: **Depressive symptoms** defined by The Center for Epidemiologic Studies Depression Scale (CESD) ≥ 16 or taking antidepressant.
- Outcomes: All-cause mortality, cardiovascular mortality

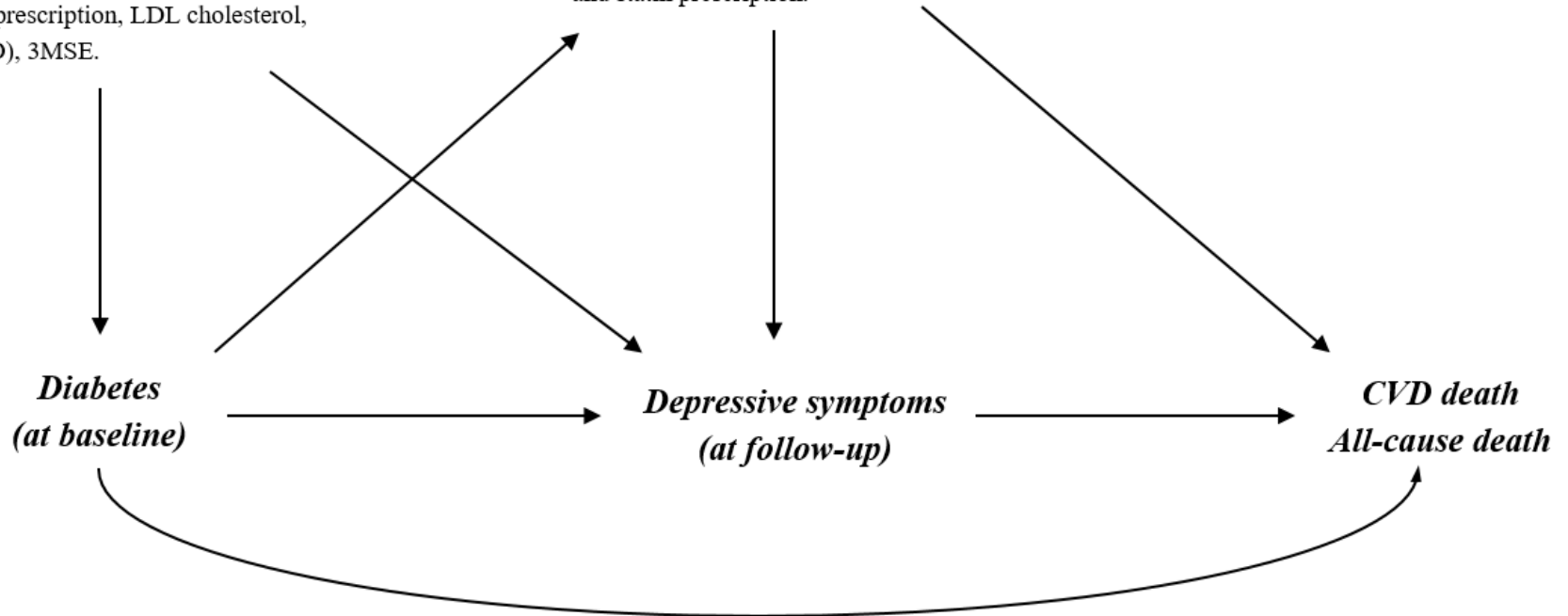
Causal Diagram

Baseline covariates

age, sex, country of birth, education levels, married, smoking, alcohol, occupation, physical activity, BMI, waist circumference, hypertension, statin prescription, LDL cholesterol, (CESD), 3MSE.

Follow-up (time-varying) covariates

BMI, waist circumference, hypertension, cardiovascular disease, and statin prescription.



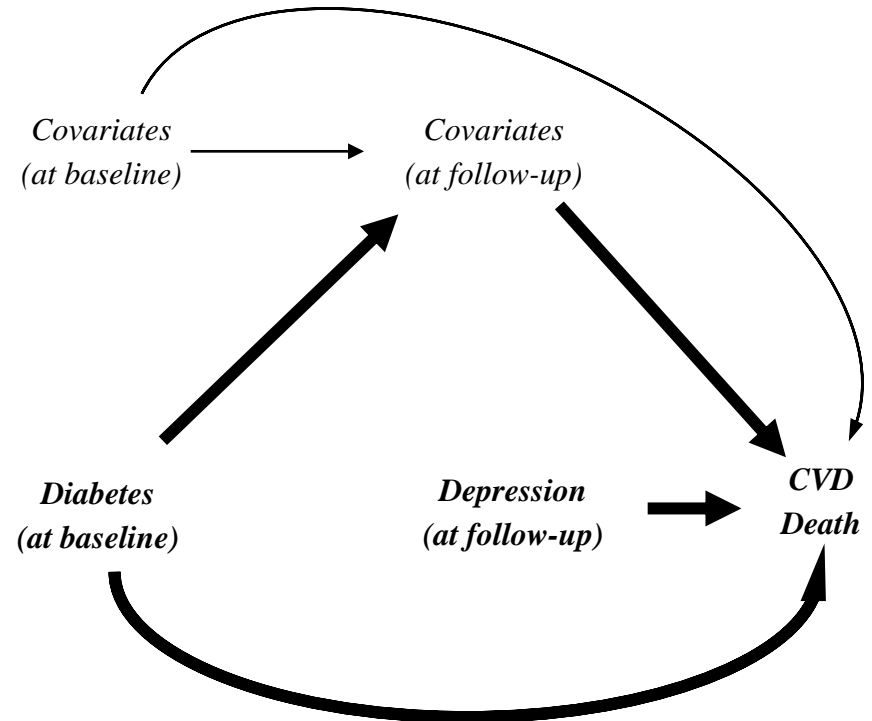
Statistical analysis

- **Marginal structural model (MSM)** approach to estimate the joint effects of diabetes and depressive symptoms on cardiovascular and all-cause mortality.
- Cox proportional hazards models were applied using the final weight and adjusting for potential confounders.

Marginal Structural Model

Inverse probability of treatment weighting (IPTW)

$$= \frac{\frac{pr[DM_1]}{pr[DM_1|Cov_1]} \times pr[Dep_2]}{pr[Dep_2|DM_1,Cov_1,Cov_2]}$$



Robust 95% confidence intervals were estimated by 1,000 bootstrapped samples.

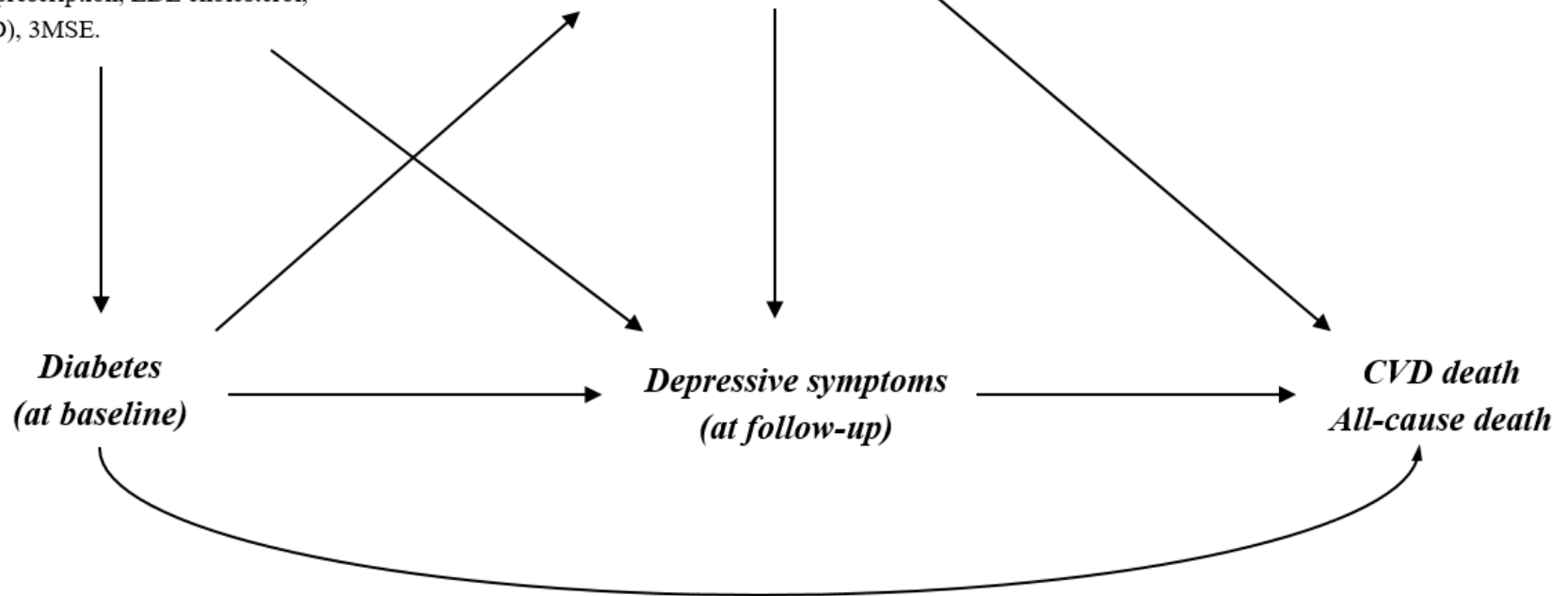
Q: G-computationを用いる場合はどの予測モデルを作る？

Baseline covariates

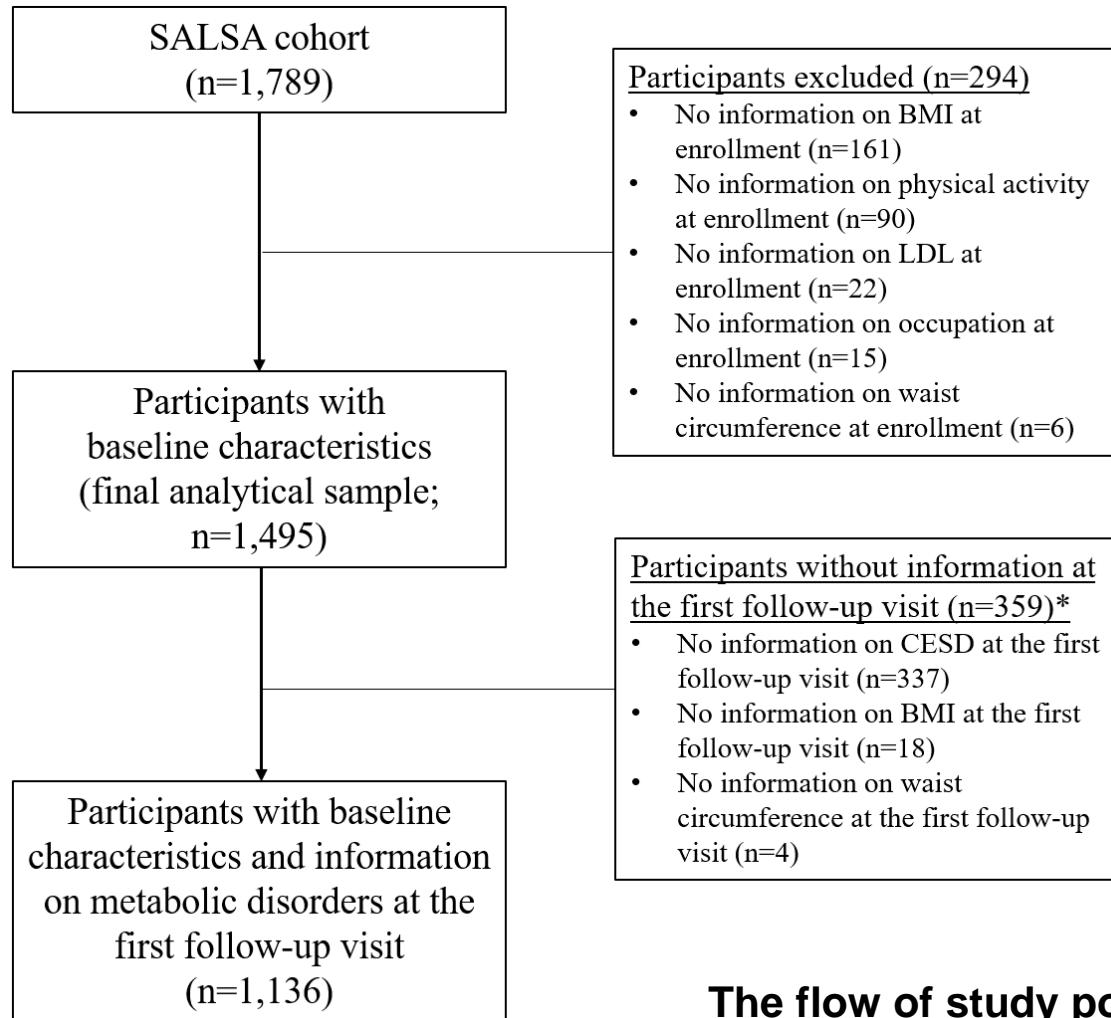
age, sex, country of birth, education levels, married, smoking, alcohol, occupation, physical activity, BMI, waist circumference, hypertension, statin prescription, LDL cholesterol, (CESD), 3MSE.

Follow-up (time-varying) covariates

BMI, waist circumference, hypertension, cardiovascular disease, and statin prescription.



Results



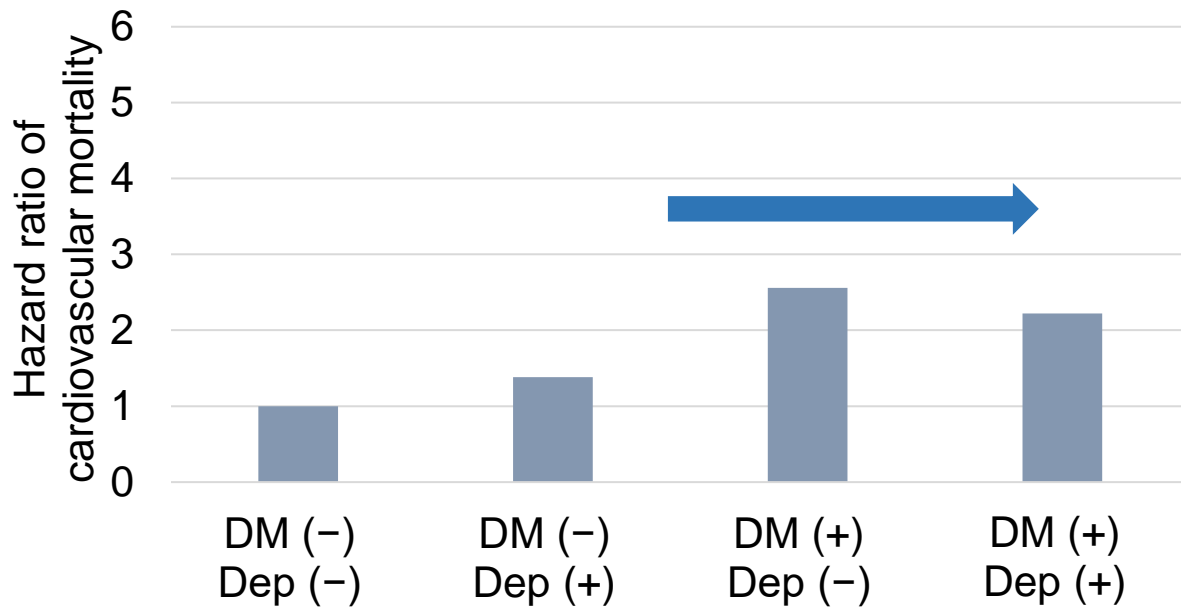
The flow of study population

We did not find the interaction between diabetes (at baseline**) and depressive symptoms (**at baseline**) for mortality.**

Outcomes		Cardiovascular mortality		All-cause mortality	
Diabetes at enrollment	Elevated depressive symptoms at enrollment	Number of Events	Adjusted HR (95% CI)	Number of Events	Adjusted HR (95% CI)
No	No	62/724	Ref	111/724	Ref
Yes	No	78/318	2.56 (1.75 to 3.65)	106/318	2.12 (1.59 to 2.87)
No	Yes	40/283	1.38 (0.86 to 2.18)	67/283	1.29 (0.94 to 1.79)
Yes	Yes	38/170	2.22 (1.39 to 2.18)	57/170	2.12 (1.49 to 3.16)
HR for the interaction term (multiplicative scale)			0.64 (0.35 to 1.18)		0.78 (0.48 to 1.27)
RERI (additive scale)			-0.68 (-2.00 to 0.48)		-0.26 (-1.23 to 0.59)

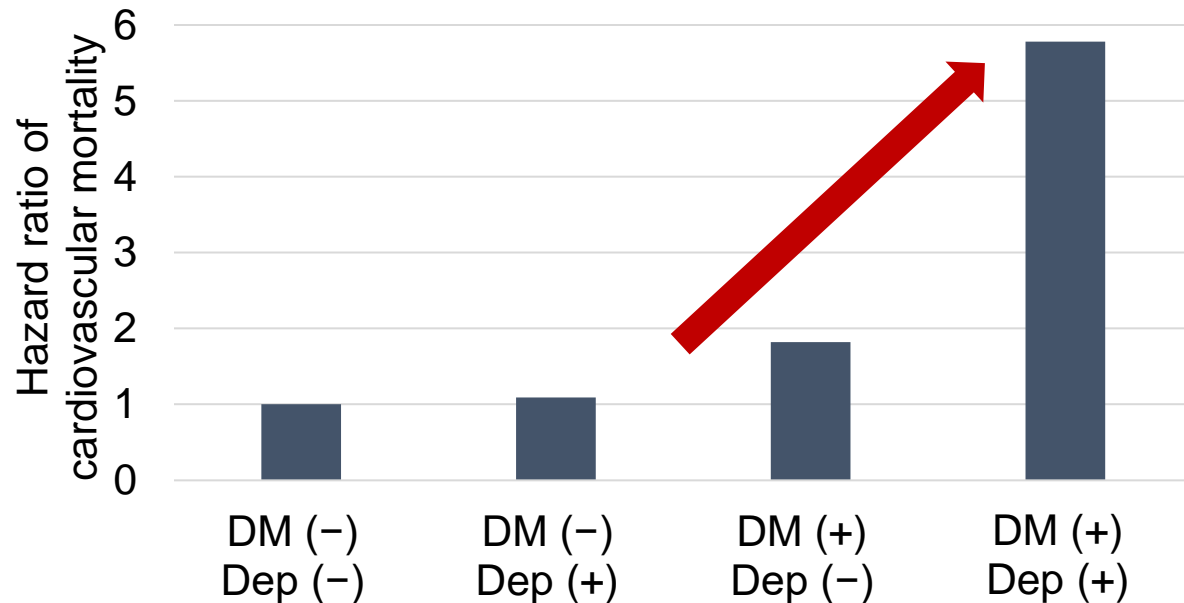
Diabetes (at baseline) and subsequent depressive symptoms (at follow-up) were jointly associated with CVD mortality.

Outcomes		Cardiovascular mortality		All-cause mortality	
Diabetes at enrollment	Elevated depressive symptoms at follow-up	Number of Events	Adjusted HR (95% CI)	Number of Events	Adjusted HR (95% CI)
No	No	50/601	Ref	86/601	Ref
Yes	No	43/249	1.82 (1.12 to 3.02)	64/249	2.10 (1.36 to 3.30)
No	Yes	21/179	1.09 (0.46 to 2.17)	38/179	1.13 (0.65 to 1.88)
Yes	Yes	34/107	5.78 (3.02 to 11.97)	43/107	4.32 (2.41 to 7.31)
HR for the interaction term (multiplicative scale)			2.94 (1.07 to 8.39)		1.80 (0.81 to 4.35)
RERI (additive scale)			3.79 (1.05 to 9.81)		2.02 (0.01 to 5.08)

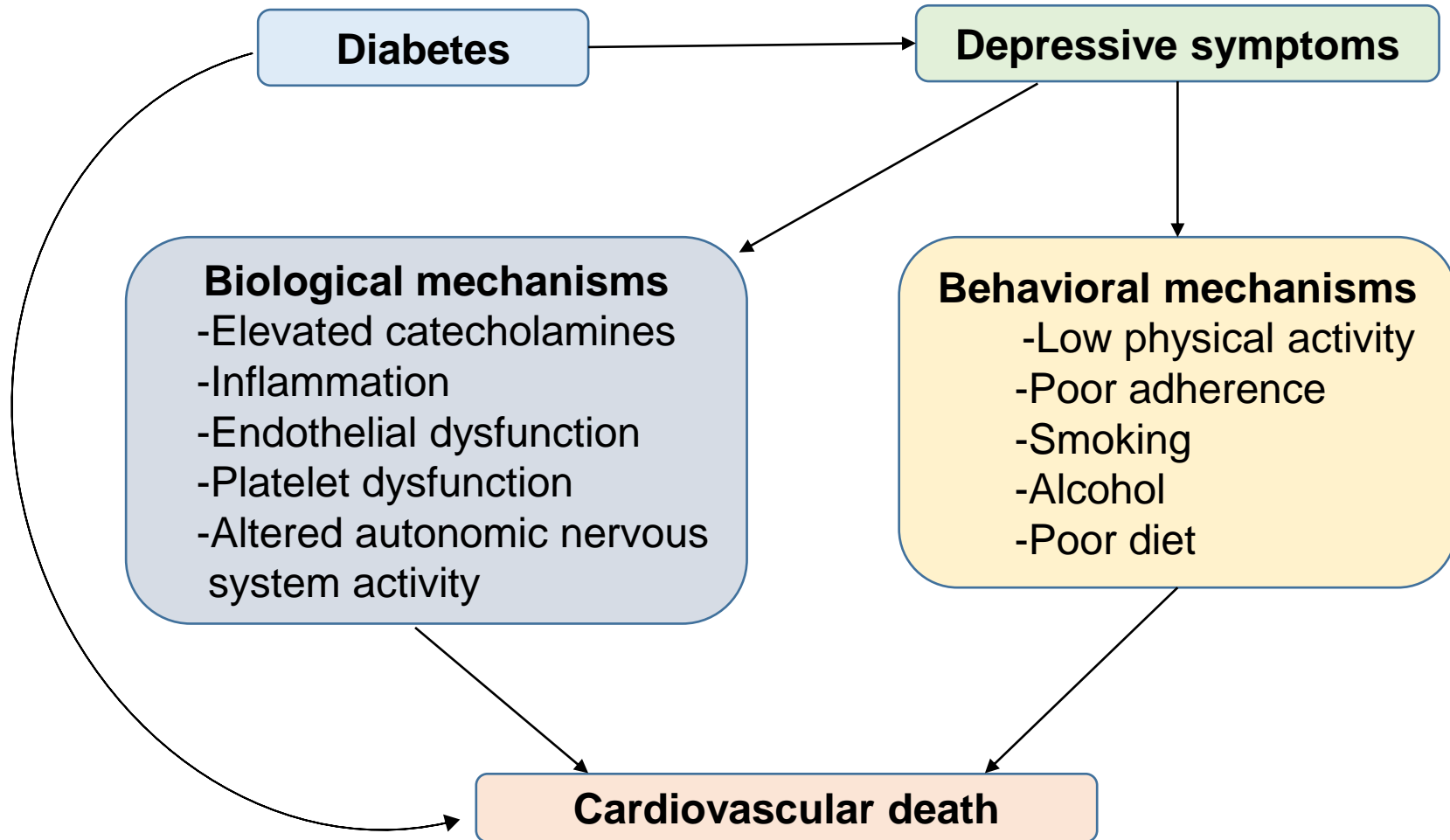


Diabetes **at baseline** and Depressive symptoms **at baseline without** adjusting for time-varying confounders

Diabetes **at baseline** and Depressive symptoms **at follow-up with** adjusting for time-varying confounders



Possible mechanisms



Nat Rev Cardiol **14**, 145–155 (2017).

Summary

- After controlling for time-varying confounders, diabetes and subsequent depressive symptoms **were jointly associated with cardiovascular mortality.**
- Mental health management after a diagnosis of diabetes may be critical to promote cardiovascular health among older Mexican Americans.

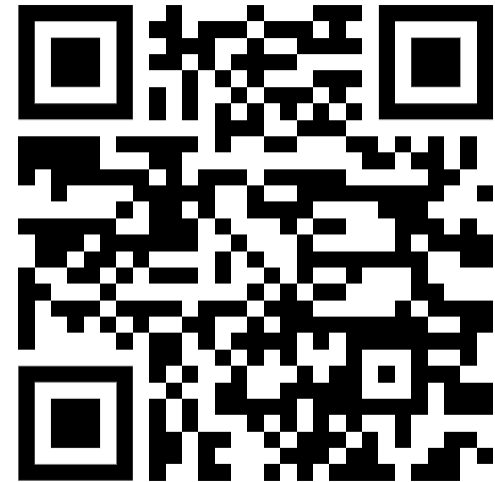
本日のアジェンダ

- Recap (DAG, Probability)
- 時間変動する曝露の効果を推定する
 - Inverse probability of treatment weight
 - G-computation
- 実例紹介 (Sustained Treatment, Different time point, Subsequent symptoms)
- $+\alpha$: 時間変動するリスクを推定する

Low HbA1c levels and all-cause or cardiovascular mortality among US adults





International Journal of Epidemiology, 2021, 1373–1383
doi: 10.1093/ije/dyaa263
Advance Access Publication Date: 30 December 2020
Original article

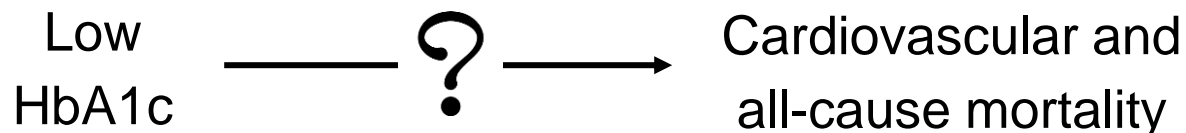


Miscellaneous

Low HbA1c levels and all-cause or cardiovascular mortality among people without diabetes: the US National Health and Nutrition Examination Survey 1999–2015

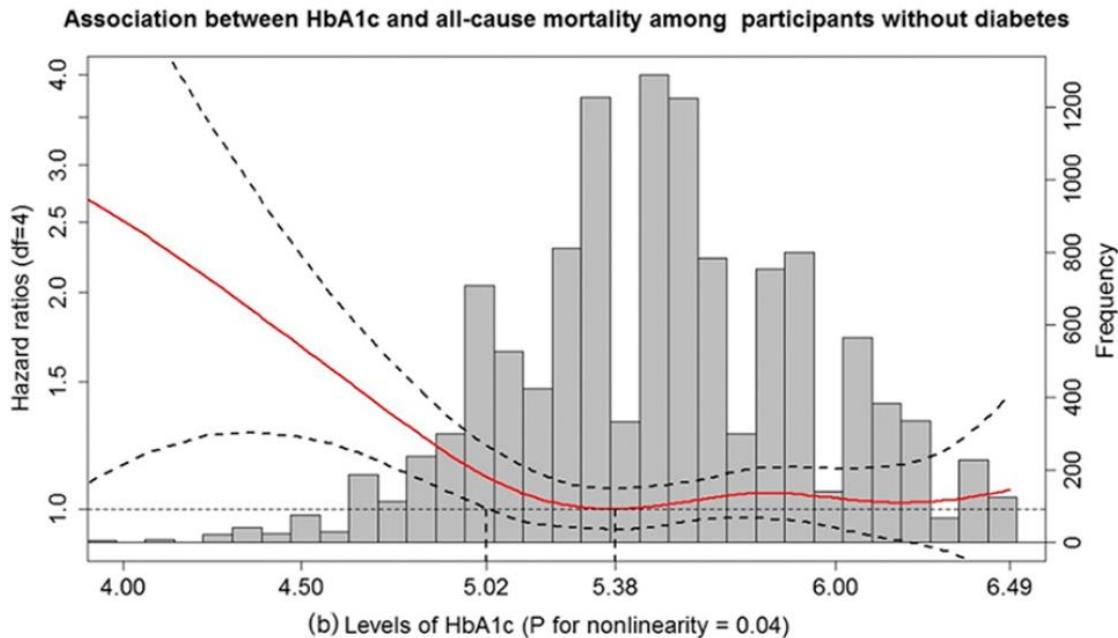
Kosuke Inoue,^{1*} Roch Nianogo,¹ Donatello Telesca,² Atsushi Goto ,³
Vahe Khachadourian,⁴ Yusuke Tsugawa,^{5,6} Takehiro Sugiyama,^{7,8}
Elizabeth Rose Mayeda ¹ and Beate Ritz^{1,9,10}

Int J Epidemiol . 2021;dyaa263.



Introduction

- Glycated hemoglobin (HbA1c) is one of the biomarkers to diagnose and manage diabetes.
- Previous studies showed the association between **low HbA1c and the increased risk of death.**



J Clin Endocrinol Metab.
2019;104(8):3345-3354

- **Several factors** (e.g., socioeconomic status, diet, exercise, biomarker, comorbidities, and medication) **are associated with both HbA1c and mortality.**

Lancet. 2016;387(10027):1513-1530.

- It is often challenging to **integrate all these dimensions and accurately infer the causality** among these factors, HbA1c, and adverse health outcomes, using traditional statistical methods.

To elucidate the causal relationship between low HbA1c and all-cause or cardiovascular mortality using machine learning within the causal structural framework.

National Health and Nutrition Examination Survey (NHANES)

- A large-scale, multistage, **nationally representative** survey of the civilian noninstitutionalized population in the US.
- Structured interview data and physical examination results are released in **two-year cycles**.
- This study included **39,453 participants** aged ≥ 20 years in the NHANES 1999-2014 **linked to mortality data** in 2015.

Measurement of variables

➤ Exposure:

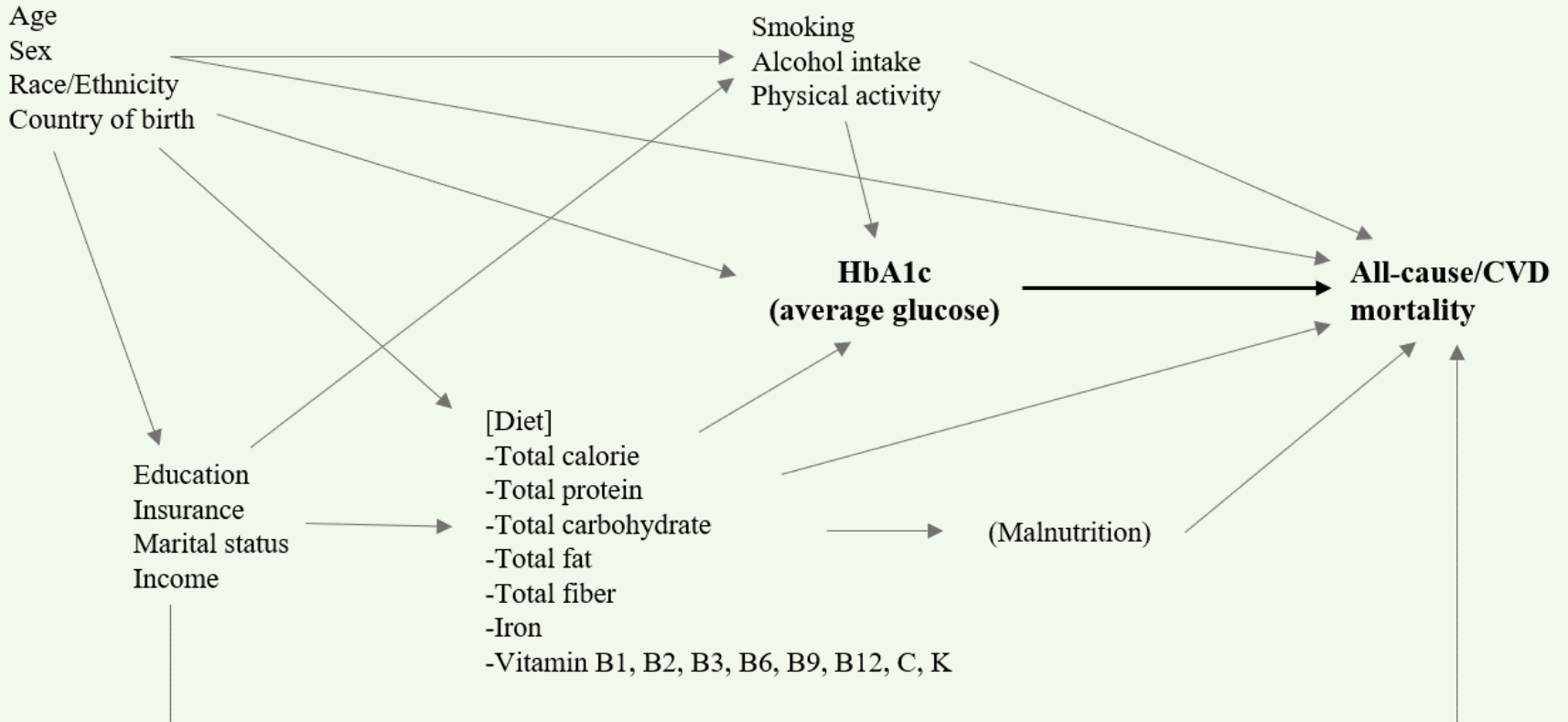
Low HbA1c,	4.0-4.9%
Mid-level HbA1c,	5.0-5.6%
Prediabetes,	5.7-6.4%
Diabetes,	≥6.5% or taking antidiabetic therapies

➤ Covariates: Demographics, social, lifestyle, and clinical information (72 variables selected through DAG). Missing data were imputed by random forest.

➤ Outcomes: All-cause mortality, cardiovascular mortality

Causal diagram

Demographic characteristics & Lifestyle & Diet



Parametric g-formula in survival analysis

- Parametric g-formula algorithm was used to estimate **the risk of death at 5 and 10 years under each HbA1c group.**

Fit the outcome prediction models
using the exposure and covariates

Predict the values of the potential outcomes
under a hypothetical intervention on the exposure

Estimate the average marginal effect of
exposure on outcome (risk ratio, risk difference)

Parametric g-formula in survival analysis

Application Steps

Step 1: Arrange data with a person-time structure.

ID	Month	HbA1c	Mortality
1	0	Low	0
1	1	Low	0
...
1	59	Low	0
1	60	Low	1
2	0	Mid-level	0
2	1	Mid-level	0
...
2	149	Mid-level	0
2	150	Mid-level	0

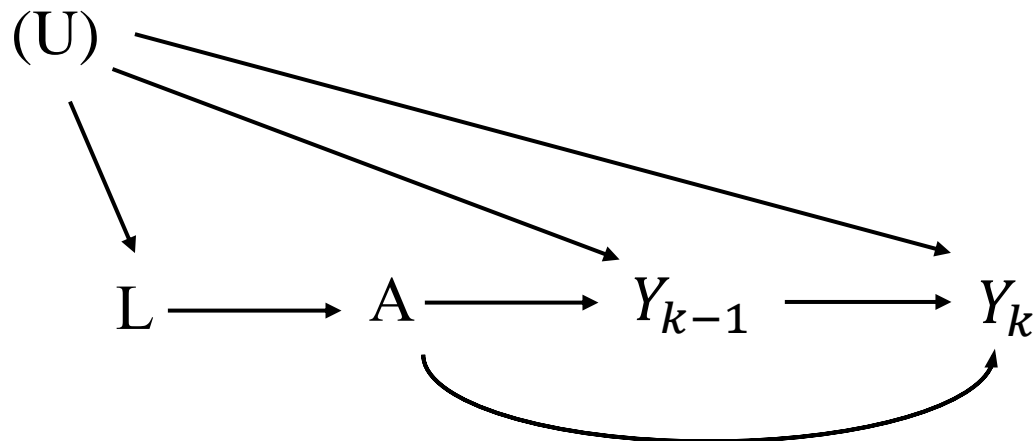
Parametric g-formula in survival analysis

Application Steps

Step 1: Arrange data with a person-time structure.

Step 2: Parametric estimation of conditional hazards

$$\text{logit Pr}[Y_{k+1}=1|Y_k=0, A, L] = \theta_{0,k} + \theta_1 A + \theta_2 Ak + \theta_3 Ak^2 + \theta_4 L$$
$$(\theta_{0,k} = \theta_0 + \theta_5 k + \theta_6 k^2)$$



Parametric g-formula in survival analysis

Application Steps

Step 1: Arrange data with a person-time structure.

Step 2: Parametric estimation of conditional hazards

Step 3: Create copies of the original sample under hypothetical intervention on the exposure.

Parametric g-formula in survival analysis

Application Steps

Step 1: Arrange data with a person-time structure.

Step 2: Parametric estimation of conditional hazards

Step 3: Create copies of the original sample under hypothetical intervention on the exposure.

Step 4: Using the model in step 2, calculate the potential outcome under each hypothetical intervention.

Parametric g-formula in survival analysis

Application Steps

Step 1: Arrange data with a person-time structure.

Step 2: Parametric estimation of conditional hazards

Step 3: Create copies of the original sample under hypothetical intervention on the exposure.

Step 4: Using the model in step 2, calculate the potential outcome under each hypothetical intervention.

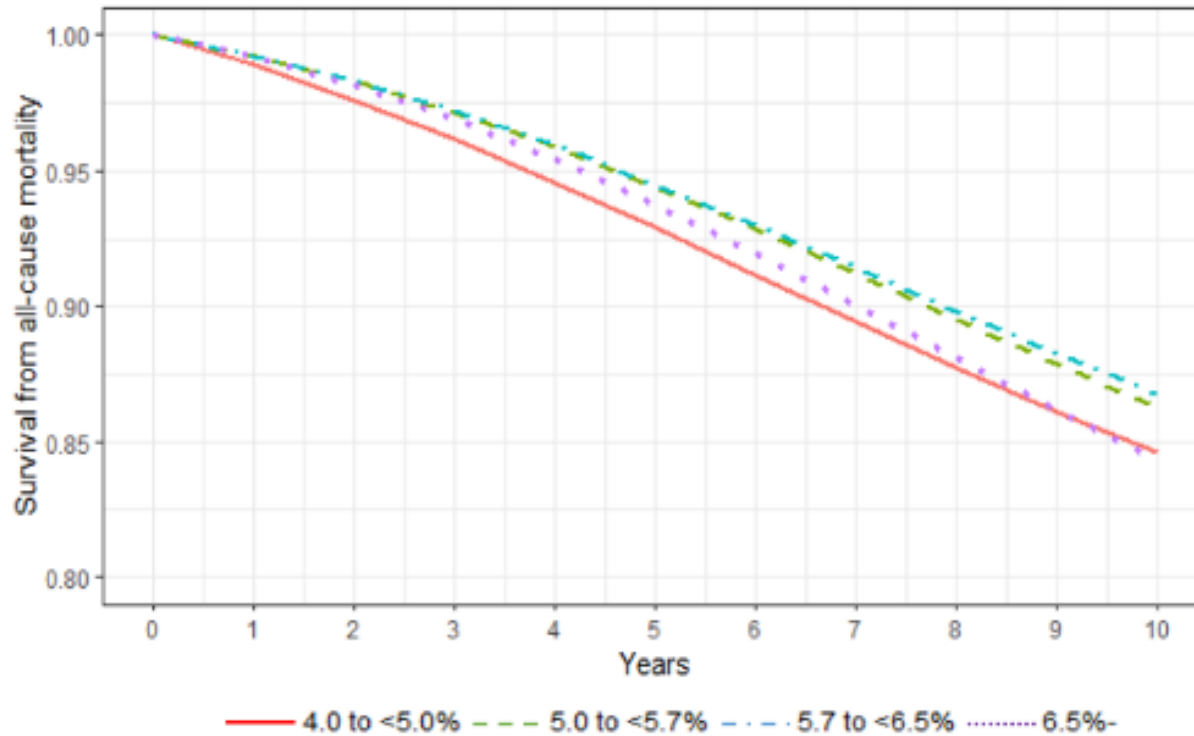
Step 5: Computation of weighted average of survivals

- Robust 95% confidence intervals were estimated by repeating the analysis on 1,000 bootstrapped samples.

Results

- Mean age: 49.5 years
- Male: Female = 1:1
- Median duration of follow-up: 7.5 years
- All-cause deaths: 5,118 (13%)
Cardiovascular deaths: 1,116 (3%)
- Low HbA1c group were more likely to be younger, female, Non-Hispanic White, highly educated, and never smoker
- Prediabetes and diabetes group were generally older, Non-Hispanic Black, less educated, smoker, and have comorbidities.

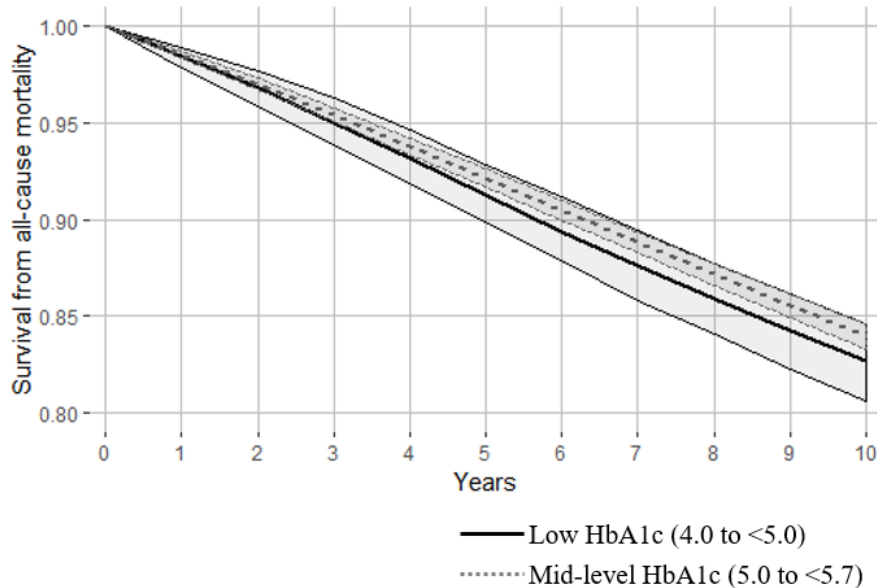
Low HbA1c showed increased risk of all-cause mortality



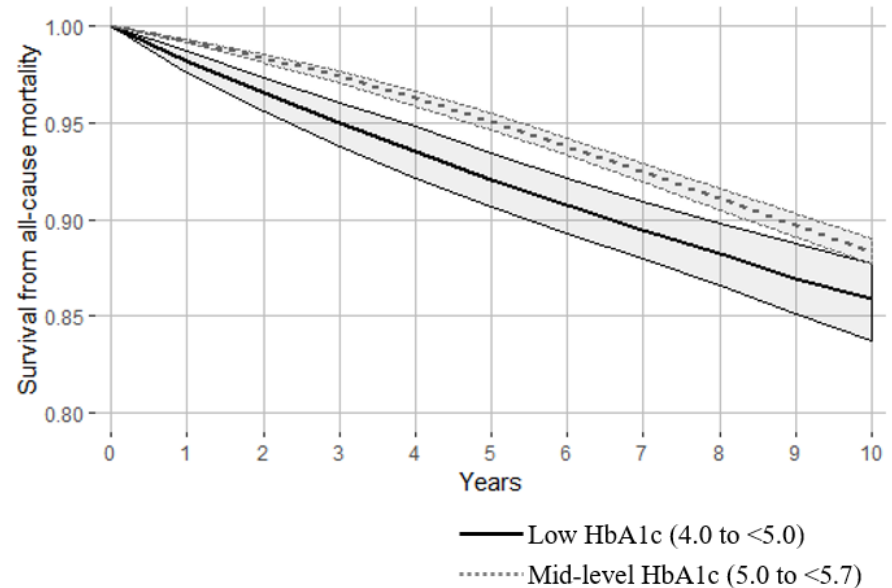
Low HbA1c vs Mid-level HbA1c	Follow-up periods	
	5 years	10 years
Adjusted risk ratio (95% CI)	1.30 (1.16 to 1.48)	1.12 (1.03 to 1.22)
Adjusted risk difference (95% CI)	+1.83% (1.02 to 2.97)	+1.66% (0.35 to 3.00)

The association between low HbA1c and all-cause mortality was larger among females

Male



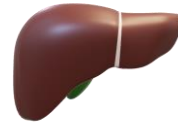
Female



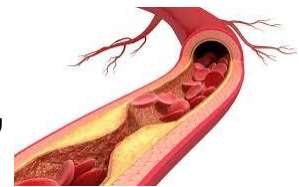
Low HbA1c vs Mid-level HbA1c Adjusted risk ratio (95% CI)	Follow-up periods	
	5 years	10 years
Male	1.11 (0.88 to 1.32)	1.07 (0.96 to 1.22)
Female	1.61 (1.30 to 1.90)	1.21 (1.04 to 1.40)

Possible mechanisms

- Poor health status among people with low HbA1c
 - malnutrition
 - unfavorable profiles of red blood cell related factors
 - inflammation
 - decreased liver function
 - an early stage of chronic disease



- Hypoglycemia
 - ⇒ sympathoadrenal activation, inflammation, and endothelial dysfunction
 - ⇒ chronic disease and cardiometabolic diseases.



Summary of this study

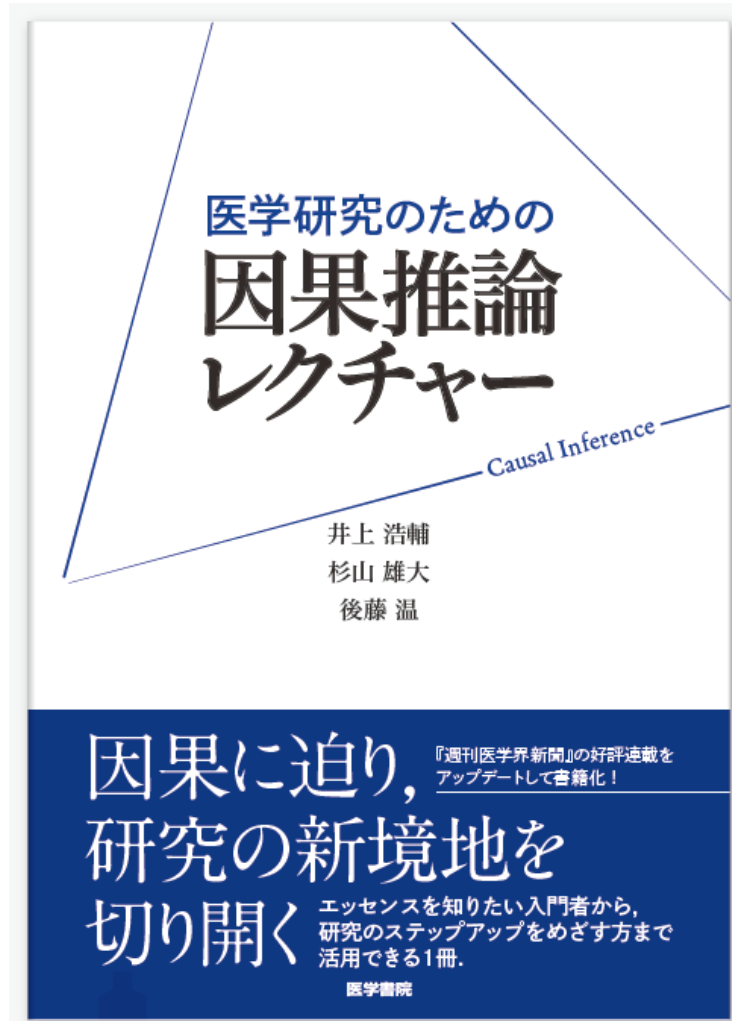
- Using the large national database of US adults, we found that **low HbA1c was associated with increased risk of all-cause mortality** at 5 and 10 years of follow-up.
- The association was **stronger among females** than males, particularly at 5 years
- Our findings highlight **that low HbA1c among people without diabetes may need to be carefully monitored.**

本日のポイント

時間変動する曝露の効果を推定する際は

- DAGを用いることで変数の複雑なかわりを整理できる。
- 時間変動する交絡因子について考える必要がある。
- Inverse probability of treatment weight、G-computationといった手法が使える。
- 問いを明確にすることが重要である。

ご清聴いただき、ありがとうございました！



Web: <http://endoepi.net>

Twitter: @ki_endoepi

Email: inoue.kosuke.2j@kyoto-u.ac.jp

2024/4 医学書院より発売予定

References

1. 臨床研究・疫学研究のための因果推論レクチャー [第7回] 時間とともに変化する曝露を扱う
2. Pearl, J., & Mackenzie, D. (2018). *The book of why: the new science of cause and effect*. Basic books.
3. Miguel, A., HERNAN, R., & James, M. (2023). *Causal inference: what if*.
4. Zhong, Q. Y., Gelaye, B., VanderWeele, T. J., Sanchez, S. E., & Williams, M. A. (2018). Causal model of the association of social support with antepartum depression: a marginal structural modeling approach. *American journal of epidemiology*, 187(9), 1871-1879. Peter Austin
5. Lee, B. K., Lessler, J., & Stuart, E. A. (2011). Weight trimming and propensity score weighting. *PloS one*, 6(3), e18174.
6. Daniel, R. M., Cousens, S. N., De Stavola, B. L., Kenward, M. G., & Sterne, J. A. C. (2013). Methods for dealing with time - dependent confounding. *Statistics in medicine*, 32(9), 1584-1618.
7. Shiba, K., Torres, J. M., Daoud, A., Inoue, K., Kanamori, S., Tsuji, T., ... & Kawachi, I. (2021). Estimating the impact of sustained social participation on depressive symptoms in older adults. *Epidemiology*, 32(6), 886-895.
8. Inoue, K., Goto, A., Kondo, N., & Shinozaki, T. (2022). Bias amplification in the g-computation algorithm for time-varying treatments: a case study of industry payments and prescription of opioid products. *BMC Medical Research Methodology*, 22(1), 1-10.
9. Inoue, K., Mayeda, E. R., Nianogo, R., Paul, K., Yu, Y., Haan, M., & Ritz, B. (2021). Estimating the joint effect of diabetes and subsequent depressive symptoms on mortality among older Latinos. *Annals of Epidemiology*, 64, 120-126. Inoue, K et al. (2021).
10. Inoue, K., Nianogo, R., Telesca, D., Goto, A., Khachadourian, V., Tsugawa, Y., ... & Ritz, B. (2021). Low HbA1c levels and all-cause or cardiovascular mortality among people without diabetes: the US National Health and Nutrition Examination Survey 1999–2015. *International journal of epidemiology*, 50(4), 1373-1383.