

# 生命保険数理における医療データ活用 - 因果推論手法の導入

PwC Japan 有限責任監査法人\*) 板倉 兼介 君

PwC Japan 有限責任監査法人\*) 尾上 辰徳 君



\*) 2023年12月、PwC あらた有限責任監査法人とPwC 京都監査法人が合併し、PwC Japan 有限責任監査法人に法人名称変更。

司会 それでは、定刻になりましたので、セッションA-2、「生命保険数理における医療データ活用 - 因果推論手法の導入」を開始させていただきます。本セッションの司会を担当させていただきます明治安田生命の清水です。よろしくお願いいたします。

本セッションは、PwC あらた有限責任監査法人の板倉兼介さんと尾上辰徳さんより、保険数理において因果推論を導入する試みについて、医療データ利活用の文脈でお話しいたします。板倉さんは、CERAであり、PwC 保険アシュアランス部門のディレクターとして、保険会社・公共クライアントの医療データの利活用支援に注力しております。尾上さんは、板倉さんのチームにおいて、データサイエンティストとして活躍しております。発表後に、時間の許す限り、皆様からいただいた質問に答える形にしていきたいと思っております。会場とスライドから質問を取り上げていきますので、スライドで質問する場合には、質問の記入をお願いいたします。

それでは、よろしくお願いいたします。



## アジェンダ

- 1 ナイーブな比較の問題
- 2 統計的因果推論
- 3 ユースケース
- 4 本日のまとめ
- 5 Appendix

2

板倉 ありがとうございます。皆様、おはようございます。本日は、沢山の皆様にお集まりいただき、われわれの取り組みに興味を持っていただき、ありがとうございます。コロナ禍が明けて、このように対面でお話できることを非常に楽しみにしておりました。対面ということもあって非常に緊張しておりますので、お手柔らかにというところでございます。

本日は、データサイエンスを活用する他業界・保険業界外の場所で既に一般的に用いられている手法ではあるのですが、保険業界では広くは浸透していないと思われる因果推論をご紹介します。保険業界、特に生命保険ビジネスにおいても活用できるのではないかと、というお話をさせていただきます。機械学習を用いる手法もあるのですが、今回は、結果が安定しやすく説明力が高い、統計的な手法に終始させていただきます。前半・後半ございますが、ご質問は最後にお受けいたしますので、よろしくお願いいたします。

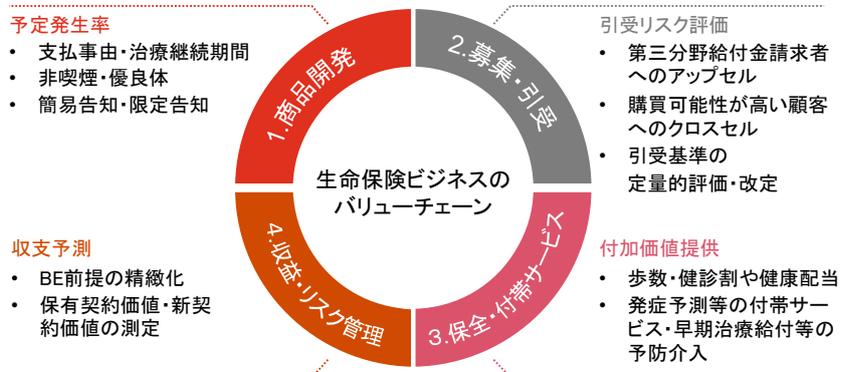
こちらが、本日のアジェンダです。まずセッション1では、医療データを分析する上で共通の課題を特定してまいります。セッション2では、統計的因果推論とは何かということ、初心者向けにご紹介したいと思います。3番めのユースケースでは、分析の設計から結果の考察までを実践的に行っているユースケースを、尾上の方からご紹介させていただきます。最後にセッション4にて、生命保険の実務において活用することを念頭に置いて、振り返りを行いたいと思います。お時間の関係でご紹介できない詳細部分が非常に沢山ございますので、その分はAppendixに記載しておりますので、ご興味のある方は、適宜見ながら聞いていただければと思います。

# 1

## ナイーブな比較の問題

### 医療データ活用機会

医療データを活用できる場面は増えている<sup>(\*)</sup>。実際の場面では特定のトリガーが発生率水準等に与える影響を観察データから抽出し、意思決定を促進する分析が必要



PwC <sup>(\*)</sup>: 網羅的に列挙する事を意図していない

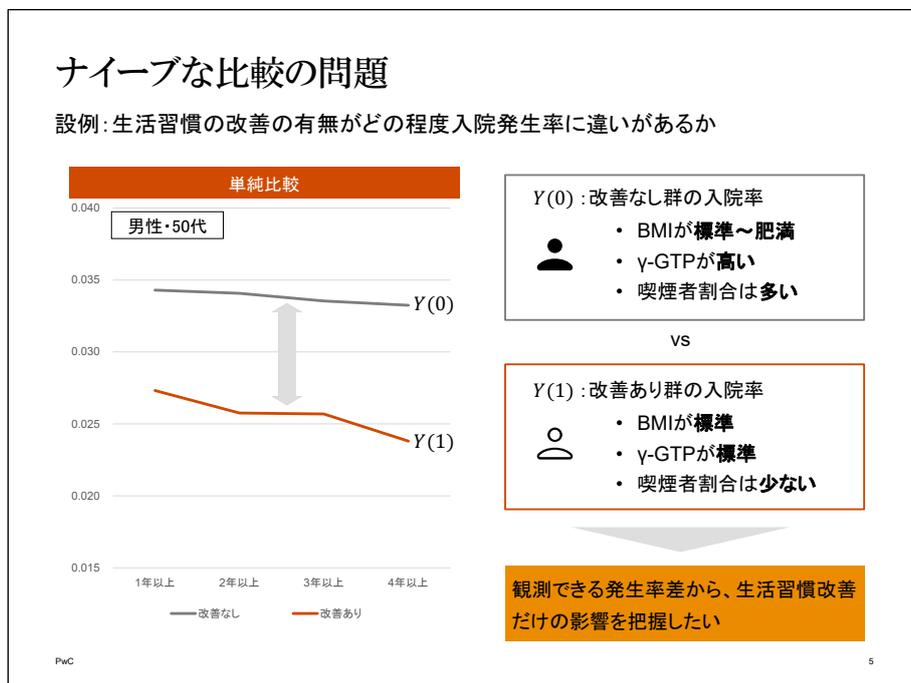
4

それでは最初に、ナイーブな比較の問題を特定していきたいと思います。

今、ヘルスケアの保健とインシュアランスの保険の共通部分が日に日に増えている状況ではございますが、それに伴って医療データ活用という場面は広がっているということ、生命保険ビジネスのバリューチェーンに沿って示したものになっています。

1 番目の商品開発から、2 の募集・引受、3 の保全・付帯サービス、4 の収益・リスク管理と、これらがお互いにオーバーラップしながらビジネスサイクルが回っていくようなイメージですが、代表的な使い方としては、ここで挙げられるような予定発生率の算出の基礎になるケースや、アンダーライティング基準の見直しの定量評価、付帯サービス導入の効果測定などにも有用になってきます。それぞれの場面では、公的統計

の発生率水準や、自社の標準体と比較してどの程度支払い率が高いかなど、基準群Aと、念頭にある群@  
 団Bとを比較することで、何らかの示唆を得ることが多いのではないかと思います。



では、どこが問題なのかということですが、例として、生活習慣の改善で入院発生率にどの程度違いがあるかを比較することを考えていきます。分からない記号などもあると思いますが、一旦、脇に置いておいていただき、後のユースケースでご紹介しますので、グレーの線とオレンジの線の差にご注目いただければと思います。

こちらのグレーの線は、生活習慣改善ありの群団Aであり、オレンジの線は改善なしの群団Bの入院発生率を示しています。生活習慣の改善が、どの程度入院発生率の差に影響を与えたのかということに興味があるとしましょう。実際にグレーの群団とオレンジの群団は、BMI、γ-GTP、喫煙習慣など、体況に当たるものが実際に悪いということが分かると思います。実際に発生率の差には、これら複数の要素の影響が入り込んでいて、生活習慣改善だけの影響ではないことが想定されますが、そのときに、この差は本当に見るべき差なのかというところが、問題意識のスタート地点です。生活習慣の改善に注目して、その要因を取り出したいということが、本日ご紹介する因果推論になります。

# 2 統計的因果推論

では、因果推論とは何かというところに入っていきます。

### 因果推論のフレームワーク

原因が結果に与える影響を、要因に変化を与える事で推計する

- 介入: ある要因を変化させて結果へ影響を与えること
- 因果関係・効果: 介入により結果変数Yが変化する関係・その強さ

確率変数名	記号	例
共変量ベクトル	$\vec{X} = (X_1, X_2, X_3, \dots)$	年齢, 性別, BMI, 血圧, 血糖値, ...などが成分
介入	$Z \in \{0,1\}$	投薬, 手術・治療, 喫煙・飲酒習慣の有無
アウトカム 結果変数	$Y(0), Y(1)$	入院・手術等の支払事由発生の有無、率

**p5の例**

全体N人 ( $i=1, 2, \dots, N$ )

生活習慣の改善

介入群  $Z_i = 1$   $E[Y(1)]$

対照群  $Z_i = 0$   $E[Y(0)]$

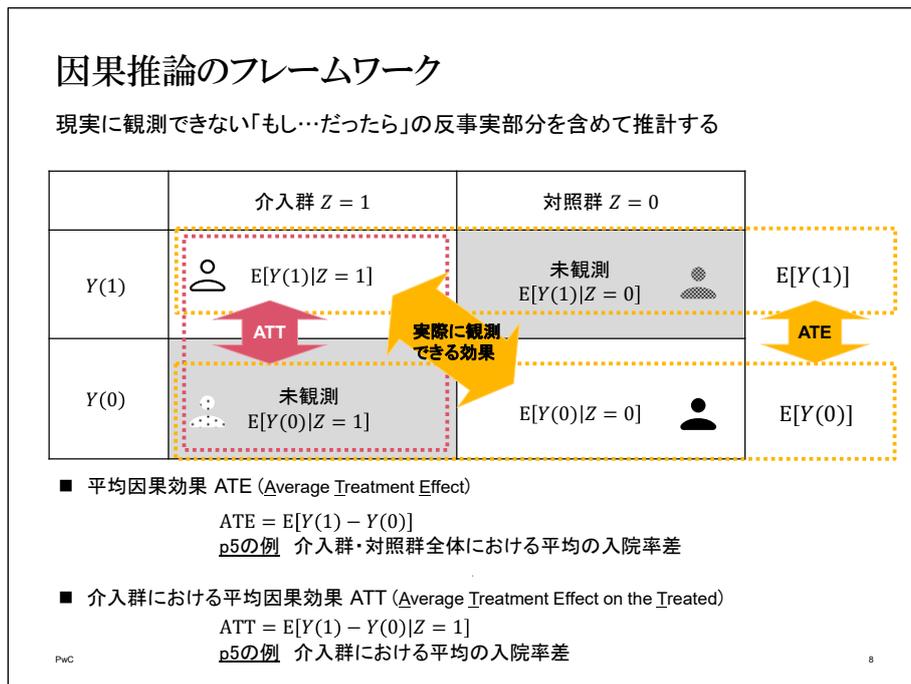
因果効果

因果推論は、文学的な表現でいくと、「原因が結果に与える影響を、要因に変化を与えることで推計する」ということになります。言葉の定義ですが、「介入」は、ある要因に変化を与えることとして定義します。一方で「因果関係」および「効果」は、介入によって結果変数が変化する関係および、その強さと定義することにします。

以下、これらの概念を数学的に記述していかなければいけないので、記号を定めたいと思います。まず、要因の集合を共変量と定めます。こちらは共変量ベクトル、ベクトル値で、主にここに記載されているような年齢、性別、BMI、血圧、血糖値などが成分になってまいります。次に介入ですが、先ほど言葉でご説明し

た介入を  $Z$  として、0、1の2変数とします。これは、例えば投薬や手術・治療、喫煙・飲酒習慣の有無などをイメージしていただければと思います。最後にアウトカム変数ですが、介入に応じて、 $Y(0)$ 、 $Y(1)$  という記号を定めます。こちらは、支払事由の発生や、その率のようなものをイメージしていただければと思います。個人単位で見ると、 $Y(0)$ 、 $Y(1)$  と書いたのですが、片方しか観測できないということに注意していただければと思います。

先ほどの例でいきますと、具体的にイメージしていただければと思いますが、全体  $N$  人に対して、観察データにおいて生活習慣を改善した人は、 $Z$  が1で、「介入群」と呼ばれる群に分類されます。片方の介入されていない人たちを、「対照群」と名前をつけます。期待値ベースで、この群団間の両者の差を因果効果であると定義します。

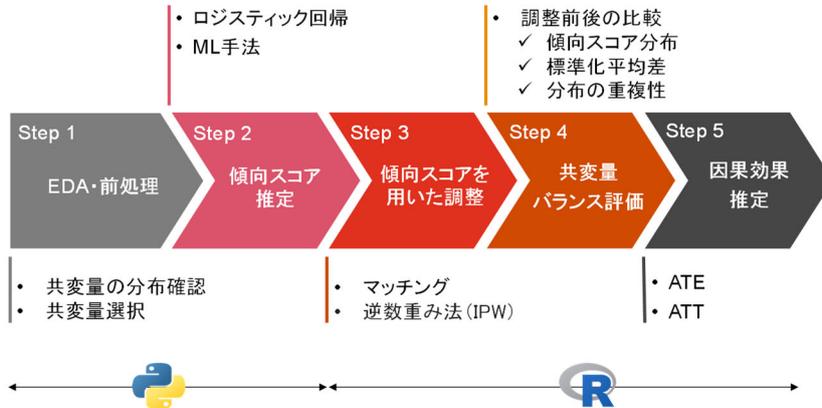


介入有無によってアウトカム変数が二つ出てきたということを思い出していただきたいのですが、実際には観測されないグレーな部分が二つ出てきてしまっている。こちらは、白い部分が実際に観測できる部分で、ナイーブに比較した場合は、ここの条件付き期待値の差が、見られる所になってきます。一方で、本当に見たい所はどこかというところ、ATE と書いている、全ての人を含めた期待値の差になっていまして、これを平均の因果効果と呼んで定義することにします。一方で、介入した人たちだけに注目したいときもございまして、これを ATT と呼んで、 $Z$  が1に条件を付けたときの条件付き期待値の差と定めることにします。

今日のご説明を割愛しますが、グレーの部分が発生しても因果効果を推定できるように、2群への割り振りをランダムに取ってデータを収集するという、RCT というデータ収集法方法もあるのですが、今日は、観察データを基に因果効果を推定するというアプローチをご説明いたします。

## 傾向スコアを用いた因果効果の推定手順

一般的な予測モデリング手法の流れと同様だが、Step2とStep5の2段階の推計を行うのが特徴的



PMC 分析環境についてはAppendix p50を参照

9

それでは、今回われわれが行った、傾向スコアというツールを導入して因果効果を推定する分析手順をご説明いたします。一般的な予測モデリングと同じように、EDA や前処理というところから始まっていくのですが、データセットから分析の対象である二つの群を抽出して、それを探索するというのがステップ1です。ステップ2は、因果推論の鍵になる傾向スコアをモデリングするステップです。ロジスティック回帰やマシンラーニング手法と書かれていますけれども、一般的にロジスティック回帰を使うことがよくあるということで、今回はロジスティック回帰を使っています。

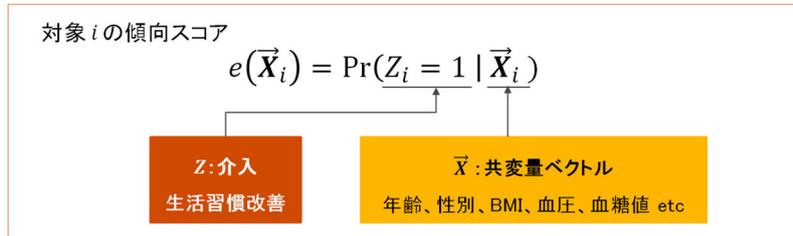
ステップ3ですが、傾向スコアを用いて調整するということをやるのですけれども、知りたい効果に応じて共変量を調整したり、マッチングと呼ばれる手法を取って、マッチする相手を選んでくるというようなことをやります。こちらに関しては、また後ほど詳しくご説明します。ステップ4では、調整前後の共変量を比較して、共変量の調整がきちんとできているということ、これらを見て確認する。そして、最後に因果効果を推定するという流れになっています。右上に、今、どこのステップにいるのかということをお示ししておきますので、どこにいるか迷ったら、それを見ていただければと思います。

われわれが取ったモデリング手法としては、全てPythonで完結することもできるのですが、Scratchから作らなければいけないということと、傾向スコアによる因果推論はRのライブラリーが充実していますので、それを使えるように、ステップ2までをPythonとしています。

## Step2: 傾向スコア



与えられた共変量ベクトルに対して介入が割り当てられる確率



- 傾向スコアを推定するモデルは、ロジスティック回帰を用いることが多い  
目的変数:  $Z$ , 説明変数:  $\vec{X}$  として、 $e(\vec{X}) = \Pr(Z = 1 | \vec{X})$  より

$$\ln\left(\frac{e(\vec{X})}{1 - e(\vec{X})}\right) = \beta_0 + \vec{\beta} \cdot \vec{X}$$
$$e(\vec{X}) = \frac{1}{1 + \exp\{-\beta_0 + \vec{\beta} \cdot \vec{X}\}}$$

$\beta_0$ : 切片,  $\vec{\beta}$ : 偏回帰係数ベクトル

- 真の傾向スコア  $e(\vec{X})$  に対して、推定後の傾向スコアを  $\hat{e}(\vec{X})$  と表記する
- 対象  $i$  の観察データ  $\{\vec{x}_i, y_i, z_i\}$  に対して、 $\hat{e}_i = \hat{e}(\vec{x}_i)$  とかく

PWC

10

それでは、ステップ2の傾向スコアですが、これが今回のお話の肝になるところで、それがどのようなワークするのかということが一番重要なところですので、こちらと次のページでご説明します。傾向スコアは、こちらの算式のとおり、与えられた共変量ベクトルを固定したときに、介入  $Z$  が割り当てられる確率として定義します。この心は、そのあとに出てくるのですが、多変量ベクトルである共変量を各成分ごとに調整することは一般に非常に難しいですが、次元の傾向スコアに沿って調整することが可能になる。そのような目的で、これを定義しています。

まだピンとこないと思うので、話を少し進めますが、観察データから傾向スコアをどのように推定するかというところを、もう少し具体的にご説明します。先ほどロジスティック回帰を使うとお話ししましたが、傾向スコアは確率ですので、 $\log p \div q$  という加工が線形回帰できるというような最初の式の形に書き換えて、 $e(X)$  に関して解いてあげると、シグモイド関数にこれを入れたような格好になります。

これを、確率なので、 $i=1$  から  $N$  までの観察データと、介入がそれぞれ入っていますので、その観察データから尤度比関数だと思って全部掛けてあげて、それを最尤法してあげると係数が定まります。ロジスティック回帰は結局それをやっているわけですが、それによって、真のスコアが分からなくても、観察データセットからこれを推測できる。そのようなことで傾向スコアを推定します。それを  $e$  ハットと書くことにします。先ほど観察データに対して定まるとお話ししましたが、それを対象  $i$  に対して決めたということ、 $e_i$  ハットと書くことにします。

## Step2: 傾向スコア



与えられた共変量ベクトルに対して介入が割り当てられる確率

- 「条件付き独立の仮定<sup>(\*)</sup>」の下で、傾向スコア定理により、ATE, ATTを推計できる<sup>(\*\*)</sup>

$$(Y(0), Y(1)) \perp Z | \vec{X}, 0 < e(\vec{X}) < 1 \text{ for } \forall \vec{X}$$

$$ATE \sim \frac{\sum_{i=1}^N \frac{Z_i}{\hat{e}_i} \times y_i(1)}{\sum_{i=1}^N \frac{Z_i}{\hat{e}_i}} - \frac{\sum_{i=1}^N \frac{1-Z_i}{1-\hat{e}_i} \times y_i(0)}{\sum_{i=1}^N \frac{1-Z_i}{1-\hat{e}_i}}$$

$$ATT \sim \frac{1}{|M|} \sum_{(i, m_i^c)} \{y_i(1) - y_{m_i^c}(0)\}$$

$M = \{(i, m_i^c) \in N_{\leq N} \times N_{\leq N} | m_i^c: \text{介入群対象 } i \text{ にマッチした対照群対象の番号}\}$

(\*): 強く無視できる割当 (Strongly Ignorable Treatment Assignment) 条件と呼ばれることが多い

PWC (\*\*): Rosenbaum & Rubinによる。計算詳細はAppendix p41~44を参照

11

先ほど、次元で足し上げることができるということがワークするとお伝えしたのですが、「条件付き独立の仮定」というものが非常に重要です。これは何を言っているかという、アウトカムの  $Y(0)$ 、 $Y(1)$  が、介入と条件を付けたときに、共変量が条件づけられたときに独立だという条件になっています。また、ここに0、1を取らないという条件も入っていますが、これがあることによって次元に沿って足し上げることができます。これが Rosenbaum / Rubin による大定理なのですから、それによって、具体的にこのご利益、これで推定することができるということですね。

少しごつく見えるのですが、実はそうではなくて、 $e_i$  ハット、先ほど推定した傾向スコアと、 $Z_i$  は0か1なので、1です。介入群に対して1で、対照群に対して1だというような、それぞれのこの式上の0、1というところなので、要は傾向スコアで逆数を取って加重平均を取っているという、それだけのことなので、その差で計算できてしまうということが、すごいところかと思えます。

ATTの方を見ますと、今度はもっと簡単で、アウトカム変数の実現値ですね。観測値を引き算してあげるといものの平均を取ってあげること、計算できてしまう。その取り方は少し工夫していますが、介入群の対象  $i$  に対して、 $m_i^c$  とマッチする人を対照群から選んでくるということで引き算を取っています。

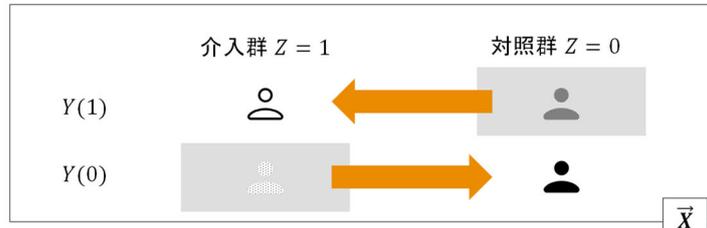
先ほどの傾向スコア定理というところは、Appendixに記載していますので、証明が気になる方は、見ていただければと思います。極力、セルフコンテインドになるように書かせていただきました。

## Step2: 条件付き独立の仮定



条件付き独立の仮定が成立している状況を検討する

例:  $\vec{X} = (50\text{歳, 男性, BMI, 血圧, ...})$  を固定



条件の確認方法:

介入 $Z$ を共変量 $\vec{X}$ で説明できているか

傾向スコア $e_i$ により、介入 $Z$ の予測精度を確認(AUC, Appendix p45参照)

共変量 $\vec{X}$ の分布はバランスしているか

2群間で共変量の分布の差がないことを確認(p15, Step4参照)

PwC

12

では、先ほどの条件付き独立の仮定が重要だとお伝えしたのですが、それがどのように成立しているのか、具体的に見たいと思います。例えば $X$ が、50歳男性、BMIいくつ、血圧いくつという、ご自身の状況をイメージしていただければと思いますが、自分が生活習慣を改善していない状況のときに、元々改善していた人と同じような入院発生率になることを認められるかという状況です。生命保険のよくある文脈では、このようなものは同じだと認めることになっていると思われると思いますが、 $X$ によっては、そのようなことは無い。また、介入の設定によってはそのようなことはないということは、後ほどユースケースでご説明します。

この条件をどのように確かめるのかということですが、実は直接確かめる方法がなく、間接的な方法が二つ提示されています。まず一つめが、介入 $Z$ を共変量で説明できるかということで、具体的に何を言っているかということ、傾向スコア $e_i$ が $i$ から $N$ まであります。介入も1から $N$ まであります。その二つのベクトルを並べてAUCを取ることで、傾向スコアの予測精度が十分であると確認します。先ほど、DeSCヘルスケアさんとミュンヘン再保さんのセッションでこのあたりはご説明されていたと思うので、定義などは、そちらを参照していただければと思います。

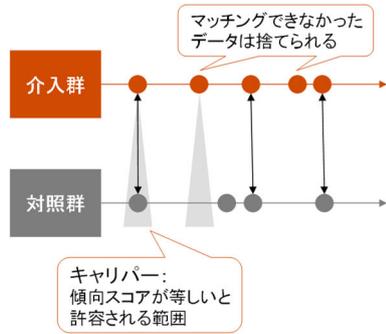
もう一つは、共変量の分布がバランスしているかということですが、傾向スコアを導入した結果、2群間で共変量の分布の差がなくなっているかを実際に確認するというステップで、経験的にそのようなものが成り立っていると、これが成り立っているということが観察できます。最近、こちらのAUCが高ければいいというものではないというお話がよく出ていて、こちらの方を重視している皆さんが多いので、本日はこちらを中心に説明いたします。

### Step3: 共変量調整～マッチング



2つの群で傾向スコアが等しいと見なせる対象者をマッチさせて、2群の差の平均を推定値とする方法

#### 1:1マッチングのイメージ



#### ■ ATTの推計

$$\frac{1}{|M|} \sum_{(i, m_i^c)} \{y_i(1) - y_{m_i^c}(0)\}$$

$M = \{(i, m_i^c) \in N_{=N} \times N_{=N} | m_i^c: \text{介入群対象 } i \text{ にマッチした対照群対象の番号}\}$

#### ■ メリット

- 直観的な理解が容易
- 因果推定が容易

#### ■ デメリット

- キャリパー設定に恣意性が入る
- 標本数の減少

PWC

13

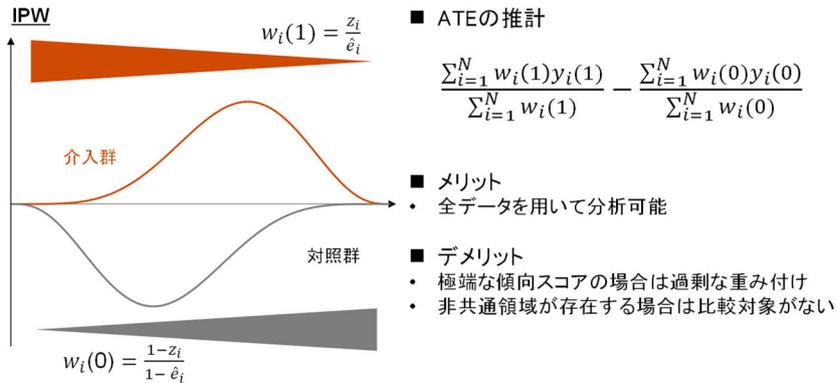
では、マッチングと IPW という方法があるとお話ししましたが、マッチングをご説明いたします。一言で申し上げますと、二つの群で傾向スコアが近い対象者をマッチさせて、2群の差の平均値を取るということを推定値とする方法です。傾向スコアが既に振られたイメージをいただいて、介入群・対照群それぞれに傾向スコアが小さい方から順に並べていっていただいて、似たスコアとなる対照のペアを取って、期待値の差を取っていくのですが、まず「近い」とは何かという話と、どこまでを近いと思うかという話があります。今回採用しているものは、ユークリッド距離で近さを取っていることと、傾向スコアが等しいという、キャリパーは分散の 0.2 倍まで設定しています。1対1にマッチするようなイメージの絵をしているのですが、このような感じです。

先ほど ATT を計算する式をご説明しましたが、このようなもので推計されるということで、非常に直感的に理解ができることと、因果推定が割と具体的にイメージできる。一方でデメリットとしては、マッチできなかったデータを捨ててしまうことや、キャリパーの設定で若干恣意的になってしまうところがございます。

### Step3: 共変量調整～IPW



傾向スコアの逆数による重み付け平均を用いて、因果効果の推定を行う方法



この方法を改善するという人も考えた人がいて、それが、こちらの IPW という手法です。まず、IPW とは何の略かという、逆確率重み付け手法。Inverse Probability Weighting の略です。傾向スコアの逆数で重み付けをして、因果効果を推定するという事です。先ほど  $e_i$  ハット分の  $Z_i$  だと書きましたが、それを重みとして定めてあげると、まさに重みに関する加重平均でその差を取るという格好になっています。これをやるとなぜいいのかというと、介入群では若干右に寄っていて、対照群では左に寄っているという絵になっていますが、この差がより左に寄って、より右に寄るといふ格好になりやすいので、きちんと apple-to-apple の比較になりやすいということなんです。

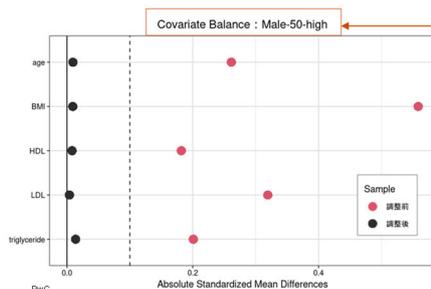
メリットとしては、先ほどのマッチングで落ちてしまうサンプルがなくなりますので、全てのデータセットを使って分析ができます。一方で、分布の状況によっては、極端な傾向スコアの場合、例えば 0 に集中している、1 に集中しているというようなケースでは、過剰に変なところに重み付けをしてしまったり、共通部分があまりないときに、本当にそれで比較していいのかということもございます。ですから、傾向スコアの分布をよく見なければいけません。

## Step4: 共変量のバランス評価



傾向スコアを用いて調整した後、2群の共変量のバランスが改善したかを評価する

50代男性	IPW調整前			IPW調整後		
	降圧薬使用 (n=5,746)	降圧薬未使用 (n=15,992)	標準化平均差 SMD	降圧薬使用 (n=21,420)	降圧薬未使用 (n=21,922)	標準化平均差 SMD
年齢(歳)	54.0	53.3	0.2575	53.5	53.5	0.0094
BMI	25.8	23.9	0.5132	24.5	24.5	0.0093
HDL	57.6	60.4	0.1850	59.3	59.5	0.0081
LDL	120.9	130.3	0.3266	127.4	127.5	0.0039
triglyceride	145.8	125.6	0.1844	132.8	131.5	0.0137



$$\text{標準化平均差} = \frac{|\bar{x}_{Z=1} - \bar{x}_{Z=0}|}{\sqrt{\frac{s_{Z=1}^2 + s_{Z=0}^2}{2}}}$$

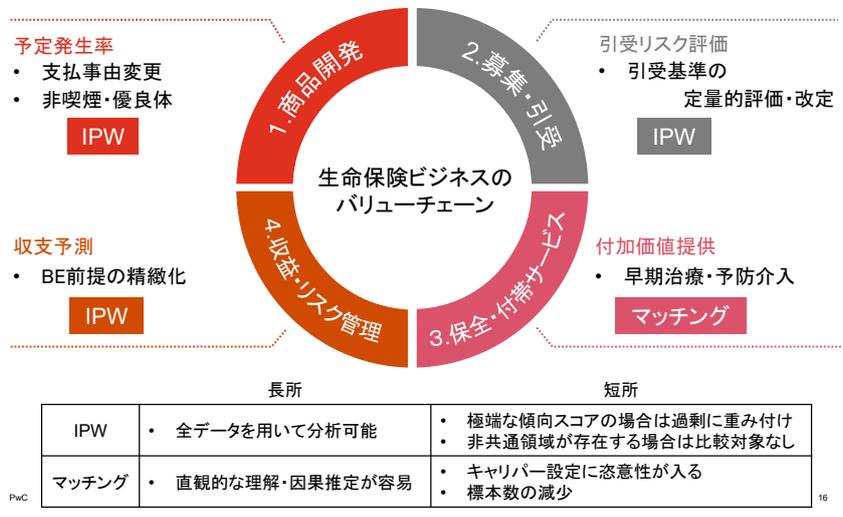
$\bar{x}_{Z=1}$  : 介入群の標本平均  $s_{Z=1}^2$  : 介入群の標本分散  
 $\bar{x}_{Z=0}$  : 対照群の標本平均  $s_{Z=0}^2$  : 対照群の標本分散

次のステップ4の共変量のバランス評価ですが、元々共変量がバランスしていないからナイーブな比較の問題だということだったので、実際に最後は調整されているということを見るために、標準化平均差というところをもって確認します。

算式上は、「介入群の標本平均」－「対照群の標本平均」を標本分散で割るとしているのですが、調整前でこのようになっていたものが、調整後にはこのようになっていると確認できるステップになります。例えばBMIを取っていただくと、0.5程度になっていたものが、調整したことにより、きちんと0.0093になっていることが、左下のグラフで確認できます。バランスができているということ、標準化平均差の0.1程度というところで判定することがよくありますので、われわれも今回はそれに倣っています。きちんと左側の点線の内側に入っていることを確認できます。

## 因果推論 (IPW / マッチング) の活用機会

全データを活用できるため、IPWを使う事が望ましい場面が多いが、マッチングも利用できる場面がある。どちらを使う場合も前提条件を満たすかの確認が重要



前半をまとめていきたいのですが、冒頭の医療データの活用機会ということで、実際に IPW やマッチングをどのように使っていけばいいのか、お示した表になっています。商品開発や引受査定などの文脈では、考慮しないサンプルがあるというケースが不適当になるケースも多く、予定発生率を作るときに想定される人が全員入っているのか、引受査定の際に公平性が大丈夫なのかというところですね。全員をきちんと考えなければいけないということで、IPW に軍配が上がる人が多いのですが、一方で、介入した対照の人のことを考えればいいのかというところで、例えば付帯サービスのケースですね。早期介入や予防介入でマッチングを使っていくということも、考えられるかと思います。

そこで、目的に応じて、どの効果を見るのか。ATE なのか、ATT なのかということも選択していくわけですが、いずれにせよ、プログラムを実行して結果を見るだけではなくて、共変量分布やモデル化後の傾向スコアをよく見て、因果推論の前提を満足するのかが検討することが、非常に重要な分析手法になっています。

では、前半は以上で、後半は、尾上より具体的なユースケースをご紹介します。お願いします。

# 3

## ユースケース

### 事例紹介

DeSCヘルスケア社提供のデータセットを用いて実施(\*)



#### <BE前提の設定>

- コロナ罹患者の未罹患者入院率に対する倍率は？  
～コロナ罹患者の入院率は、未罹患者対比で2.3倍以内～



#### <引受査定>

- 降圧薬使用を引受査定上の加点材料とするべきか？  
～降圧薬使用者の3CI広義死亡率は、未使用者対比で1.7倍以内～



#### <商品・付帯サービス開発>

- 糖尿病予備軍の人が生活習慣を改善した際の入院率低減度は？  
～生活習慣改善の継続期間に応じて20%以上低減～

PwC (\*)受領データの詳細はAppendix p47を参照

18

尾上 では、改めまして、PwCの尾上と申します。よろしくお願ひいたします。私から、具体的に今回取り上げた三つの事例をご紹介させていただきます。なお、本分析においては、DeSCヘルスケア社様提供のデータセットを用いて行っています。データの概要については、次のページでご説明させていただきます。

はじめに、このスライドにおいて、どのような場面を想定した事例なのか、その事例の内容と、どのような手法を用いたのかをご説明させていただき、最後に推定の結果は、このページでお話しさせていただければと思います。一つめの事例は、収支予測におけるベスト・エスティメイト前提の設定の活用を想定した事例です。介入をコロナの罹患として、コロナ罹患者の入院発生率が、未罹患者と比べてどの程度差があるのかを見た事例になっています。出ばなをくじくようですが、こちらの事例に関しては、傾向スコアを用いた

推定は行っておりません。理由は、後ほどご説明させていただきます。残る二つの事例については、傾向スコアを推定して調整を行った事例です。

二つめの事例ですけれども、引受査定場面を想定したもので、血圧を下げるために処方される降圧薬を、引受査定上考慮すべきか否かという検討事例になっています。この事例は、調整方法として IPW を用いています。理由としては、マッチングの場合、13 ページのイメージ図を思い出していただければと思うのですが、マッチングしなかった人、要するにアンマッチングの人を、分析対象から除いてしまうことになります。アンダーライティングの場面ですと、集団全体のリスク評価を行う必要があるため、今回のケースでは、集団全体の効果を推定できる IPW を採用しています。

三つめは、商品や付帯サービスの開発を想定した事例です。題材としては、生活習慣の改善と入院率の関係を取り上げました。生活習慣の改善を行うことによって、どの程度入院発生率を低減することができるかを見たものです。この事例では、調整方法として、もう一つのマッチングを採用しています。理由としては、生活習慣の改善に資する取り組みが、どの程度注目したいイベントに影響を与えているかを見たいためです。今回、見たい効果は、入院発生率としています。従って、取り組み自体の効果が見られればよいので、取り組みを行った人、すなわち介入を受けた人の効果を推定できるマッチングを採用しています。

では、改めてこのページで、推定の結果をお話しさせていただければと思います。一つめの結果は、全年齢において、コロナ罹患者の入院発生率は、未罹患者に比べて 2.3 倍以内でカバーできるという推定結果が得られました。二つめの結果は、定義は後ほどご説明しますが、降圧薬使用者の 3CI 広義死亡率が、未使用者に比べて 1.7 倍でカバーできるという結果が得られました。三つめの事例は、生活習慣の改善を行うことで、20%程度、入院発生率を低減することができるという結果が得られました。

## DeSCヘルスケアが保有するデータ



DeSCヘルスケアは以下3種類の保険制度にかかるレセプトデータを保有しています

- 健保(健康保険組合)
- 国保(国民健康保険)
- 後期高齢(後期高齢者医療制度)

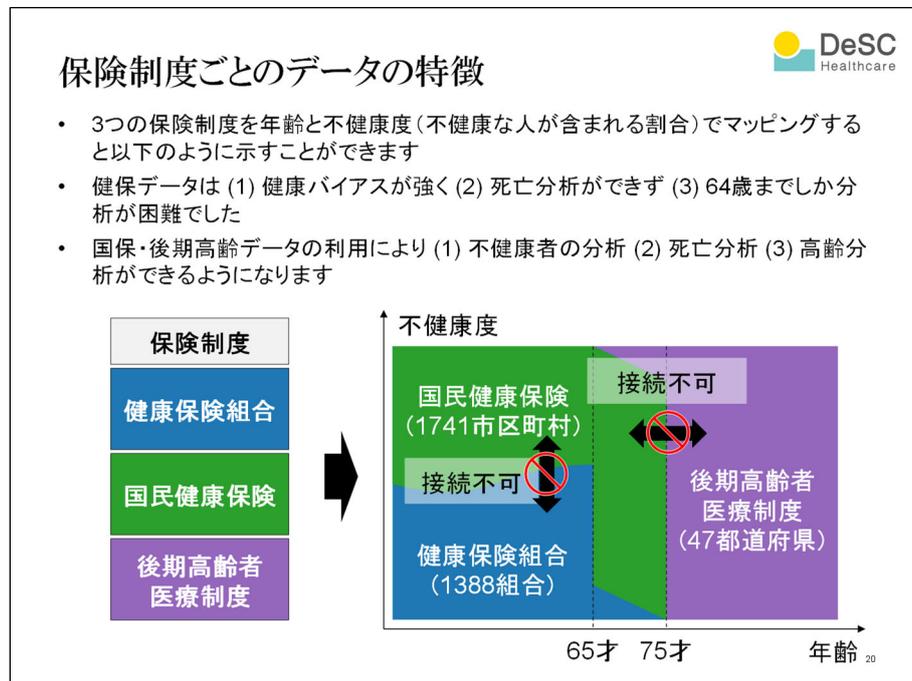
保険制度	保険者	被保険者
健康保険組合	企業	企業の従業員と家族 基本的に64歳以下
国民健康保険	市区町村	自営業者等企業に属さない者 74歳以下
後期高齢者 医療制度	都道府県	基本的に75歳以上 (65~74歳は一定の障がいあり)

※ DeSCヘルスケアが保有するデータの最新情報はDeSCヘルスケア会員制メディア (<https://www.desc-life-actuary.jp/>)にて確認できます(2023年11月1日公開予定)

19

では、詳細を説明させていただく前に、今回の推定に用いたデータに関して、2枚ほどご説明させていただければと思います。繰り返しになりますが、今回の分析に用いたデータは、DeSCヘルスケア社様提供の保険者制度別のレセプトデータになっています。対象となる保険制度は、健保、国保、後期高齢の三つです。個人 ID で紐づけることによって、傷病や、今回の例に挙げている降圧薬などの投薬といった、各種レセプト、

健診・問診データをひもづけて分析することができます。もちろん年度を与えた分析も可能になっています。データの最新情報については、スライド下部にございます URL よりご確認いただくことができますので、興味のある方は、ご確認いただければと思います。



加えて、このデータセットの特徴を2点ほど補足させていただきます。まず大前提として、健保のデータですと死亡の情報を含んでおりませんので、死亡の分析ができないということと、後期高齢部分の分析ができないという課題点があります。そのうえで、一つめの特徴として、国保のデータを用いることで死亡の分析が可能である点。2点めは、読んで字のごとくですが、後期高齢のデータを用いることで高齢部分の分析が可能であるということが、このデータセットの特徴です。われわれが今回受領したデータセットの詳細については、Appendix に掲載してございますので、ご興味のある方は、後ほどご確認いただければと思います。

## 事例1:コロナ罹患後入院

### ■ 命題

- コロナ罹患後の入院発生率は未罹患患者に対してどの程度上昇するか

### ■ キーワード

- 介入:コロナ罹患
- アウトカム:全入院



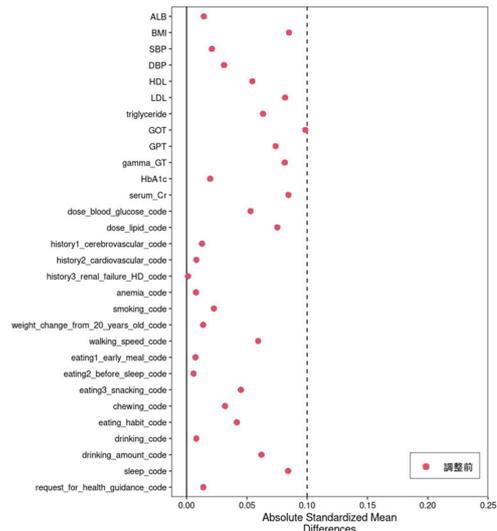
PwC

それでは、前置きが長くなってしまいましたが、具体的に事例の紹介に移りたいと思います。まず一つめの事例です。この事例では、コロナの罹患を介入とし、全入院をアウトカムとした推定を行っています。コロナの罹患が、入院にどの程度影響を及ぼすかを見た事例です。全入院とありますが、今回の集計では、いわゆる見なし入院は含んでおりません。その点をご留意いただければと思います。また、本セクションの冒頭で述べましたとおり、傾向スコアを推定して調整を行っていない事例になっています。その理由も含めて、詳細をご説明していきたいと思います。

## 事例1:標準化平均差



男性・50代



PwC

22

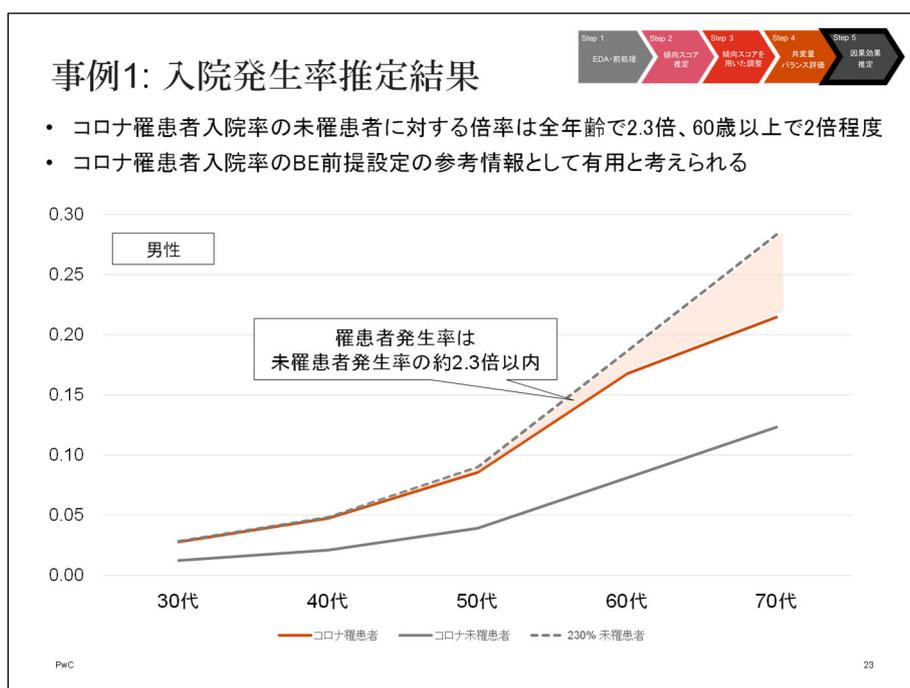
まず、先ほどの9ページの手順に沿って、傾向スコアを推定する前に、介入別、今回で言うとコロナの罹患有無別に共変量の分布を調べてみました。それがこちらになります。こちらは、標準化平均差をプロット

したものです。見方のおさらいですが、各プロットの項目が 0.1 よりも小さければ、2 群間において分布の偏りがないとされていました。上の方の項目が健診項目で、下が問診項目になっています。この問診項目の詳細については、Appendix に説明がございますので、適宜ご参照いただければと思います。また、比較した 2 群間の各数値に関しても Appendix に掲載してございますので、どの程度差があるのか、確認いただけたと思います。

見てお分かりのとおり、傾向スコアを推定する以前に、既に 2 群で分布の偏りが生じていないことが分かるかと思います。つまり、コロナの介入に対して、現在得られている健診・問診項目であれば、背景因子がそろっているということになります。ですから、傾向スコアを推定して調整する必要なく、単純に集計したコロナの罹患有無別の発生率の差が、介入における効果と見なせると考えられると思います。

ちなみに、この状態で仮に傾向スコアを推定すると、AUC は 0.56 と非常に低く、得られている健診・問診項目からではコロナの罹患が分類できないことも分かりました。直感的に、今回のコロナのケースで言うと、職業や居住地、また、コロナ禍におけるリモートなのか、在宅なのかといった勤務状況などを仮に得ていたとして、プロットすれば、0.1 よりも超えるのではないかと思います。また、それを踏まえて傾向スコアを推定すれば、もう少し精度は上がるのではないかと考えています。

そのように、観測できていない要素、未観測の因子というのが影響を受けていないかなどは、EDA や分析の設計時に確認が必要です。もちろん、その項目を含めて何か解釈していいのかということも、併せて確認が必要かと思います。

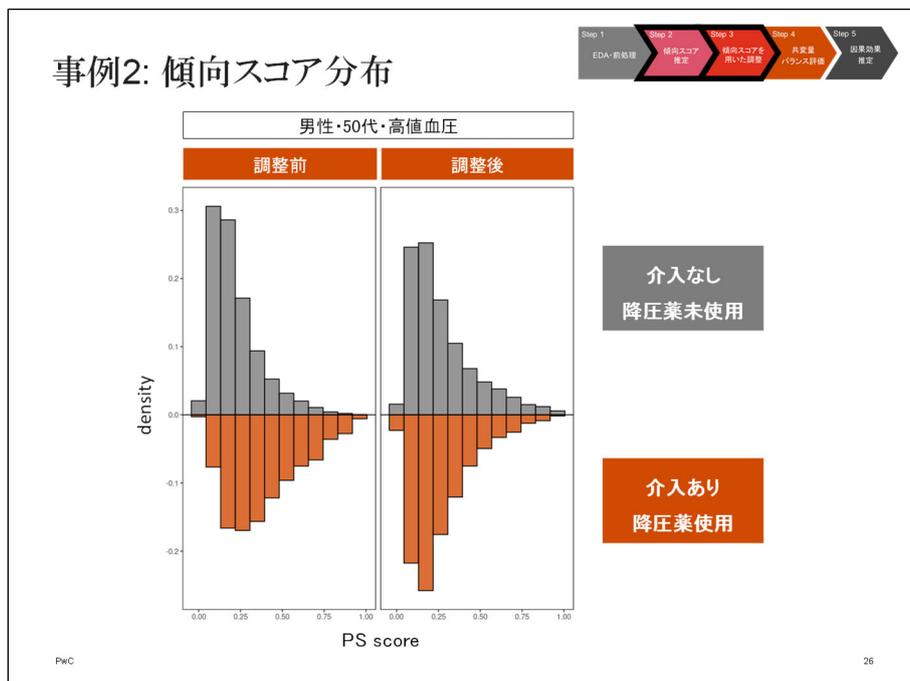


では、コロナ罹患有無別の入院発生率の集計結果です。灰色が未罹患患者で、オレンジ色が罹患患者を表しています。おおむね全年齢において、罹患患者の入院率は、未罹患患者に比べると 2.3 倍程度、特に 60 歳以上では、2 倍程度の差が確認できました。この結果を受けて、例えば未罹患患者の発生率をベースとして、罹患患者の発生率を 2.3 倍程度としたもの。こちらをベスト・エスティメイトの前提として用いるなどが、活用方法としては考えられると思います。以上が一つめの事例で、傾向スコアを用いないケースとしてご紹介させていただきました。



安定させるためにも、健保も分母に含めて集計を行っています。

二点めは、血压分類の定義です。今回、降圧薬の使用を考慮したあとの状態として、高値血压の人に限って分析を行っています。どのようなことかという、高血圧で降圧薬を使って高値血圧にいる人と、元々高値血圧にいて、降圧薬を使用せずに高値血圧に居続けている人たちを、同じリスクの評価、同じ評点として見なしてもよいのかという問いに対して、因果推論手法、今回は傾向スコアを用いて検討していくことになります。



具体的な結果の紹介に移りたいと思います。この事例の場合、EDA の時点で既に2群で偏りが生じていたので、スライド上では、傾向スコアのご紹介からになります。代表で、男性の50代における調整前後の介入有無別の傾向スコアを掲載しています。精度は0.72と、先ほどに比べて高くなりました。確認いただきたいことは、傾向スコアの分布の形状です。左側の調整前ですと、未使用者なので、傾向スコア、すなわち降圧薬を使用しているであろう確率が低いので、左側に分布の形状が偏っている様子が見て分かるかと思います。これが調整後ですと、同じような形状になっており、分布がそろっているといえるかと思います。

ただ、背景因子がそろっているかどうかに関しては、この結果からは分からないので、標準化平均差をプロットして確認していきたいというのが、次のページです。

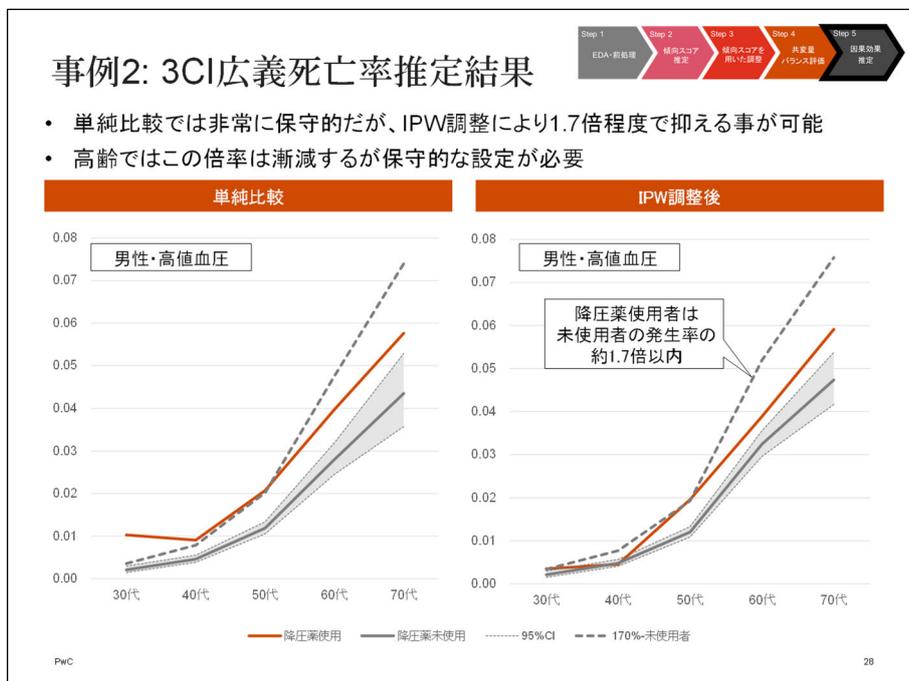
## 事例2: 標準化平均差



黒色のプロットが調整後で、コロナの事例から追加でプロットさせていただきました。健診項目に関しては、調整前の段階で2群の偏りが生じていることが確認できますが、調整後は、いずれの項目も0.1以内に収まっていることが見て分かると思います。これによって、2群において背景因子がそろっていることが確認できたので、このもとで推定結果を見ていきたいと思います。

## 事例2: 3CI広義死亡率推定結果

- 単純比較では非常に保守的だが、IPW調整により1.7倍程度で抑える事が可能
- 高齢ではこの倍率は漸減するが保守的な設定が必要



左のグラフが調整前、右が単純集計です。右のグラフが、IPWによる調整後のグラフになっています。灰色の実線が未使用者、オレンジの実線が使用者を表しています。調整前後で、使用者・未使用者の差、つまりATEは小さくなったものの、未使用者の95%信頼区間からは外れているので、アウトカムに対する降圧薬の使用の効果が確認できるかと思えます。年代によっても異なりますが、全体では、灰色の未使用者の発生率

をベースラインとして使用者の発生率を評価すると、約 1.7 倍でカバーできるという結果を得ました。  
年齢群によって倍率が異なる様子は、次のページでご説明させていただきます。

### 事例2: 3CI広義死亡率推定結果

男性 高値血圧	単純比較			IPW調整後		
	降圧薬 使用	降圧薬 未使用	倍率 <sup>(*)</sup>	降圧薬 使用	降圧薬 未使用	倍率 <sup>(*)</sup>
30代	0.0103	0.0021	4.8-4.9	0.0035	0.0021	1.6-1.7
40代	0.0091	0.0046	1.9-2.0	0.0046	0.0048	0.9-1.0
50代	0.0207	0.0119	1.7-1.8	0.0196	0.0120	1.6-1.7
60代	0.0397	0.0280	1.4-1.5	0.0389	0.0325	1.1-1.2
70代	0.0576	0.0435	1.3-1.4	0.0591	0.0474	1.2-1.3

PWC (\*)：倍率は降圧薬使用者 / 降圧薬未使用者として算出

29

こちらは、未使用者をベースラインとして、どの程度の水準で使用者の発生率をカバーできるのかということ、倍率という形で表したものです。左側の単純集計ベースに比べて、右側の傾向スコアを用いて影響を推定したものを確認すると、全体的に倍率が抑えられている様子が確認できるかと思えます。つまり、単純集計、いわゆる実態では、保守的な水準であったということが分かりました。加えて、もう1点の特徴としては、高年齢になるにつれて、倍率が漸減していく様子も確認できました。

活用のイメージとしては、引受査定において、評点の見直しへの活用。例えばですが、未使用者の評点を下げる、もしくは使用者の評点を上げる、もしくは到達年齢において評点の見直しを行うなど、方向性は幾つか考えられるかと思えますが、活用できるのではないかと考えています。ここまでが、二つめの事例の紹介になります。

### 事例3: 糖尿病と生活習慣

#### ■ 命題

- 糖尿病予備軍の人が生活習慣を改善したら入院率の低減割合はどの程度か

#### ■ キーワード

- 境界型糖尿病
- 介入: 生活習慣の継続的な改善
- アウトカム: 全入院

#### ■ 手法

- マッチング

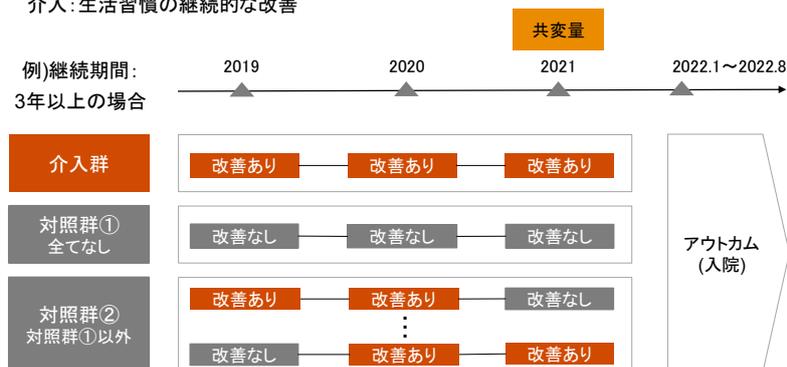


PwC

では、最後の事例の紹介になります。最後は、生活習慣の改善が、入院の発生率をどの程度低減することができるのかを見た事例です。対象となる集団としては、糖尿病の一步手前の、境界型糖尿病の人を対照にしています。理由は、生活習慣の改善の取り組みがある程度必要な層を対象とするべく、今回は境界型糖尿病を対象の集団として設定しました。

### 事例3: 介入の定義

- 共変量: 2021年に行われた健診・問診のデータ
- 介入: 生活習慣の継続的な改善



対照群②は以下の人を含む

- 改善あり→改善なし…途中でやめてしまった人
  - 改善なし→改善あり…対象とする年数を継続していない人
- ⇒以降「継続努力必要群」と呼称

PwC

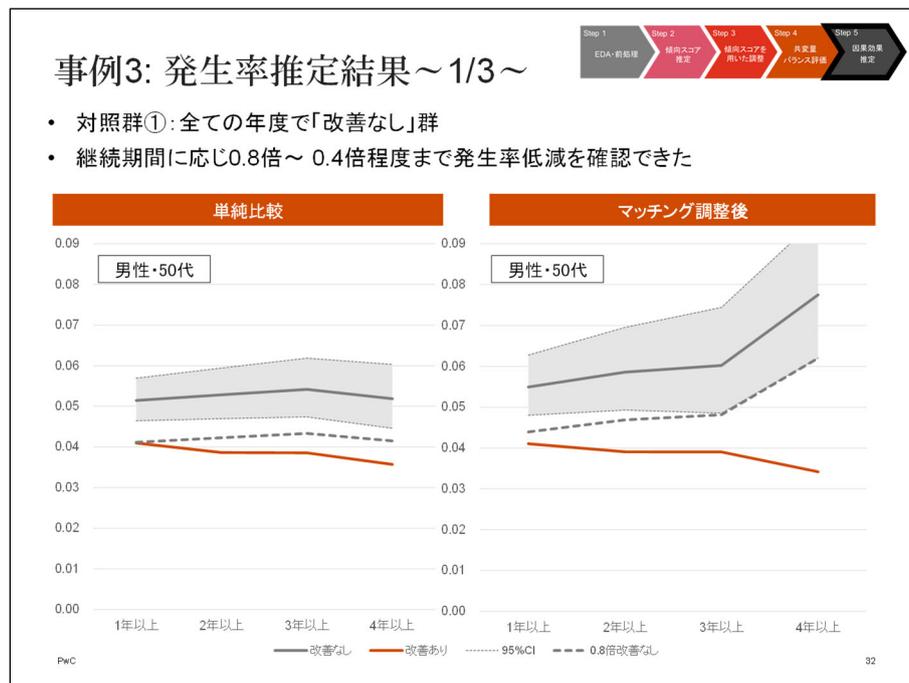
31

はじめに、この事例における介入の定義をご説明したいと思います。介入は、複数年度にわたって継続的な生活習慣の改善があることとしています。このスライドにおいては、継続期間が3年以上の場合を例に、記載しています。従って、介入群は、3年間全ての年度で改善ありの人たちの集団になります。

このケースでは、対照群を二つ設定してみました。一つは、介入群の真逆。逆に全ての年度で改善なしと、

いわゆる改善の意思がないような集団を対照群①として設定しています。もう一つ、対照群②として、①以外なのですけれども、2パターン対象としています。改善ありから改善なしのように、途中でやめてしまった人。もう1パターンとしては、改善なしから改善ありのような改善はしているのだけれども、今回のケースで言うと3年以上までは達していない。そのような人たちを、まとめて対照群②と設定しています。この集団のことを、「継続努力が必要な群」と呼ぶことといたします。

もう1点、アウトカムですが、データの都合上、1月から8月という8か月間の集計になっていますので、次にお見せする集計結果に関しては、年率に換算したものを掲載してございます。



では、具体的に結果の紹介に移りたいと思います。傾向スコアと標準化平均差に関しては、事例の二つめと同様の傾向を示していましたので、スライドからは除いて、Appendixの方に掲載してございます。ご興味のある方は、後ほどご確認いただければと思います。この事例では、推定結果からのご説明になります。

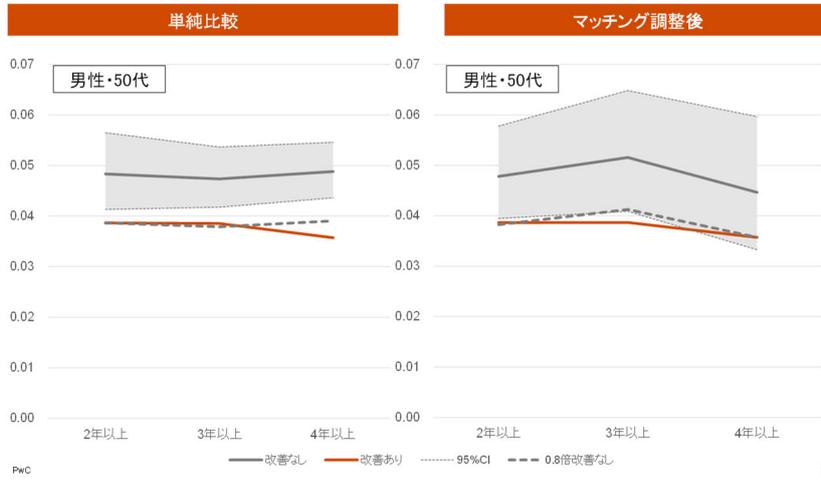
まず、対照群として、全ての年度で改善なしの集団との比較になります。単純比較において、介入群と対照群の差が確認できるかと思いますが、調整後は、より一層その差が広がっている様子が、このグラフからも見て分かるかと思いますが、つまり、生活習慣改善の継続的な取り組みによる効果が、単純な比較よりも大きいことが分かりました。また、継続期間が延びていくにつれて、どの程度低減できるかという倍率も、0.8倍から0.4倍程度まで低減していくことも分かりました。この倍率に関しては、次の対照群も併せて、2ページ先でご紹介させていただければと思います。

この結果から、活用のイメージとしては、ベースラインを全くやっていない人と設定した場合に、発生率水準は20%程度ディスカウントするなど、発生率水準の設定への活用が、一つ考えられるかと思いますが、また、付帯サービスの文脈で考えれば、継続期間が延びるにつれて介入による効果が見られるので、継続期間に応じたインセンティブを与える仕組みの導入など、施策導入の材料としても活用できるのではないかと考えています。

### 事例3: 発生率推定結果～2/3～



- 対照群②: 継続努力必要群
- 継続的な改善が必要だが、それでも0.8倍程度まで発生率低減を確認できた



では、続いて、対照群の二つめです。継続努力が必要な集団との比較を見てみようと思います。こちらは、先ほどとは異なり、2群の差は、95%信頼区間から大きく外れるほどには生じませんでした。それでも、点線が0.8倍ですが、0.8倍程度までは低減できることが確認できました。継続期間が延びるにつれて発生率が低減していくことに関しては、一つ前の対照群①と同じ傾向を示しています。この場合、継続期間によっては信頼区間に含まれてはいますが、差があるということは確認できていますので、どの程度ディスカウントするのかということは、考えどころかと思えます。ただ、付帯サービスの提供を想定しますと、このような層に対しては継続的にフォローしていく必要があるという、直感的なイメージと検証結果が一致したということになります。

### 事例3: 発生率推定結果～3/3～



マッチング調整後						
男性 50歳	対照群①: 全ての年度で「改善なし」群			対照群②: 継続努力必要群		
	改善なし	改善あり	倍率 <sup>(*)</sup>	改善なし	改善あり	倍率 <sup>(*)</sup>
1年以上	0.0366	0.0274	0.7-0.8	-	-	-
2年以上	0.0391	0.0260	0.6-0.7	0.0319	0.0258	0.7-0.8
3年以上	0.0401	0.0260	0.6-0.7	0.0344	0.0258	0.7-0.8
4年以上	0.0517	0.0228	0.3-0.4	0.0298	0.0238	0.7-0.8

PwC <sup>(\*)</sup>:倍率は改善あり/改善なしとして算出

では、最後に、二つの対照群における発生率の倍率を見てみたいと思います。紙面の都合上、ともにマッチング調整後の倍率を記載しています。改善なしをベースラインとして、どの程度低減が可能かを倍率という形で示しています。一つめの全て改善なし群との比較では、継続期間が3年までは20%程度低減、4年以上ですと60%程度低減できるという推定結果が得られました。継続する期間に応じて、インセンティブを調整するなどの対応が考えられるかと思います。

継続努力が必要な集団との比較では、期間を通じて約20%程度低減できることが確認できました。こちらは、繰り返しになりますけれども、集団の特性を考慮して、どのように、どの程度インセンティブを与えていくかということは、考える余地があるかと思います。以上までが、具体的なユースケースを想定した事例の紹介になります。

### 事例紹介～サマリー～



<BE前提の設定>

- コロナ罹患者の未罹患者入院率に対する倍率
- ✓ 60歳未満で2.3倍、60歳以上で2倍程度



<引受査定>

- 降圧薬使用を引受査定上の加点材料とするか
- ✓ 3CI広義死亡で到達年齢に応じた評点の見直しを検討可能



<商品・付帯サービス開発>

- 糖尿病予備軍の人が生活習慣を改善した時の入院率低減度合
- ✓ 対照群の健康意識に応じた入院率低減効果を認められる
- ✓ 介入施策導入の検討材料として利用可能

PWC
35

最後に、まとめとして事例の結論を振り返って、本セクションを終わりにしたいと思います。サマリーになります。一つめのコロナの事例では、コロナ罹患者の入院発生率は、未罹患者に比べて60歳未満で2.3倍程度、60歳以上で2倍程度に設定するなど、コロナ罹患者のベスト・エスティメイト前提の設定に有用であるというお話をさせていただきました。また、この事例では、傾向スコアを用いることが適さない事例として、理由も併せてご説明させていただきました。

残りの二つの事例につきましては、いずれも単純な比較では背景因子がそろっていないことによって、注目したい介入の影響を見ることができないケースのご紹介でした。二つめの降圧薬のケースですと、年齢に応じて効果があることが確認できましたので、到達年齢に応じた評点の見直しなど、検討材料に用いることができるのではないかとのお話をさせていただきました。

最後の生活習慣改善の事例では、対照群として二つ設定させていただきました。全く行っていない無関心層と、改善してきている、もしくは途中でやめてしまったというような、継続努力が必要な群の二つに分けて、効果を推定しました。効果の度合いは異なるものの、いずれの対照群の比較においても発生率の低減は確認できまして、施策導入の検討材料にこの結果が有用ではないかという話をさせていただきました。以上が、実際の保険ビジネスにおける活用を想定して、三つのユースケースとしてご紹介をさせていただきます。

した。

最後に、本日の全体のまとめを板倉よりお話しさせていただき、クローズとさせていただければと思います。



## まとめ

- 高齢群団の体況、死亡転帰など、利用可能な情報に広がりが増えた事により、保険会社で医療データを活用するシーンはさらに増加している
- 因果推論手法により、複合的なリスク要因において一つの要因を紐解くことで、医療データを更に有効活用する事が出来る
- 因果推論手法の前提条件や結果の妥当性については結果を出すだけでなく検証が重要。IPW / マッチングそれぞれで、共変量分布や傾向スコアなどを見てデータと向き合う必要があり、アクチュアリーに向いている
- 介入の効果測定にも分析結果を踏まえて意思決定を行う際には、結果の解釈を踏まえてステークホルダー間で調整する
- 医療データ利活用に限らず、因果効果の推定は保険会社のバリューチェーンで有用。より広い文脈で有効活用される事が期待できる

PwC

37

板倉 それでは、われわれの取り組みをまとめさせていただきます。まず、高齢群団の体況や死亡の転帰情報など、利用可能な情報は、今、広がっていき、保険会社で医療データを活用するシーンは、更に増えているという状況だと思います。その際に、今日ご紹介したような因果推論手法によって、複合的なリスク要因はございますが、その一つの要因をひもとくことで、医療データを更に有効活用できるのではというお話をいたしました。

一方で、因果推論手法は、いわゆるマジカルステッキではなくて、前提条件や結果の妥当性について試行を続けなければいけない、検証を続けなければいけないというところがございます。IPW、マッチングという手法をご紹介しましたが、共変量分布や傾向スコアの分布などを見て、データとしっかり向き合わなければいけない。これこそ、まさにアクチュアリーに向いているのではないかと考えています。

一方、介入の効果測定ということで分析結果をご提示したところではあるのですが、それを直接皆様のビジネスに使われる際には、保険者集団とも違うというところもありますので、結果をどのように解釈するか、ステークホルダー間で調整が必要なところでもございます。まさにそのようなところにも、アクチュアリーの仕事があるのではないかと思います。

今日は医療データ活用というお話をしたのですが、実はこの因果効果の推定は、医療データに限らず、保険会社のバリューチェーンの至る所で使える手法なのではないかと思っております。より広い文脈で有効活用してほしいと思っております。私からは以上です。本日は、ご清聴ありがとうございました。

司会 それでは、質問に入ります。会場で質問のある方は、挙手をお願いいたします。よろしいですか。

次に、スライドから質問を読み上げさせていただきます。まず一つめですが、「事例1について、職業等の未観測のデータ項目がコロナ罹患に影響を与えていないか、EDA等で確認が必要とおっしゃられたと理解したのですが、具体的には、どのようにすれば確認できるのでしょうか。直感的に不可能、難しいのではないかという気がしています」という質問が来ております。お願いします。

尾上 ありがとうございます。誤解を生じさせるコメントになってしまったかと思うのですが、そもそも分析を設計する際に、どのような要素がアウトカムや介入に影響を及ぼすかということは、事前にドメイン知識等を用いて設計が必要かと思えます。今回、紙面上、掲載はしていないのですが、感度分析という方法もございまして、未観測の因子の影響がどの程度あるのかを見る方法もございまして、そのような方法も加味して、未観測の因子の影響を見ることができると思えます。以上です。

司会 ありがとうございます。それでは、スライドからもう1問、質問が来ております。「マッチングについて、素朴な方法として、共変量の空間で距離を測って、最近傍にある非介入群の個体を取ってくる方法も考えられると感じました。傾向スコアベースでマッチングする方法の利点には、どのようなことがあるのでしょうか」。お願いします。

板倉 コメント頂きありがとうございます。共変量空間で距離を取る場合、冒頭に申し上げた比較上の課題で、両者の背景因子がそろっていないために、apple-to-appleの比較が難しくなることと思われまます。傾向スコアを導入するのは、まさにそれを解消するためにやっています。こちらでお答えになっていますでしょうか。以上です。

司会 ありがとうございます。それでは、もう1問お願いいたします。「RWDの分析のためのデータについて、アカデミアや研究機関であれば、研究目的ということで入手しやすい面があると思いますが、民間企業で商品開発に利用するとなると、商用利用目的となり、簡単には入手しにくいものでもあると思います。健保、国保、後期高齢者のおのおのについて、どのようなデータを入手したのでしょうか」という質問でござい

ます。

板倉 ありがとうございます。今回は DeSC ヘルスケア社様からデータセットを提供いただいているのですが、われわれがお答えできるようなお話ではないのですが、同意を取っていると理解はしておりまして、詳細に関しては、あちらにお伺いいただければと思います。

司会 すみません。答えにくい質問で、ありがとうございました。それでは、もう1問お願いいたします。「本研究は、保険の観点ではなく、医学的研究結果としても意義があるように思います。医学との共同研究等の可能性について、お考えがあれば伺いたく存じます」という質問です。

板倉 ありがとうございます。われわれ PwC としては、国民医療データの分野でもご支援をしたりしておりまして、そのような文脈で医学的研究成果としてお話しすることは、可能性としてはあるかと思えます。ご提案いただいたとおり、今後、もし有用であるということであれば、結果を公にするということも、一つあるかと思えます。

司会 どうもありがとうございます。最後にもう1問お願いします。「対照群②を介入群として、対照群①と分析した結果はございますでしょうか。3年の生活習慣の改善は長いので、3年未満だと、全く改善しない人との程度異なるか、知りたいと思いました」というご質問です。

尾上 ありがとうございます。実際に対照群を二つ設定した背景としては、対照群どうしを比較して、全くやっていない人と、ある程度やっているのだけれども途中でやめてしまった人たちを比較することで、どのような層にアプローチすればよいのかも併せて結果として見られればということで、比較はしております。紙面上、Appendix にも今回は追加していませんが、この場で結果がすぐに思い出せないのが、回答できないのですが、比較自体はしていますし、設計上、比較することは可能です。すみません、回答になっていないのですけれども。

司会 ありがとうございます。それでは、スライドの質問は、これで全てお答えいただきました。ありがとうございました。

有益なお話を、どうもありがとうございました。聴講している皆さんも、非常に参考になったのではないかと思います。少し早いですが、時間になりましたので、このセッションを終了したいと思います。ご清聴ありがとうございました。

<次ページ以降に参考文献と Appendix を掲載させていただきましたので、ご参照ください>

## 参考文献

### 原論文：

- [1] Rosenbaum, P.R. & Rubin, D.B. (1983), The central role of the propensity score in observational studies for causal effects, *Biometrika* (1983), 70, 1, pp. 41-55 <https://www.math.mcgill.ca/dstephens/SISCER2021/Articles/Rosenbaum-Rubin-Bka83.pdf>
- [2] Rubin, D.B. (1985). The Use of Propensity Score in Applied Bayesian Inference. In J.M. Bernardo, M.H. DeGroot, D.V. Lindley, & A.F.M. Smith (Eds.), *Bayesian Statistics 2* (pp.463-472). North-Holland/Elsevier Science Publisher B.V. <https://cir.nii.ac.jp/crid/1570854175568240000>
- [3] Rosenbaum, P.R. (1987) Model-Based Direct Adjustment. *Journal of the American Statistical Association*, 82, 387-394. <https://www.jstor.org/stable/2289440>

### 日本語書籍：

- [4] 星野 崇宏 (2009) 『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 (シリーズ確率と情報の科学)』, 岩波書店.
- [5] 安井 翔太 (2020) 『効果検証入門～正しい比較のための因果推論/計量経済学の基礎』, 技術評論社.

PwC

38

# 5

## Appendix

PwC

## Appendix

- ATE算出詳細
- AUCと傾向スコア分布の関係
- マッチングの設定
- 今回利用したデータの内容
- 問診項目一覧
- 分析環境
- 集計条件
- 事例2: 集計条件イメージ図
- 事例3: 傾向スコア分布
- 事例1: 標準化平均差
- 事例2: 標準化平均差
- 事例3: 標準化平均差
- 事例3: マッチング率

PwC

40

## ATE算出詳細([1]Rosenbaum & Rubin 1983)

### Definition 1: バランシングスコア

バランシングスコア $b(\bar{X})$ とは以下の条件を満たす共変量の関数である:  
 $Z \perp \bar{X} | b(\bar{X})$

### Theorem 1: 傾向スコアのバランシング性

傾向スコアはバランシングスコアである。つまり $Z \perp \bar{X} | e(\bar{X})$ を満たす

Proof

条件 $Z \perp \bar{X} | e(\bar{X})$ は $\Pr(Z = 1 | \bar{X}, e(\bar{X})) = \Pr(Z = 1 | e(\bar{X}))$ と同値であるためこれを示す。  
このために、両辺が $e(\bar{X})$ になることを示す。

左辺)

$$\begin{aligned}\Pr(Z = 1 | \bar{X}, e(\bar{X})) &= E[Z | \bar{X}, e(\bar{X})] \\ &= E[Z | \bar{X}] \quad (\because e(\bar{X}) \text{ は } \bar{X} \text{ の関数より}) \\ &= \Pr(Z = 1 | \bar{X}) \\ &= e(\bar{X}) \quad (\because \text{傾向スコアの定義})\end{aligned}$$

右辺)

$$\begin{aligned}\Pr(Z = 1 | e(\bar{X})) &= E[Z | e(\bar{X})] \\ &= E[E[Z | \bar{X}, e(\bar{X})] | e(\bar{X})] (\because \text{反復条件付け}) \\ &= E[e(\bar{X}) | e(\bar{X})] \quad (\because E[Z | \bar{X}, e(\bar{X})] = e(\bar{X})) \\ &= e(\bar{X})\end{aligned}$$

PwC

41

## ATE算出詳細([1]Rosenbaum & Rubin 1983)

### Theorem 2: バランシングスコアの粗性

すべてのバランシングスコア $b(\vec{X})$ に対して $f(b(\vec{X})) = e(\vec{X})$ を満たす関数 $f$ が存在する

#### Proof

背理法で示す。バランシングスコア $b(\vec{X})$ に対して $f(b(\vec{X})) = e(\vec{X})$ を満たす関数 $f$ が存在しないと仮定する。

このとき、 $b(\vec{X}_1) = b(\vec{X}_2)$ かつ $e(\vec{X}_1) \neq e(\vec{X}_2)$ を満たす共変量 $\vec{X}_1, \vec{X}_2$ が存在する。

一方、

$$e(\vec{X}_1) = E[Z | \vec{X}_1] = E[Z | b(\vec{X}_1), \vec{X}_1] = E[Z | b(\vec{X}_1)] = E[Z | b(\vec{X}_2)] = E[Z | \vec{X}_2] = e(\vec{X}_2)$$

であり $e(\vec{X}_1) \neq e(\vec{X}_2)$ に矛盾する。

PwC

42

## ATE算出詳細([1]Rosenbaum & Rubin 1983)

### Theorem 3: 傾向スコア定理

$(Y(0), Y(1)) \perp Z | \vec{X}$ かつ $0 < e(\vec{X}) < 1$  for  $\forall \vec{X}$  のとき、 $(Y(0), Y(1)) \perp Z | e(\vec{X})$ を満たす

**Proof**  $\Pr(Z = 1 | Y(0), Y(1), b(\vec{X})) = \Pr(Z = 1 | e(\vec{X}))$ を示す。

$$\begin{aligned} \Pr(Z = 1 | Y(0), Y(1), b(\vec{X})) &= E[Z | Y(0), Y(1), b(\vec{X})] \\ &= E[E[Z | Y(0), Y(1), \vec{X}, b(\vec{X})] | Y(0), Y(1), b(\vec{X})] \\ &= E[E[Z | \vec{X}, b(\vec{X})] | Y(0), Y(1), b(\vec{X})] && (\because \text{条件付き独立の仮定}) \\ &= E[E[Z | b(\vec{X})] | Y(0), Y(1), b(\vec{X})] && (\because \text{バランシングスコアの定義}) \\ &= E[Z | b(\vec{X})] && (\because \text{条件つき期待値}) \\ &= \Pr(Z = 1 | b(\vec{X})) \end{aligned}$$

Theorem 2より $b(\vec{X})$ を $e(\vec{X})$ としても上記の関係式は成立する。

PwC

43

## ATE算出詳細([1]Rosenbaum & Rubin 1983)

### Theorem 4

$$E[Y(1)|b(\bar{X}), Z = 1] - E[Y(0)|b(\bar{X}), Z = 0] = E[Y(1) - Y(0)|b(\bar{X})]$$

#### Proof

$$\begin{aligned} E[Y(1)|b(\bar{X}), Z = 1] - E[Y(0)|b(\bar{X}), Z = 0] &= E[Y(1)|b(\bar{X})] - E[Y(0)|b(\bar{X})] \quad (\because \text{Theorem 3}) \\ &= E[Y(1) - Y(0)|b(\bar{X})] \end{aligned}$$

となる。

さらに、これの期待値をとると

$$E[E[Y(1) - Y(0)|b(\bar{X})]] = E[Y(1) - Y(0)]$$

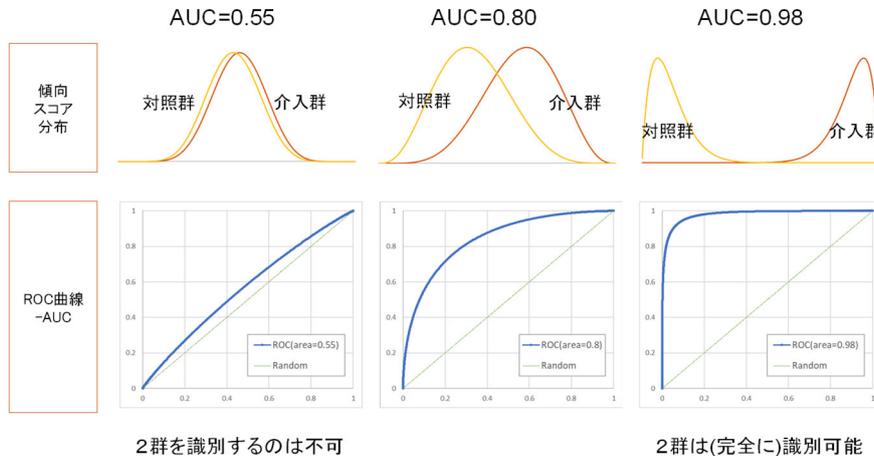
となり、ATEの不偏推定量となる。

PwC

44

## AUCと傾向スコア分布の関係

- AUCを用いて、共変量 $\bar{X}$ の条件の下で介入をどの程度識別できているかを確認する
- 傾向スコアの分布とAUCとの関係



PwC

45

## マッチングの設定

マッチング法を採用する場合には、キャリパーおよび構成比(マッチングの方法)を指定する必要

### ■ キャリパー(Caliper)の指定

キャリパー(Caliper): マッチングさせる許容領域のこと

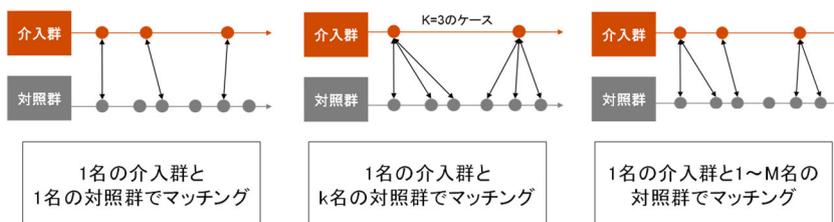
- 推奨値:  $SD \times 2.0 \sim 2.5$  をキャリパーに設定することが推奨されている  
(Rosenbaum & Rubin(1985), Austin PC(2011))
- 指定法: 一定の傾向スコアの距離に収まる人をマッチングの対象とする方法(分析者が設定)

### ■ 構成比の種類

1:1マッチング(one-to-one)

1:kマッチング(one-to-k)

1:Mマッチング(one-to-many)



PwC

46

## 今回利用したデータの内容



- DeSCヘルスケアが保有するデータのうち2018年1月以降の加入者台帳・健診・レセプトデータを用いて分析を行いました
- 保険制度ごとのデータ量・期間は以下のとおりです

保険者制度 (データソース)	健康保険組合	国民健康保険	後期高齢者 医療制度
観察可能期間	2018/01~2022/08		
加入者数※ (万人)	90.9	77.4	29.8
最長観察月数	56	56	51
観察可能人年 (万人年)	323	206	97

※ DeSCヘルスケアが保有するデータの最新情報はDeSCヘルスケア会員制メディア  
(<https://www.desc-life-actuary.jp/>)にて確認できます(2023年11月1日公開予定)

47

## 問診項目一覧 (1/2)

### ・ 各事例で用いた問診項目

データ名	問診内容
dose_blood_pressure_code	問診の質問「現在、血圧を下げる薬を使用している」 回答： 1:はい 2:いいえ
dose_blood_glucose_code	問診の質問「現在、インスリン注射又は血糖を下げる薬を使用している」 回答： 1:はい 2:いいえ
dose_lipid_code	問診の質問「現在、コレステロールを下げる薬を使用している」 回答： 1:はい 2:いいえ
history1_cerebrovascular_code	問診の質問「医師から、脳卒中(脳出血、脳梗塞等)にかかっているといわれたり、治療を受けたことがありますか」 回答： 1:はい 2:いいえ
history2_cardiovascular_code	問診の質問「医師から、心臓病(狭心症、心筋梗塞等)にかかっているといわれたり、治療を受けたことがありますか」 回答： 1:はい 2:いいえ
history3_renal_failure_HD_code	問診の質問「医師から、慢性の腎不全にかかっているといわれたり、治療(人工透析)を受けたことがありますか」 回答： 1:はい 2:いいえ
anemia_code	問診の質問「医師から、貧血といわれたことがある」 回答： 1:はい 2:いいえ
smoking_code	問診の質問「現在、たばこを習慣的に吸っている。(※「現在、習慣的に喫煙している者」とは、「合計100本以上、又は6ヶ月以上吸っている者」であり、最近1ヶ月間も吸っている者)」 回答： 1:はい 2:いいえ
weight_change_from_20_years_old_code	問診の質問「20歳の時の体重から10kg以上増加している」 回答： 1:はい 2:いいえ
exercise_habits_30_minutes_or_more_code	問診の質問「1回30分以上の軽く汗をかく運動を週2日以上、1年以上実施」 回答： 1:はい 2:いいえ
walking_physical_activity_code	問診の質問「日常生活において歩行又は同等の身体活動を1日1時間以上実施」 回答： 1:はい 2:いいえ
walking_speed_code	問診の質問「ほほ同じ年齢の同性と比較して歩く速度が速い」 回答： 1:はい 2:いいえ

PwC

48

## 問診項目一覧 (2/2)

データ名	問診内容
weight_change_1_year_code	問診の質問「この1年間で体重の増減が±3kg以上あった」 回答： 1:はい 2:いいえ
eating1_early_meal_code	問診の質問「人と比較して食べる速度が速い」 回答： 1:速い 2:ふつう 3:遅い
eating2_before_sleep_code	問診の質問「就寝前の2時間以内に夕食をとることが週に3回以上ある」 回答： 1:はい 2:いいえ
eating3_night_meals_code	問診の質問「夕食後に間食(3食以外の夜食)をとることが週に3回以上ある」 回答： 1:はい 2:いいえ (旧検査項目)
eating3_snacking_code	問診の質問「朝昼夕の3食以外に間食や甘い飲み物を摂取していますか。」 回答： 1:毎日 2:時々 3:ほとんど摂取しない
chewing_code	問診の質問「歯や歯ぐき、かみあわせなど気になる部分があり、かみにくいことがある」 回答： 1:何でも 2:かみにくい 3:ほとんどかめない
eating_habit_code	問診の質問「朝食を抜くことが週に3回以上ある」 回答： 1:はい 2:いいえ
drinking_code	問診の質問「お酒(清酒、焼酎、ビール、洋酒など)を飲む頻度」 回答： 1:毎日 2:時々 3:ほとんど飲まない
drinking_amount_code	問診の質問「飲酒日の1日当たりの飲酒量清酒1合(180ml)の目安:ビール中瓶1本、焼酎35度(80ml)、ウイスキーダブル一杯、ワイン2杯」 回答： 1:1合未満 2:1~2合未満 3:2~3合未満 4:3合以上
sleep_code	問診の質問「睡眠で休養が十分とれている」 回答： 1:はい 2:いいえ
lifestyle_habits_improvement_code	問診の質問「運動や食生活等の生活習慣を改善してみようと思いませんか」 回答： 1:改善するつもりはない 2:改善するつもりである(概ね6ヶ月以内) 3:近いうちに(概ね1ヶ月以内)改善するつもりであり、少しずつ始めている 4:既に改善に取り掛かっている(6ヶ月未満) 5:既に改善に取り掛かっている(6ヶ月以上)
request_for_health_guidance_code	問診の質問「生活習慣の改善について保健指導を受ける機会があれば、利用しますか」 回答： 1:はい 2:いいえ

PwC

49

## 分析環境

- 分析環境及び使用したパッケージ

■ 分析環境: Azure Databricks

■ 利用パッケージ・ライブラリ



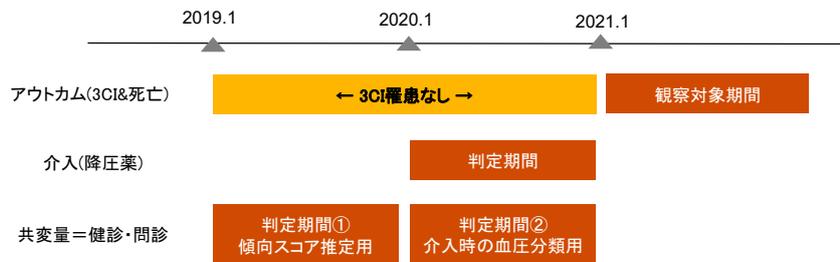
## 集計条件

適用事例	介入	アウトカム	共変量	UW条件
事例1: コロナ罹患後入院	<ul style="list-style-type: none"> <li>コロナ罹患有無</li> <li>レセプトICD10と診療行為を用いて罹患を判定</li> <li>2020年9月～2021年8月</li> </ul>	<ul style="list-style-type: none"> <li>全入院</li> <li>2021年9月～2022年8月</li> </ul>	<ul style="list-style-type: none"> <li>健診、問診データ</li> <li>2020年9月～2021年8月</li> </ul>	<ul style="list-style-type: none"> <li>アウトカム観察期間前3か月において入院がない</li> <li>全入院に「みなし入院」は含まない</li> </ul>
事例2: 血圧と降圧薬	<ul style="list-style-type: none"> <li>降圧薬使用有無</li> <li>2020年1月～2020年12月</li> </ul>	<ul style="list-style-type: none"> <li>3CI+死亡</li> <li>2021年1月～2021年12月</li> </ul>	<ul style="list-style-type: none"> <li>健診、問診データ</li> <li>2019年1月～2019年12月</li> </ul>	<ul style="list-style-type: none"> <li>アウトカム観察期間前2年間に於いて罹患がない</li> </ul>
事例3: 糖尿病と生活習慣	<ul style="list-style-type: none"> <li>生活習慣の改善の有無</li> <li>2018年1月～2021年12月</li> </ul>	<ul style="list-style-type: none"> <li>全入院</li> <li>2022年1月～2022年8月</li> </ul>	<ul style="list-style-type: none"> <li>健診、問診データ</li> <li>2021年1月～2021年12月</li> </ul>	<ul style="list-style-type: none"> <li>設定なし</li> </ul>

## 事例2:集計条件イメージ図

共変量の集計条件の補足

- 傾向スコア推定用の共変量・・・判定期間①2019年1月～2019年12月
  - 血圧分類判定用の共変量・・・判定期間②2020年1月～2020年12月
- ※ただし、降圧薬を用いている場合には、降圧薬使用後の健診・問診値を採用する

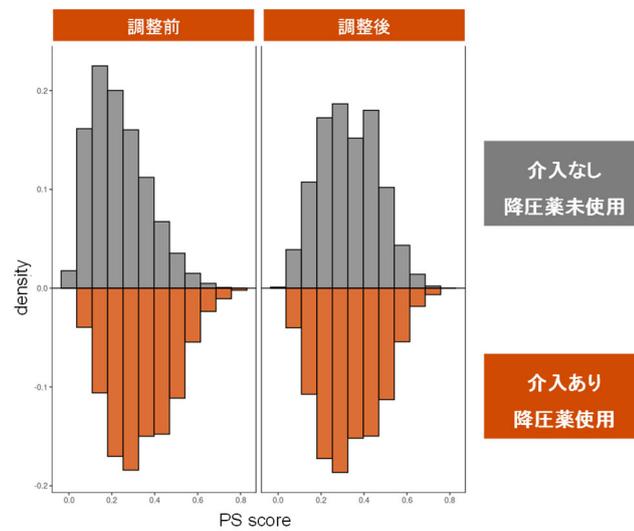


PwC

52

## 事例3: 傾向スコア分布

男性-50代-継続期間3年



PwC

53

## 事例1: 標準化平均差

50代男性	単純比較		標準化平均差 SMD
	コロナ罹患者 (n=2,337)	コロナ未罹患者 (n=83,693)	
年齢(歳)	54.2	54.2	0.014
BMI	24.5	24.2	0.081
SBP	126.0	125.7	0.021
DBP	80.9	80.6	0.031
HDL	59.1	60.0	0.055
LDL	124.9	127.4	0.082
Triglyceride	131.8	125.6	0.062
GOT	26.4	25.2	0.096
GPT	30.0	28.6	0.072
gamma_GT	55.6	50.6	0.077
HbA1c	5.5	5.5	0.073
Serum_Cr	0.6	0.6	0.019

SMD: 標準化平均差(standardized mean difference)

PwC

54

## 事例2: 標準化平均差

50代男性 高値血圧	単純比較			IPW調整後		
	降圧薬使用 (n=5,746)	降圧薬未使用 (n=15,992)	標準化平均差 SMD	降圧薬使用 (n=21,420)	降圧薬未使用 (n=21,922)	標準化平均差 SMD
年齢(歳)	54.0	53.3	0.258	53.5	53.5	0.009
BMI	25.8	23.9	0.513	24.5	24.5	0.009
SBP	129.6	124.9	0.361	126.0	126.2	0.017
DBP	83.9	80.6	0.358	81.4	81.5	0.012
HDL	57.6	60.4	0.185	59.3	59.5	0.008
LDL	121.0	130.3	0.327	127.4	127.5	0.004
Triglyceride	145.8	125.6	0.184	132.9	131.5	0.014
GOT	27.3	24.5	0.183	25.4	25.5	0.003
GPT	31.9	26.7	0.249	28.5	28.4	0.005
gamma_GT	63.2	49.8	0.206	54.8	54.0	0.014
HbA1c	5.9	5.6	0.129	5.4	5.4	0.014
Serum_Cr	0.9	0.9	0.083	0.5	0.6	0.009

SMD: 標準化平均差(standardized mean difference)

PwC

55

### 事例3: 標準化平均差

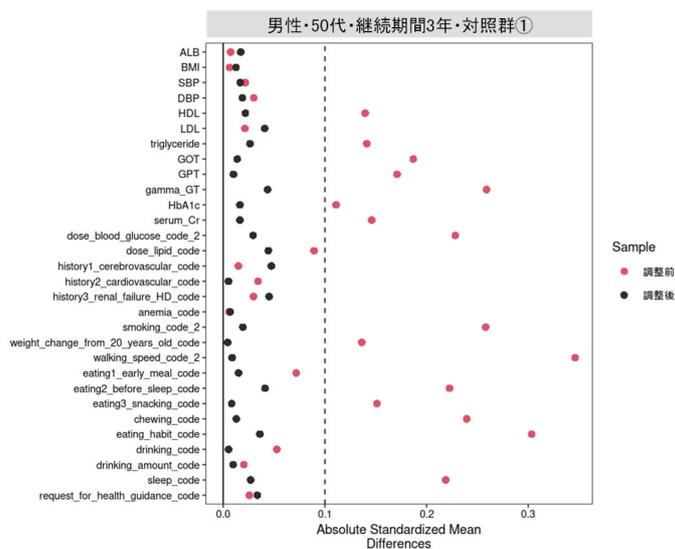
50代男性 継続期間3年以上	単純比較			マッチング調整後		
	継続改善あり	継続改善なし	標準化平均差 SMD	継続改善あり	継続改善なし	標準化平均差 SMD
	(n=934)	(n=2,658)		(n=922)	(n=922)	
年齢(歳)	54.7	54.7	0.007	54.7	54.7	0.017
BMI	25.5	25.4	0.007	25.5	25.5	0.013
SBP	129.0	128.6	0.021	129.0	128.8	0.017
DBP	82.9	82.6	0.030	82.9	82.7	0.019
HDL	59.5	57.3	0.140	59.3	59.0	0.022
LDL	125.0	125.7	0.022	125.2	123.9	0.042
Triglyceride	129.3	144.4	0.155	129.7	132.6	0.029
GOT	25.7	28.2	0.238	25.7	25.9	0.018
GPT	31.0	34.7	0.187	31.1	31.3	0.011
gamma_GT	50.2	66.6	0.368	50.5	53.2	0.062
HbA1c	5.9	5.7	0.115	5.9	5.9	0.017
Serum_Cr	0.7	0.6	0.140	0.7	0.7	0.016

SMD: 標準化平均差(standardized mean difference)

PwC

56

### 事例3: 標準化平均差



PwC

57

### 事例3: マッチング率 (対照群①②)

男性 50歳	マッチング調整後			
	対照群①: 全ての年度で「改善なし」群		対照群②: 継続努力必要群	
	改善なし	改善あり	改善なし	改善あり
1年以上	53.9%	99.8%	-	-
2年以上	41.8%	98.9%	67.9%	100.0%
3年以上	34.7%	98.7%	27.0%	99.7%
4年以上	30.4%	97.9%	16.1%	100.0%

PwC

58

# Thank you

[www.pwc.com/jp](http://www.pwc.com/jp)

© 2023 PricewaterhouseCoopers Aarata LLC. All rights reserved.  
PwC refers to the PwC network member firms and/or their specified subsidiaries in Japan, and may sometimes refer to the PwC network. Each of such firms and subsidiaries is a separate legal entity. Please see [www.pwc.com/structure](http://www.pwc.com/structure) for further details.  
This content is for general information purposes only, and should not be used as a substitute for consultation with professional advisors.