



# SASによる直接効果と間接効果の推定

○矢田 真城<sup>1\*</sup> 魚住龍史<sup>2</sup>

<sup>1</sup>エイツーヘルスケア株式会社 生物統計部第1部

<sup>2</sup>京都大学大学院医学研究科 医学統計生物情報学

Estimation of direct and indirect effects by using SAS

Shinjo Yada<sup>1\*</sup> and Ryuji Uozumi<sup>2</sup>

<sup>1</sup> *A2 Healthcare Corporation*

<sup>2</sup> *Kyoto University Graduate School of Medicine*

\**email: [yada-s@a2healthcare.com](mailto:yada-s@a2healthcare.com)*

要旨:

近年のSAS/STATでは、因果推論に関連したプロシジャがバージョンアップのたびに次々と追加されている。

本発表では、これら因果推論に関連したプロシジャのうち、CAUSALMEDプロシジャ、CAUSALGRAPHプロシジャをとりあげ、実行方法と出力結果について説明する。

キーワード:

CAUSALMEDプロシジャ, 直接効果, 間接効果, 効果の分解,  
CAUSALGRAPHプロシジャ, 因果効果の識別可能条件,  
バックドア基準, 操作変数

- CAUSALMEDプロシジャによる解析
  - 適用例の紹介
  - 総合効果の分解
  - 解析内容とSAS出力結果
- CAUSALGRAPHプロシジャによる解析
  - 適用例の紹介
  - 因果効果の識別可能条件
  - 解析内容とSAS出力結果

- CAUSALMEDプロシジャによる解析
  - 適用例の紹介
  - 総合効果の分解
  - 解析内容とSAS出力結果
- CAUSALGRAPHプロシジャによる解析
  - 適用例の紹介
  - 因果効果の識別可能条件
  - 解析内容とSAS出力結果

# 因果効果の分解と推定

- CAUSALMEDプロシジャによる解析
  - SAS/STAT 14.3 にて追加されたプロシジャのひとつ
  - 興味ある因果効果を，処理変数が中間変数を経由して反応変数に及ぼす効果（間接効果）と，経由しなかったときの効果（直接効果）に分解し，推定する
  - VanderWeeleによる4つの要素への分解と推定を実装
  - 生存時間データに対応した解析，より複雑な状況での解析には別途SASプログラムが必要（現時点では）

- 家庭環境が子供の認知能力にもたらす影響

Marjoribanks (1974) *Environments for Learning*

- 両親によりもたらされる家庭環境が、子供の認知能力に大きな影響を与える

- 子供本人の学習意欲は？

- 学習意欲も家庭環境に左右されているはず
- 認知能力に与える影響のうち、どの程度が、本人の学習意欲で説明できるだろうか

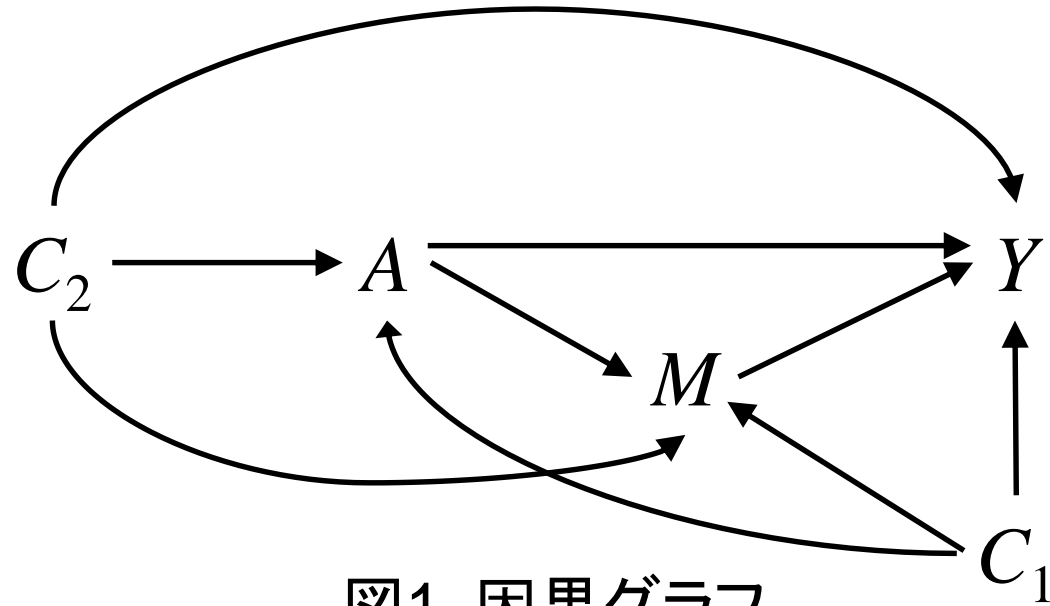


認知能力・・・IQや偏差値など数値化できる能力(学力)

# 適用例の因果グラフ

- 家庭環境が子供の認知能力にもたらす影響
- 知りたいこと

家庭環境に左右される子供の学習意欲が、どれだけ本人の認知能力に影響を与えるか



Y: 子供の認知度テストスコア  
A: 親のサポートスコア  
M: 子供の学習スコア  
C1: 家族数  
C2: 親の社会的身分スコア

図1. 因果グラフ

# 使用するSASデータセット

- Maroribanksのデータ

| 変数名        | 内容                          | 変数のタイプ |
|------------|-----------------------------|--------|
| StudentID  | 対象者識別ID                     | 数値型変数  |
| CogPerform | Y: 子供の認知度テストスコア             | 数値型変数  |
| Encourage  | A: 親のサポートスコア                | 数値型変数  |
| Motivation | M: 子供の学習スコア                 | 数値型変数  |
| FamSize    | C <sub>1</sub> : 家族数        | 数値型変数  |
| SocStatus  | C <sub>2</sub> : 親の社会的身分スコア | 数値型変数  |



# あてはめるモデル

- 下記の回帰モデルを想定

$$E[Y | A = a, M = m, C = c] = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_{41} c_1 + \theta_{42} c_2$$

$$E[M | A = a, C = c] = \beta_0 + \beta_1 a + \beta_{21} c_1 + \beta_{22} c_2$$

$Y$ : 子供の認知度テストスコア,  $A$ : 親のサポートスコア

$M$ : 子供の学習スコア,  $C_1$ : 家族数,  $C_2$ : 親の社会的身分スコア

Marjoribanksのデータに上記モデルをあてはめて,  
親のサポートスコア $A$  → 認知度テストスコア $Y$ の効果,  
親のサポートスコア $A$  → 学習スコア $M$  → 認知度テストスコア $Y$ の効果を推定する

# 潜在的な変数

- 潜在的な反応

個人ごとに有する, ある処理を割り当てられたときの潜在的な反応

$Y(a)$ : 処理  $A = a$  を受けたときに観測されるであろう反応変数

$M(a)$ : 処理  $A = a$  を受けたときに観測されるであろう中間変数

$Y(a, m)$ :  $A = a, M = m$  を受けたときに観測されるであろう反応変数

実際に観測されるのは1つだけで,

それ以外は観測されない = 事実と反する 事象をモデル化

反事実モデル (counterfactual model)

# 総合効果の分解

- 総合効果の分解：要素数2

- 処理レベル  $A = a^*$  を基準として処理レベル  $A = a$  を比較するとき

$$TE(a, a^*) = PDE(a, a^*) + TIE(a, a^*)$$

- 自然な直接効果 ( $PDE$ )

$$PDE(a, a^*) = Y(a, M(a^*)) - Y(a^*, M(a^*))$$

- 自然な間接効果 ( $TIE$ )

$$TIE(a, a^*) = Y(a, M(a)) - Y(a, M(a^*))$$

※総合効果を因果リスク差として定義した場合

# 総合効果の分解

- 総合効果の分解：要素数2（続き）

$$TE(a, a^*) = TDE(a, a^*) + PIE(a, a^*)$$

- 総合直接効果 ( $TDE$ )

$$TDE(a, a^*) = Y(a, M(a)) - Y(a^*, M(a))$$

(4)における中間変数 $M(a^*)$ を $M(a)$ としたときの効果

- 純粋な間接効果 ( $PIE$ )

$$PIE(a, a^*) = Y(a^*, M(a)) - Y(a^*, M(a^*))$$

(5)における処理レベル $a$ を $a^*$ としたときの効果

※総合効果を因果リスク差として定義した場合

- VanderWeeleによる総合効果の分解：要素数3

VanderWeele (2013b) *Epidemiology*

- 処理レベル  $A = a^*$  を基準として処理レベル  $A = a$  を比較するとき

$$TE(a, a^*) = CDE(a, a^*; m^*) + PAI(a, a^*; m^*) + PIE(a, a^*)$$

- 制御された直接効果 (*CDE*)

中間変数を  $M = m^*$  に制御(固定)したときに、  
処理レベルを  $A = a^*$  から  $A = a$  に変化させたときの効果

- 生物学的交互作用に起因する効果 (*PAI*)

生物学的交互作用を定量化した効果

- VanderWeeleによる総合効果の分解：要素数4

VanderWeele (2014) *Epidemiology*

$$TE(a, a^*) = CDE(a, a^*; m^*) + INT_{ref}(a, a^*; m^*) + INT_{med}(a, a^*) + PIE(a, a^*)$$

- 生物学的交互作用に起因する効果

$$PAI(a, a^*; m^*) = INT_{ref}(a, a^*; m^*) + INT_{med}(a, a^*)$$

- $INT_{ref}$  : 参照交互作用 (reference interaction)

$$\sum_m \{Y(a, m) - Y(a^*, m) - Y(a, m^*) + Y(a^*, m^*)\} I(M(a^*) = m)$$

- $INT_{med}$  : 媒介交互作用 (mediated interaction)

$$\sum_m \{Y(a, m) - Y(a^*, m) - Y(a, m^*) + Y(a^*, m^*)\} \{I(M(a) = m) - I(M(a^*) = m)\}$$

※総合効果を因果リスク差として定義した場合

# 間接効果に関する指標

- 間接効果が総合効果に対して与える影響の指標

- Proportion mediated

中間変数を経由して反応変数に与える影響がどの程度あるか

$$E[TIE(a, a^*)] / E[TE(a, a^*)] \quad (\text{純粹な間接効果} / \text{総合効果})$$

- Proportion eliminated

VanderWeele (2013a) *Epidemiology*

中間変数がある値  $m^*$  に制御(固定)したときに, 総合効果のうちどの程度除去できるのかを表す指標

$$\{E[TE(a, a^*)] - E[CDE(a, a^*; m^*)]\} / E[TE(a, a^*)]$$

(総合効果 - 制御された直接効果) / 総合効果

# CAUSALMEDプロシージャによる解析

- SASプログラム例

プログラム1

Y: 反応変数, A: 処理変数, M: 中間変数

```
proc causalmed data = Cognitive decomp poutcomemod pmedmod ;
```

decomp... 要素数の指定がなければ2つ/3つ/4つの要素へ分解

poutcomemod... Yへの回帰式での回帰係数についての統計量を出力

pmedmod... Mへの回帰式での回帰係数についての統計量を出力

```
model CogPerform = Encourage | Motivation ;
```

modelステートメントの左辺にYを, 右辺にMとAを指定  
MとAの交互作用をモデルに含めたい場合,  
A|M, または A M A|M と指定



# CAUSALMEDプロシージャによる解析

- SASプログラム例(続き)

プログラム1

Y: 反応変数, A: 処理変数, M: 中間変数

**mediator** Motivation = Encourage ;

mediatorステートメントの左辺にMを, 右辺にAを指定

**covar** FamSize SocStatus ;

covarステートメントでモデルに含めたい共変量を指定

**bootstrap** bootcii(bc) nsamples = 1000 seed = 4989 ;

各因果効果推定値の標準誤差と信頼区間をブートストラップ法に基づいて算出したい場合に指定

# ブートストラップ法に基づく信頼区間

- ブートストラップ法

観測されたデータから、リサンプリングすることにより生成された擬似的なデータを用いて、パラメータ推定量の統計的誤差や統計量の分布を推定する方法



- ブートストラップ法に基づく信頼区間の構成

- パーセンタイル法に基づく信頼区間 Efron (1981) *Can J Stat.*

- $BC_a$  信頼区間: バイアス修正加速化信頼区間 Efron (1987) *J Am Stat Assoc.*

本資料では,  $BC_a$  信頼区間において加速定数  $a$  を 0 としたときの  $BC_a$  信頼区間 (bias-corrected confidence interval) を,  $BC$  信頼区間と表記する.

・・・など

- 制御された直接効果を例にとると

Efron (1987) *J Am Stat Assoc.*

$n$ 個の個体から、制御された直接効果(CDE)の推定値 $\widehat{CDE}$ が得られたとする。このとき

Step 1: 大きさ $n$ のブートストラップ標本を,  $B$ 組反復生成させる

- 大きさ $n$ のブートストラップ標本

区間(0,1]上で発生させた $n$ 個の一樣乱数 $u_1, u_2, \dots, u_n$ を用いて, 観測されたデータから,  $[nu_1]+1$ 番目,  $[nu_2]+1$ 番目,  $\dots$ ,  $[nu_n]+1$ 番目にあたるデータを抽出することで構成される。

※  $[\cdot]$ : かぎ括弧の中の値をこえない最大の整数値

- 制御された直接効果を例にとると(続き)

Efron (1987) *J Am Stat Assoc.*

Step 2: 得られる $B$ 個の $CDE$ の推定値を

$\widehat{CDE}_{(1)}^* \leq \widehat{CDE}_{(2)}^* \leq \dots \leq \widehat{CDE}_{(B)}^*$  と大きさの順に並べ替えた上で,  
 $CDE$ の $100(1-2\alpha)\%$ 信頼区間を $(\widehat{CDE}_{(B\hat{\alpha}_1)}^*, \widehat{CDE}_{(B\hat{\alpha}_2)}^*)$  と推定する.

$\hat{\alpha}_1, \hat{\alpha}_2$  は,  $\Phi()$ を標準正規分布に従う確率変数の累積分布関数,  
 $a$ を加速定数,  $z_\alpha$ を標準正規分布の $100\alpha\%$ 点として

$$\hat{\alpha}_1 = \Phi\left(z_0 + \frac{z_0 + z_\alpha}{1 - a(z_0 + z_\alpha)}\right), \quad \hat{\alpha}_2 = \Phi\left(z_0 + \frac{z_0 + z_{1-\alpha}}{1 - a(z_0 + z_{1-\alpha})}\right)$$

バイアス修正定数 $z_0$ は  $z_0 = \Phi^{-1}\left(\frac{1}{B} \sum_{b=1}^B I(\widehat{CDE}_{(b)}^* \leq \widehat{CDE})\right)$

$\widehat{CDE}^*(b)$ は $b$ 番目のブートストラップ標本に基づく $CDE$ の推定値( $b = 1, 2, \dots, B$ ),  
 $I()$ はカッコの中の式が成立すれば1を, それ以外は0をとる指示関数

# CAUSALMEDプロシージャによる実行結果

- 因果効果の推定結果

表1. Maroribanksのデータに対する因果効果の推定結果

| 効果                        | 推定値     | ブートストラップ法 |                 |
|---------------------------|---------|-----------|-----------------|
|                           |         | 標準誤差      | 95%BC信頼区間       |
| 総合効果                      | 6.8421  | 0.2311    | 6.3603 7.2954   |
| 制御された直接効果                 | 4.1797  | 0.0561    | 4.3278 4.5540   |
| 自然な直接効果                   | 4.1509  | 0.0554    | 4.0527 4.2820   |
| 自然な間接効果                   | 2.6912  | 0.2375    | 2.2219 3.1670   |
| Percentage mediated (%)   | 39.3325 | 2.2153    | 34.9090 43.7548 |
| Percentage eliminated (%) | 38.9128 | 2.2902    | 30.8853 39.9757 |

BC信頼区間・・・bias-corrected confidence interval.

# CAUSALMEDプロシージャによる実行結果

- 総合効果の分解結果

表2. Maroribanksのデータに対する総合効果の分解及び推定結果

| 効果                                  | 推定値     | ブートストラップ法 |           |         |
|-------------------------------------|---------|-----------|-----------|---------|
|                                     |         | 標準誤差      | 95%BC信頼区間 |         |
| 総合効果                                | 6.8421  | 0.2311    | 6.3603    | 7.2954  |
| 制御された直接効果                           | 4.1797  | 0.0554    | 4.0847    | 4.3119  |
| 参照交互作用                              | -0.0287 | 0.0034    | -0.0353   | -0.0222 |
| 媒介交互作用                              | 0.0574  | 0.0054    | 0.0462    | 0.0677  |
| 純粹な間接効果                             | 2.6338  | 0.2325    | 2.1740    | 3.0993  |
| Portion attributable to interaction | 0.0287  | 0.0023    | 0.0234    | 0.0328  |

Portion attributable to interaction: 生物学的交互作用に起因する効果 (PAI)

# 因果効果の分解と推定

- 平均的な効果

$$\text{制御された直接効果} = (\theta_1 + \theta_3 m)(a - a^*)$$

$$\text{自然な直接効果} = (\theta_1 + \theta_3 \beta_0 + \theta_3 \beta_1 a^* + \theta_3 \beta_2 c)(a - a^*)$$

etc.

CAUSALMEDプロシジャでは, 特に指定しない限り,

$$a = (\text{データ平均値}) + 0.5, a^* = (\text{データ平均値}) - 0.5,$$

$$m = (\text{データ平均値}), \text{共変量} = (\text{データ平均値})$$

として算出される. これらを特定の値に指定したい場合への対応として,  
**EVALUATEステートメント**が用意されている.

# EVALUATEステートメント

- 評価変数, 変数レベルを指定したいとき

- 処理変数, 中間変数, 共変量について, ある特定の値を指定するためのステートメント
- ひとつのプロシジャの中で複数指定できる

/\* 家族数を5に指定 \*/

```
evaluate 'C1 = 5' FamSize = 5 ;
```

/\* 家族数を2, 親の社会的身分スコアをデータ最大値に指定 \*/

```
evaluate 'C1 = 2 C2 = max' FamSize = 2 SocStatus = max ;
```



# EVALUATEステートメント使用例

- SASプログラム例

中間変数である子供の学習意欲スコアを,  
データ平均値  $+1.5 \times$  [データの標準偏差] に制御した場合

## プログラム2

```
proc causalmed data = Cognitive ;  
  model CogPerform = Encourage | Motivation ;  
  mediator Motivation = Encourage ;  
  covar FamSize SocStatus ;  
  bootstrap bootci(bc) nsamples = 1000 seed = 4989 ;  
  evaluate 'motivation = mean + 1.5SD' Motivation = 1.5(SD) ;  
run ;
```

# EVALUATEステートメント使用例

- SAS出力結果

表3. 子供の学習意欲が高い集団での因果効果の推定結果

| 効果                        | 推定値     | ブートストラップ法 |           |         |
|---------------------------|---------|-----------|-----------|---------|
|                           |         | 標準誤差      | 95%BC信頼区間 |         |
| 総合効果                      | 6.8421  | 0.2311    | 6.3603    | 7.2954  |
| <u>制御された直接効果</u>          | 4.4206  | 0.0561    | 4.3278    | 4.5540  |
| 自然な直接効果                   | 4.1509  | 0.0554    | 4.0527    | 4.2820  |
| 自然な間接効果                   | 2.6912  | 0.2375    | 2.2219    | 3.1670  |
| Percentage mediated (%)   | 39.3325 | 2.2153    | 34.9090   | 43.7548 |
| Percentage eliminated (%) | 35.3916 | 2.2902    | 30.8853   | 39.9757 |

中間変数の制御を, データ平均値から, データ平均値  $+1.5 \times$  [データの標準偏差] に変えたことで, 制御された直接効果の推定値が少し高くなっている.

# EVALUATEステートメント使用例

- SAS出力結果(続き)

表4. 子供の学習意欲が高い集団での総合効果の分解と推定結果

| 効果   | 推定値     | ブートストラップ法 |           |         |
|--|---------|-----------|-----------|---------|
|  |         | 標準誤差      | 95%BC信頼区間 |         |
| 総合効果                                       | 6.8421  | 0.2311    | 6.3603    | 7.2954  |
| <u>制御された直接効果</u>                           | 4.4206  | 0.0561    | 4.3278    | 4.5540  |
| <u>参照交互作用</u>                              | -0.2696 | 0.0094    | -0.2880   | -0.2509 |
| 媒介交互作用                                     | 0.0574  | 0.0054    | 0.0462    | 0.0677  |
| 純粹な間接効果                                    | 2.6338  | 0.2325    | 2.1740    | 3.0993  |
| <u>Portion attributable to interaction</u> | -0.2122 | 0.0069    | -0.0259   | -0.1989 |

中間変数の制御を, データ平均値からデータ平均値  $+1.5 \times$  [データの標準偏差] に変えたことで, 制御された直接効果, 参照交互作用, Portion attributable to interactionが変わる.

- CAUSALMEDプロシジャによる解析
  - 適用例の紹介
  - 総合効果の分解
  - 解析内容とSAS出力結果
- CAUSALGRAPHプロシジャによる解析
  - 適用例の紹介
  - 因果効果の識別可能条件
  - 解析内容とSAS出力結果

# 因果効果の識別可能条件

- CAUSALGRAPHプロシジャによる解析
  - SAS/STAT 15.1 にて追加されたプロシジャのひとつ
  - 興味のある因果効果を識別し、因果効果を推定するために必要となる変数集合を選択する
  - 調整基準(バックドア基準を包含するconstructive backdoor criterionと同一)、条件付き操作変数による変数選択が可能
  - 未観測変数が存在する場合にも対応
  - 変数間の因果関係は解析者が事前に想定

- PFASの曝露が母乳期間に与える影響
- 望まない早期離乳を誘導する母乳不足
  - 母乳不足の原因のひとつに環境化学物質があるのでは
  - パーフルオロアルキル化合物(PFAS)の影響



- 母親の血清中PFAS濃度の増加に伴う、多産婦の母乳期間低下

*Fei et al. (2010) Scand J Work Environ Health.*

- 初産婦及び多産婦を対象とした出生コホートデータによる解析

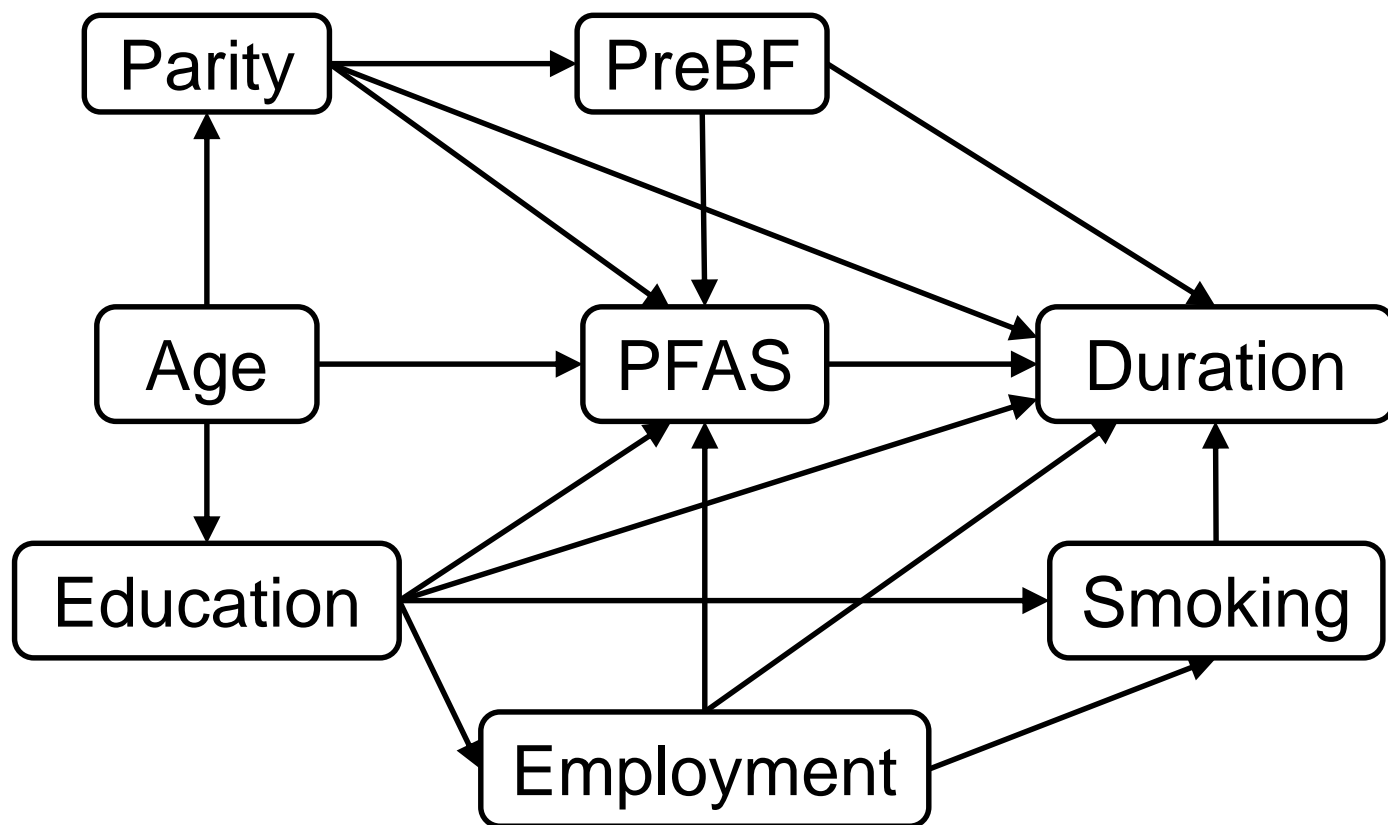
PFASと母乳期間の交絡要因：母乳育児経験の有無を考慮

*Timmermann et al.(2017) Reprod Toxicol.*

- PFASの曝露が母乳期間に与える影響
  - PFASの曝露の母乳期間への因果効果の推定
  - 因果効果を識別し推定する際に含めるべき共変量の同定
  - 出生コホート研究のデータ
    - 反応変数・・・母乳期間
    - 処置変数・・・PFASの曝露
    - 共変量・・・ 母体年齢, 母親の学歴, 母親の就業状況, 妊娠中の喫煙の有無, 過去の母乳期間, 妊娠中のアルコール摂取有無, 妊娠中のBMI

# 因果ダイアグラム

- PFASの曝露と母乳期間との因果関係を表現



- PFAS…母親のPFASへの曝露
- Duration…母乳期間
- PreBF…過去の母乳期間
- Parity…出産経験の有無
- Age…母体年齢
- Education…母親の学歴
- Employment…母親の就業状況
- Smoking…妊娠中の喫煙の有無

図2. PFASの曝露と母乳期間の因果ダイアグラム



# CAUSALGRAPHプロシージャによる解析

- SASプログラム例

## プログラム3

```
proc causalgraph method = adjustment ;  
  model "Figure 2"  
    PFAS ==> Duration,  
    PreBF ==> PFAS Duration,  
    Parity ==> PreBF Duration PFAS,  
    Age ==> Parity PFAS Education,  
    Education ==> PFAS Duration Smoking Employment,  
    Employment ==> PFAS D  
    Smoking ==> Duration ;  
  ...
```

MODELステートメントにて、  
全ての変数の因果関係を矢線(==>)で記述  
ひと組ずつカンマ(,)で区切って指定

# 共変量調整

- 6つの共変量全てを用いた調整は妥当かな？
- CAUSALGRAPHプロシジャで確認

## プログラム4

```
proc causalgraph method = adjustment ;  
  model "Figure 2"  
    PFAS ==> Duration,  
    ...  
    Smoking ==> Duration ;  
  identify PFAS ==> Duration ;  
  testid "All covariate" PreBF Parity Age Education Employment Smoking ;  
run ;
```

オプションMETHODで  
因果効果の識別条件を指定

IDENTIFYステートメントにて、  
処理変数 ==> 反応変数 ;と記述

TESTIDステートメントにて、調べたい変数を記述

# 因果効果の識別可能条件

- バックドア基準

非巡回有向グラフ $G$ において,  $X$  は $Y$  の非子孫であるとする.

このとき, 以下の2つの条件を満たす変数集合 $Z$  は $(X, Y)$  についてバックドア基準を満たすという.

i)  $Z$  は $X$  の子孫ではない.

ii) グラフ $G_{\underline{X}}$ において,  $Z$  が $X$  と $Y$  を有向分離する.

( $G_{\underline{X}}$  :  $X$  から出る全ての矢線を取り除いたグラフのこと)

黒木 訳 (2009) 統計的因果推論

# 因果効果の識別可能条件

- 条件付き操作変数

非巡回有向グラフ $G$ において、以下のA1, A2を満たす変数集合 $Z$ を、変数集合 $W$ を与えたときの $(X, Y)$ に対する条件付き操作変数であるという。

(A1)  $Z$ と $W$ に含まれる任意の頂点は、 $X$ 及び $Y$ の子孫ではない。

(A2) グラフ $G_X$ において、 $W$ は $Z$ と $Y$ を有向分離するが、 $W$ は $Z$ と $X$ を有向分離しない。

- 線形構造方程式モデルの場合、 $W$ を与えたときの条件付き操作変数 $Z$ を用いることで、 $X$ から $Y$ への因果効果が識別可能

# 因果効果の識別可能条件

- 調整基準 (adjustment criterion)

Shpister *et al.*(2010) *Proc of 21th Conf on UAI.*

非巡回有向グラフ $G$ において, 変数集合 $Z$ が以下の条件: B1, B2を満たすとき,  $Z$ を $(X, Y)$ に対する調整基準と定義する.

(B1)  $Z$ は, グラフ $G_{\underline{X}}$ における $X$ から $Y$ への有向パス上にある変数の子孫を含まない.

(B2)  $Z$ は, グラフ $G$ における全ての $X$ から $Y$ への非有向パスをブロックする.

- Constructiveバックドア基準 (constructive backdoor criterion) と同一

Van der Zander *et al.*(2014) *Proc of 13th Conf on CI.*

- 因果効果を識別する全てのかつ唯一の変数集合を検出する基準

# PROC CAUSALGRAPHステートメント

- オプションMETHODにより因果効果の識別条件を指定
  - METHOD = ADJUSTMENT
    - デフォルトの設定であり, 調整基準が適用される
  - METHOD = BACKDOOR
    - バックドア基準が適用される
    - バックドア基準を満たす変数集合は調整基準を満たすが, その逆は必ずしも成り立たない.
  - METHOD = IV
    - 条件つき操作変数が適用される

# 共変量調整の妥当性

- 6つの共変量全てを用いた調整は妥当かな？
- SAS出力結果

表5. 共変量調整の妥当性確認結果

| 共変量調整のテスト: All covariate |     |     |    |     |           |            |        |       |         |
|--------------------------|-----|-----|----|-----|-----------|------------|--------|-------|---------|
| PFAS の因果効果 (Duration)    |     |     |    |     |           |            |        |       |         |
| モデル                      | サイズ | 有効  | 最小 | 共変量 |           |            |        |       |         |
|                          |     |     |    | Age | Education | Employment | Parity | PreBF | Smoking |
| Figure 2                 | 6   | Yes | No | *   | *         | *          | *      | *     | *       |

- カラム「有効」が”Yes” ⇒ 指定した変数による調整は妥当
- カラム「最小」が”No” ⇒ 因果効果識別の最小の変数集合ではない

# 因果効果を識別する変数集合

- 因果効果を識別し推定するために必要な最小の変数集合は？
- CAUSALGRAPHプロシジャで検討

## プログラム5

```
proc causalgraph method = adjustment minimal ;  
  model "Figure 3"  
    PFAS ==> Duration,  
    ...  
    Smoking ==> Duration ;  
  identify PFAS ==> Duration ;  
  testid "All covariate" PreBF Parity Age Education Employment Smoking ;  
run ;
```

TESTIDステートメントを削除し、  
CAUSALGRAPHステートメントにて  
オプションMINIMALを指定



# 因果効果を識別する変数集合

- 因果効果を識別し推定するために必要な最小の変数集合は？
- SAS出力結果

表6. 調整基準を満たす最小の変数集合

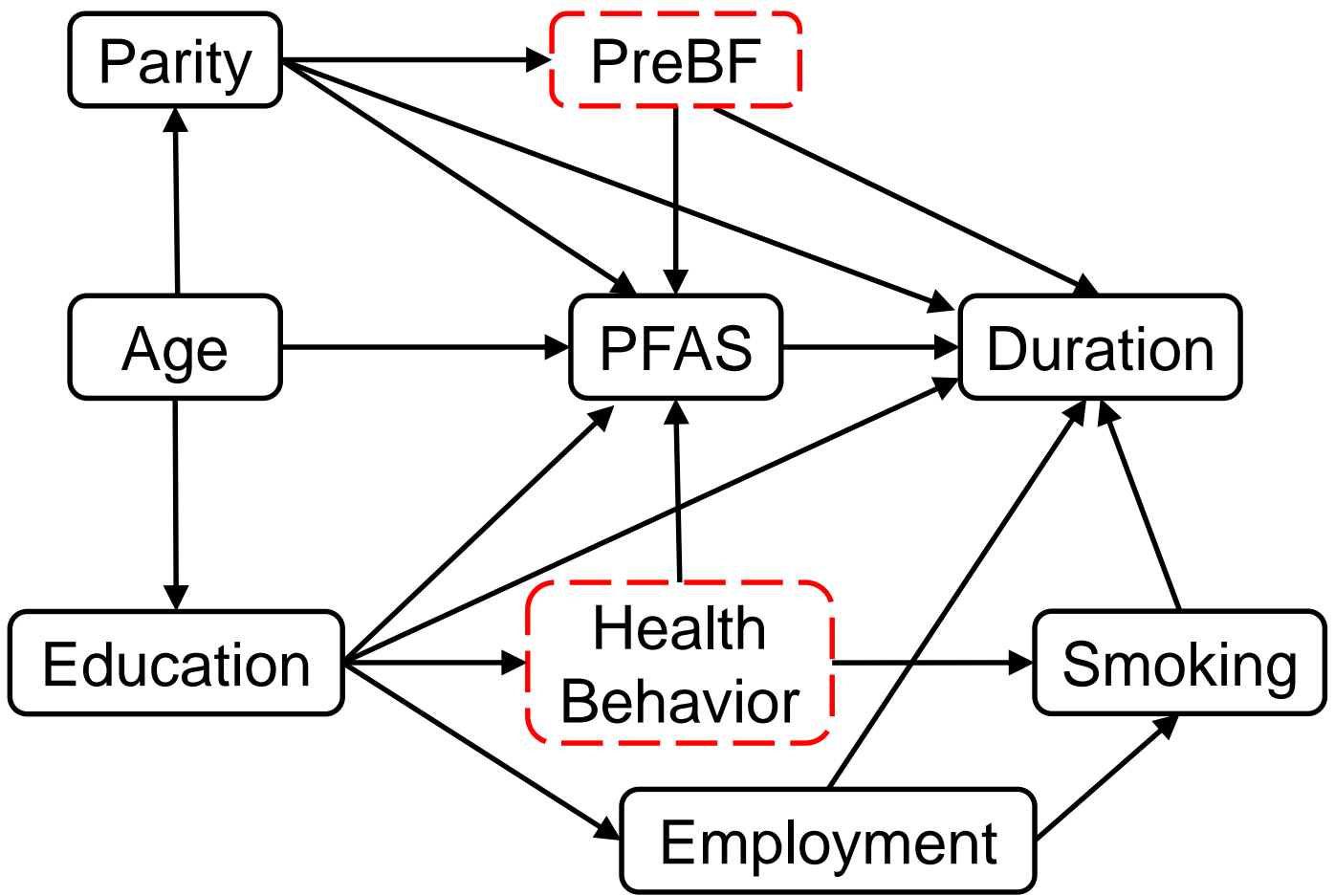
| Covariate Adjustment Sets for Figure 3 |              |     |           |            |        |       |         |
|--|--------------|-----|-----------|------------|--------|-------|---------|
| PFAS の因果効果 (Duration)                  |              |     |           |            |        |       |         |
| サイズ                                    | 最小           | 共変量 |           |            |        |       |         |
|  |              | Age | Education | Employment | Parity | PreBF | Smoking |
| 1                                      | 4 <b>Yes</b> |     | *         | *          | *      | *     |         |

- カラム「最小」が”Yes”

因果効果を識別し推定するために最低限必要となる変数の集合に対してのみ，\*が表示される。

# 未観測変数を含んだ因果ダイアグラム

## ● PFASの曝露と母乳期間との因果関係



- PFAS...母親のPFASへの曝露
- Duration...母乳期間
- PreBF...過去の母乳期間 (未観測)
- Parity...出産経験の有無
- Age...母体年齢
- Education...母親の学歴
- Employment...母親の就業状況
- Smoking...妊娠中の喫煙の有無
- Health Behavior...健康管理活動 (未観測)

図3. PFASの曝露と母乳期間の因果ダイアグラム(未観測変数あり)

(未観測)

# 未観測変数を含めた因果効果の識別

- SASプログラム例

## プログラム6

```
proc causalgraph method = adjustment ;  
  model "Figure 4"  
    PFAS ==> Duration,  
    ...  
    PreBF ==> PFAS Duration,  
    Parity ==> PreBF Duration PFAS,  
    HealthBehavior ==> PFAS Duration Smoking,  
  ;  
  identify PFAS ==> Duration ;  
  unmeasured PreBF HealthBehavior ;  
run ;
```

MODELステートメントでは  
未観測変数も含めて変数間  
の因果関係を記述

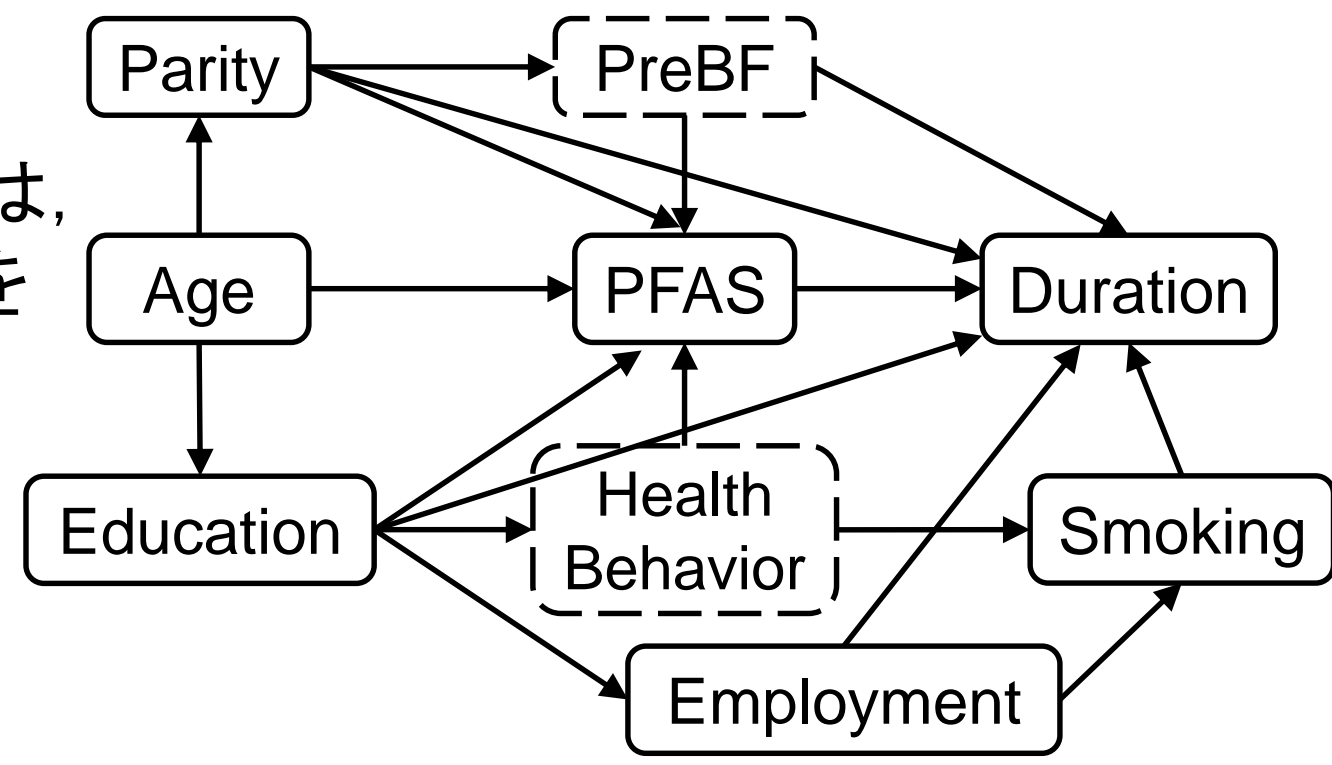
UNMEASUREDステートメントにて  
未観測変数を指定

# 未観測変数を含めた因果効果の識別

- SAS出力結果

NOTE: There are **no adjustment sets** satisfying the specified criteria for Figure 4

Figure 4 (右記の因果モデル)では、PFASのDurationへの因果効果を識別するために必要な変数が存在しません



PreBF, Health Behaviorは未観測変数

➡ METHOD = IV と指定し、条件付き操作変数を使用

# 条件付き操作変数

- SAS出力結果

表7. 条件付き操作変数

インストルメント変数: Figure 4

PFAS の因果効果 (Duration)

インストルメント

条件付き

Age

Education

Employment

Parity

Smoking

1 Age

\*

\*

\*

Education, Employment, Parityを与えたときの条件付き操作変数として Ageが存在する

- 総合効果の分解と推定
  - 直接効果と間接効果への分解
  - CAUSALMEDプロシジャによる解析
- 因果効果の識別可能性問題
  - 有向グラフを用いた因果効果の識別可能条件
  - CAUSALGRAPHプロシジャによる解析

# 参考文献

1. Efron, B. (1981). Nonparametric standard errors and confidence intervals. *The Canadian Journal of Statistics La revue canadienne de statistique* **9**: 139–172.
2. Efron, B. (1987). Better Bootstrap confidence intervals. *Journal of the American Statistical Association* **82**: 171–185.
3. Fei, C., McLaughlin, J., K., Lipworth, L., and Olsen, J. (2010). Maternal concentrations of perfluorooctanesulfonate (PFOS) and perfluorooctanoate (PFOA) and duration of breastfeeding. *Scand J Work Environ Health* **36**: 413–421.
4. Marjoribanks, K. *Environments for Learning*. London: National Foundation for Educational Research Publications. 1974.
5. Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika* **82**: 669–710.
6. Pearl, J. (2009). *Causality: Models, Reasoning, and Inference. The 2nd Edition*. Cambridge University Press. (黒木学記. (2009). 統計的因果推論—モデル・推論・推測—. 共立出版).
7. Rosenbaum P. R, and Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**: 41–55.
8. Rothman, K. J. (2012). *Epidemiology: An Introduction. The 2nd Edition*. Oxford University Press. (矢野栄二・橋本英樹・大脇和浩 訳. (2013). ロスマンの疫学—科学的思考への誘い. 篠原出版新社.)
9. SAS Institute Inc. (2020). *SAS/STAT® 15.2 User's Guide*. SAS Institute Inc., Cary, NC.
10. Shpitser, I., VanderWeele, T., and Robins, J., M. (2010). “On the Validity of Covariate Adjustment for Estimating Causal Effects”. *In Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, 527–536. Corvallis, OR: AUAI Press.
11. Timmermann, C., A., G., Budtz-Jørgensen, E., Petersen, M., S., Weihe, P., Steuerwald, U., Nielsen, F., Jensen, T. K., and Grandjean, P. (2017). Shorter duration of breastfeeding at elevated exposures to perfluoroalkyl substances. *Reproductive Toxicology* **68**: 164–170.

12. Van der Zander, B., Li´skiewicz, M., and Textor, J. (2014). “Constructing Separators and Adjustment Sets in Ancestral Graphs.” In *Proceedings of the Thirtieth Conference on Causal Inference: Learning and Prediction*, 11–24. Corvallis, OR: AUAI Press.
13. VanderWeele, T. J., and Vansteelandt, S. (2009). Conceptual issues concerning mediation, interventions and compositions. *Statistics and Its Interface* **2**: 457–468.
14. VanderWeele, T. J. (2013a). Policy-relevant proportions for direct effects. *Epidemiology* **24**: 175–176.
15. VanderWeele, T. J. (2013b). A three-way decomposition of a total effect into direct, indirect, and interactive effects. *Epidemiology* **24**: 224–232.
16. VanderWeele, T. J. (2014). A unification of mediation and interaction: a 4-way decomposition. *Epidemiology* **5**: 749–761.
17. 魚住龍史・矢田真城・山本倫生・川口淳. (2017). SASによる傾向スコアマッチング. SASユーザー総会論文集, 263–281.
18. 黒木学 (2017). 構造的因果モデルの基礎. 共立出版.
19. 田栗正隆 (2014). 直接効果・間接効果の推定および未測定の間接効果に対する感度解析. *統計数理* **62**: 59–75.
20. 星野崇宏 (2019). 調査観察データの統計科学 —因果推論・選択バイアス・データ融合. 岩波書店.
21. 松山裕(2018). 新版医学統計学ハンドブック18. 因果推論. 522–546, 朝倉書店.
22. 矢田真城・魚住龍史 (2019). CAUSALMEDプロシジャによる媒介分析. SASユーザー総会論文集, 29–50.
23. 矢田真城・魚住龍史 (2019). CAUSALGRAPHプロシジャによる因果効果の識別. SASユーザー総会論文集, 51–72.
24. 矢田真城・魚住龍史・田栗正隆 (2020). 反事実モデルに基づく直接効果と間接効果の推定. *計量生物学* **40**: 81–116.
25. 矢田真城・浜田知久馬 (2017). EMBアルゴリズムを用いたMultiple Imputation. SASユーザー総会論文集, 198–208.



ご清聴頂き，誠にありがとうございました。

