

# 異質性と時間依存性を考慮した介入効果評価 — 医療データを用いた因果推論分析 —

SOMPO ひまわり生命保険株式会社 尾上 辰徳

日本アクチュアリー会 2025年度 年次大会  
プレゼンテーション A-2

## 異質性と時間依存性を考慮した介入効果評価 — 医療データを用いた因果推論分析 —

SOMPOひまわり生命保険株式会社

尾上 辰徳

2025年11月7日

※本発表内容に関しては、いずれも個人の見解であり、所属会社の見解ではありません

司会 それでは、定刻になりましたので、セッションAの2、「異質性と時間依存性を考慮した介入効果評価」。副題として「医療データを用いた因果推論分析」についてのセッションを始めます。本日、司会を務めます、メットライフ生命の河野と申します。よろしくお願いいたします。

発表者の尾上様は、国内生命保険会社にて、決算業務や商品開発業務に従事されました。その後、データ分析支援を行うコンサルティング企業を経て、監査法人にて医療健康データのR&Dや付帯サービスの導入支援に携わられました。おとし、今回と同様の因果推論のテーマで、年次大会で発表いたしました。現在は、SOMPO ひまわり生命保険株式会社でDX推進部に所属され、デジタル施策の検証や効果特定など、データサイエンティストとして活躍されています。本日は、時間の許す限り、皆様のご質問に答える形にしたいと思います。会場とSlidoから質問を取り上げていきますので、Slidoで質問する場合は質問の記入をお願いいたします。それでは、尾上様、よろしくお願いいたします。

尾上 ご紹介ありがとうございます。皆さん、改めまして、こんにちは。SOMPO ひまわり生命の尾上と申します。本セッションにご参加いただきまして、ありがとうございます。このセッションでは、「異質性と時間依存性を考慮した介入効果評価」というタイトルでお話しさせていただきます。いずれも個人の見解であって、所属会社の見解ではないという点にご留意いただければと思います。初の個人での登壇ということもあり、ページ数がかさんでいますので、早速、本題に入りたいと思います。

## 目次

1. はじめに
2. 因果推論～算式セットアップ～
3. ユースケース
  - 3-1. 単一期間における効果推定
  - 3-2. 複数期間における効果推定
4. まとめ
5. Appendix

まず、「はじめに」のセクションで、課題設定をお話しさせていただきます。次に、因果推論とは何かを簡単にご説明させていただいて、用いる算式や記号のセットアップを行います。セクション3において、具体的にユースケースを用いて、因果推論の実務への応用ということでお話をさせていただきます。最後に「まとめ」となります。

## はじめに

# Section 1

### はじめに

4

#### 背景

- ✓ 保険会社の付帯サービスは、従来の「加入者全員への一律提供」から、**健康増進・予防ニーズや性別・年代、体況に応じたカスタマイズ型サービス**へシフト
- ✓ マーケティング施策において、**ターゲティング精度の向上や個別最適化**が重視される
- ✓ これらサービスは、契約者の体況・意識・行動特性に基づいて設計されるため、「誰にどの介入を行うか」が重要な意思決定要素となる

#### 課題

##### 課題1：単純比較による問題

- 個人の体況や行動特性に基づき介入が決定されるため、単純比較では共変量の差異の影響を受ける

単純比較では“バイアス”を除去できない

##### 課題2：マーケティングに必要な「誰に何を」の最適化

- 平均効果（ATE）だけでは、異質性を無視した介入評価となり、ROIの最大化が困難

「誰に・どの施策を」適用すべきかの定量把握が困難

##### 課題3：複数期間にわたる時間依存性の問題

- 健康介入は短期では効果が表れにくく複数年評価が必要
- 各時点の状態は、直前の状態の影響を受ける

時間依存性を考慮しないと効果推定に歪みが生じる

##### 課題4：複数介入の組み合わせ効果の評価

- 実際の施策は複数介入を組み合わせるケースが多く、単一介入の評価だけでは不十分

複合介入の相乗効果を評価できない

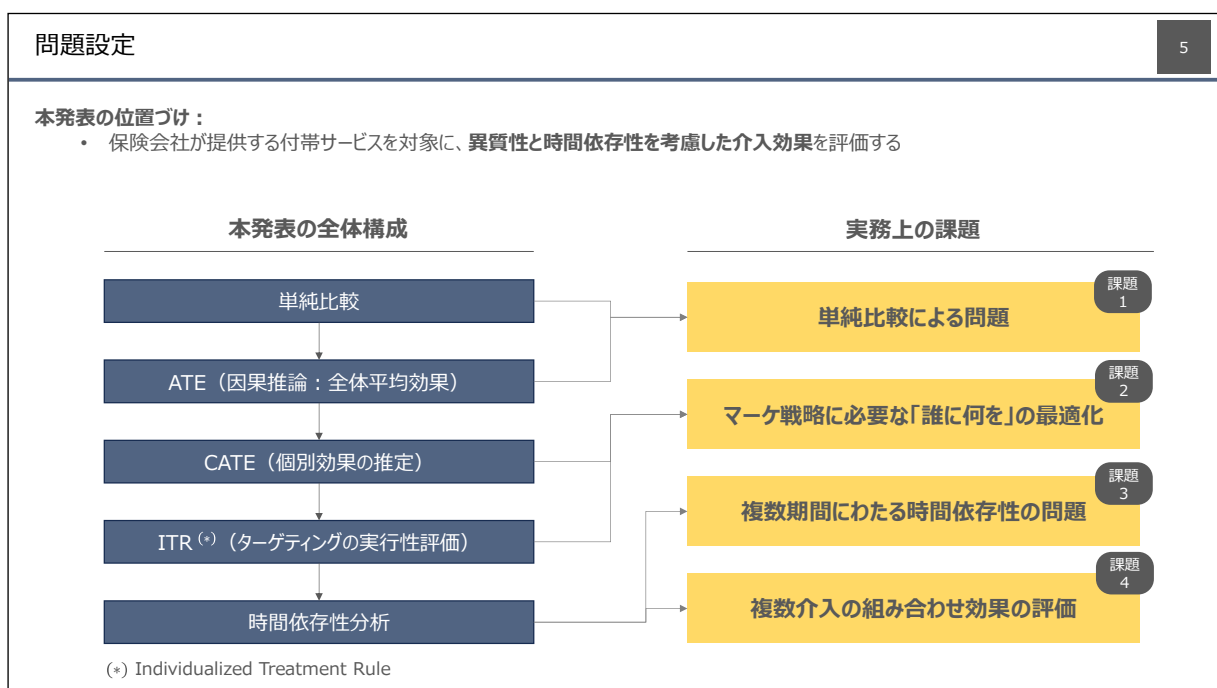
まず、そもそも今回のテーマを一言で言うと、因果推論の実務への応用になります。因果推論のフレームワーク自体は保険のバリューチェーン全体で活用可能ということは、以前、別の会社になりますが、おとし発表させていただきました。本日は、その中でも付帯サービスの評価に注目してお話しします。

「背景」に記載がございますとおり、保険会社の付帯サービス自体は、いわゆる健康増進や疾患の罹患の予防、重症化の予防など、個々の体況やニーズに応じたカスタマイズの流れにあると思っています。これら付帯サービス、いわゆる保険会社としての介入は、マーケティング色が強くて、誰に、どのような介入が効果的であるかという点が、会社の戦略的にも重要な要素であると考えています。

それを推し測る上での課題として、下に4点ほど記載させていただきました。一つめが、単純比較では、

バイアスの影響を受けて介入自体の効果を推定できないこと。二つめが、ATE（平均因果効果）だけでは、「誰に、どの施策を」という観点の定量化が難しい。三つめが、今回のような付帯サービス自体は、即効性を考えると、運動したから翌年度に疾患率が下がるという類のものではなく、じわじわ体況の変化、ひいては疾患率に影響を及ぼすということを考えると、複数期間を観察する必要があると思います。その場合、各時点の状態は前の時点の影響を受けるため、最終年度だけでの共変量、いわゆる背景因子の調整だけでは、ゆがみが生じる可能性がある。

最後に4点目として、実際の実務においては、あるサービスの一本やりというわけではなく、複数のサービスを展開していることが多いと考えると、どの順で、どの介入を提供していくか、推進していくかを考える必要があるという課題を認識しています。これらの課題に対して、今回ご説明させていただく因果推論のフレームワークを用いることで、より実態に即した介入の評価が行えます。



本分析では、それらの課題に対して、左側に記載の流れで一つの解を提示していこうと思っています。分析の流れとしては、まず、単一期間において平均効果（ATE）を推定して、そのあとに個別の効果、CATEと、介入ルール、ITRを設計・評価していきます。その後、複数期間においても同様に介入効果を検証していく流れになっています。このように段階的に分析を進めることで、先ほど挙げた四つの課題に対して、因果推論を通じた解決方法、解釈を提示していこうと思っています。

実務の施策とRWD分析には前提条件の違いがあり、以下の内容が挙げられる

実際の施策との乖離

分析上の仮定

内容	前提	現実	影響
介入の実施状況	加入者全員に一律で施策を打ったと仮定	実際は、体況・リスク・行動特性に応じて施策対象が選定される	“バイアス”を完全に再現できない
介入内容の定義	問診項目で施策内容を代用 (例：運動習慣、喫煙、飲酒、睡眠)	実際の施策は、プログラム内容・強度・頻度・フォローアップなどを含む	介入の「質」や「強度」を反映できない
介入のタイミング・頻度	年次データを基に、各年で介入有無を設定	施策は随時実施される場合があり、年次データではタイミングの精度が低い	実際の効果発現タイミングと乖離
アウトカムの定義	問診や健診データからアウトカムを設定 (例：疾病リスク)	会社の評価指標は、顧客満足度、疾病率、継続率など複合的	ROIなど実務評価と異なる可能性
施策のコスト情報	コストは一律設定 (介入コストは1,000円/1人と仮定)	実際は施策ごとにコスト構造が異なる (デジタル/対面施策、利用ツールなど)	ROI評価の精度に影響
加入者属性の偏り	健保組合・国保組合のRWDを利用	自社の加入者構成と異なる	外的妥当性が限定的

今回の分析では、DeSC ヘルスケア様にデータの提供をいただきました。実務の設定とは異なる点もあり、ここで挙げています。代表的なところで言うと、例えば実際の施策では、リスクなど、体況に応じて対象者が選定されると思いますが、今回は加入者全員に施策を打ったという仮定に基づいています。介入内容も、問診項目で代用していたり、実際の実務の強度や頻度とは異なる点をご留意ください。これらの前提条件の違いを踏まえたうえで、分析結果をご覧いただければと思います。それ以外に分析上の仮定として、下に青セルで三つほど記載しておりますが、こちらも後ほどご覧いただければと思います。ここまでの、今回の分析の背景、解決したい課題になります。

## 因果推論

# Section 2

因果推論とは「ある介入(Z)がアウトカム(Y)に与える影響を定量的に評価すること」

$$\text{ATE : 平均因果効果 (Average Treatment Effect)} = E[Y(1)] - E[Y(0)] \stackrel{\text{How?}}{\approx} E[Y|Z=1] - E[Y|Z=0]$$

反事の世界 = データからは不明

データで集計可能

	介入群 (Z = 1)	対照群 (Z = 0)
Y(1)	$E[Y(1) Z=1]$	$E[Y(1) Z=0]$ <span style="background-color: #cccccc;">観察不可</span>
Y(0)	$E[Y(0) Z=1]$ <span style="background-color: #cccccc;">観察不可</span>	$E[Y(0) Z=0]$

グレー箇所が観察できないため、単純な集計ではバイアスが生じる可能性がある

因果推論とは、観察データから未観測の反事実を推定し、真の因果効果に近づけるためのフレームワーク

では、次いで、因果推論とは何かからご説明を進めていきたいと思います。今回見たいこととしては、ある介入がアウトカムに与える影響を定量的に評価することです。例えば運動改善の施策が、死亡や疾患リスクにどれだけ影響を及ぼすかを知りたい。個人ベースですと、介入した場合と、そうでなかった場合の両方を同時に観測することはできません。つまり、反事実の世界は観測できないという制約があります。そこで、集団ベースで、期待値という形で介入群を評価していきます。具体的には、介入群と対照群のアウトカムの平均差をもって、平均因果効果（ATE）を推定します。ただし、現実には観測できない反事実、ここで言うと灰色の部分があります。それゆえに、単純な比較ではバイアスの影響を受けて真の効果を捉えることができません。

この後のページでは、こうした反事実の推定を可能にするための手法として、傾向スコアや g-formula などをご紹介します。

これらを踏まえると、ページ下部にも記載していますが、因果推論とは「観察データから未観測の反事実を推定し、真の因果効果に近づけるためのフレームワーク」とも読み変えられる、解釈できると思います。

変数名	単一期間	複数期間	今回の事例
介入	$Z_i^{(k)} \in \{0,1\}$ $k \in \{1,2,\dots,K\}$ : 介入の種類 (1: 介入, 0: 非介入)	$Z_i = (a_{i,1}^{(k_1)}, \dots, a_{i,t}^{(k_t)}, \dots, a_{i,T}^{(k_T)})$ <ul style="list-style-type: none"> <li><math>a_{i,t}^{(k_t)} \in \{1,2,3,4\}</math></li> <li><math>t \in \{1,2,\dots,T\}</math>: 時点</li> <li><math>k_t \in \{1,2,\dots,K\}</math>: 介入の種類</li> </ul> (1:改善, 2:悪化, 3:悪化継続, 4:良好継続)	<ul style="list-style-type: none"> <li>運動習慣改善</li> <li>歩行習慣改善</li> <li>禁煙</li> <li>飲酒習慣改善</li> <li>睡眠改善</li> </ul>
共変量	$X_i = (X_i^{(1)}, X_i^{(2)}, \dots, X_i^{(p)})$ (p: 共変量の個数)	$X_i = \begin{bmatrix} X_{i,1}^{(1)}, X_{i,1}^{(2)}, \dots, X_{i,1}^{(p)} \\ X_{i,2}^{(1)}, X_{i,2}^{(2)}, \dots, X_{i,2}^{(p)} \\ \vdots \\ X_{i,T}^{(1)}, X_{i,T}^{(2)}, \dots, X_{i,T}^{(p)} \end{bmatrix}$ (p: 共変量の個数, T: 時点)	<ul style="list-style-type: none"> <li>■ 基本属性               <ul style="list-style-type: none"> <li>年齢、性別、組合種別</li> </ul> </li> <li>■ 健診項目               <ul style="list-style-type: none"> <li>BMI、血圧、HbA1c,...</li> </ul> </li> <li>■ 問診項目               <ul style="list-style-type: none"> <li>投薬有無</li> <li>体重変化</li> <li>食習慣</li> <li>飲酒習慣,...</li> </ul> </li> </ul>
アウトカム	$Y_{i,T} \in \{0,1\}$ (T: 観察時点; 1: 疾患あり, 0: 疾患無し)		<ul style="list-style-type: none"> <li>3CI&amp;死亡</li> </ul>

ここでは、本分析で用いる変数の説明を行っています。個別に算式を読み上げることはしませんが、登場人物としては、「介入」と「共変量」と「アウトカム」の三つです。介入のうちの単一期間ですが、取りうる事象としては、複数期間に記載しています改善や悪化、それらの継続の四つを取りえますが、後ほど説明します傾向スコアを用いた推定では介入群と対照群を設定する必要があるため、いずれかを選択したうえでイチゼロで設定する必要があることから、単一期間では、取りうる値としては1と0ということで表記しております。共変量は、年齢や性別などの属性値と、健診・問診などの情報を含みます。これらはバイアスを調整する上で必要な情報です。アウトカムは、単一期間・複数期間を問わず、観察するタイミングでの疾患の有無になりますので、共通で記載しています。

定義：個人*i*の傾向スコア(Propensity Score)

$$\hat{e}(X_i) = P(Z_i = 1 | X_i)$$

- マッチングペアの作成：  
介入群の*i*に対してPSが近い対照群の個体*j*(*i*)を選択

$$j(i) = \arg \min_{j \in \mathcal{C}(i)} |\hat{e}(X_i) - \hat{e}(X_j)|$$

- ・ キャリパー条件を満たす対照群集団

$$\mathcal{C}(i) = \{j : Z_j = 0, |\hat{e}(X_i) - \hat{e}(X_j)| \leq \delta\}$$

本来知りたい  
効果

$$ATE = E[Y(1) - Y(0)] \dots (*)$$

- マッチング後のアウトカム：

$$Y_i^{treated} = Y_i(Z_i = 1), \quad Y_j^{matched\ control} = Y_{j(i)}(Z_{j(i)} = 0)$$

- マッチング後のATE…(\*)の近似値

$$\widehat{ATE}_{match} = \frac{1}{n_1} \sum_{i: Z_i=1, \exists j(i) \in \mathcal{C}(i)} (Y_i^{treated} - Y_{j(i)}^{matched\ control})$$

ここで、 $n_1 = \sum_{i=1} \mathbf{1}_{\{Z_i=1, \exists j(i) \in \mathcal{C}(i)\}}$ ：キャリパー条件を満たしマッチングした介入群の人数

推定  
フロー

傾向スコア算出

ロジスティック回帰や  
機械学習モデルを  
用いて $e(X_i)$ を推定

マッチング

$\hat{e}(X_i)$ が近い介入群と  
対照群をペアリング  
(Caliper)

ATE算出

マッチング後のペアで  
アウトカム差を計算し、  
平均化してATEを推定

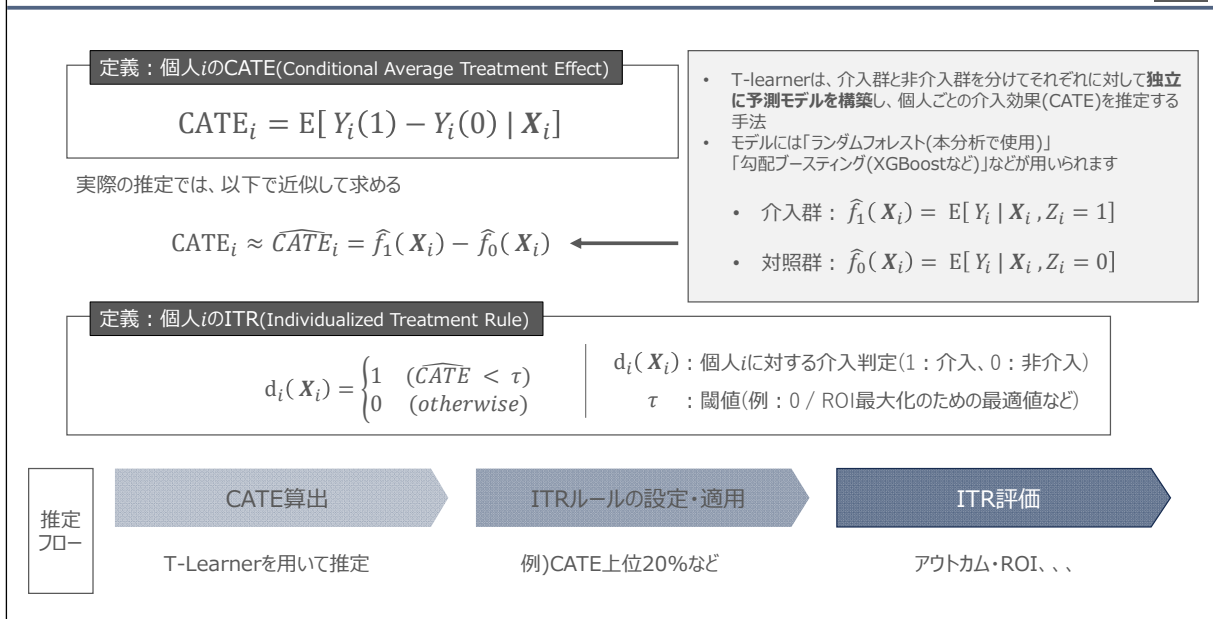
[事後検証]  
PS分布・SMD確認

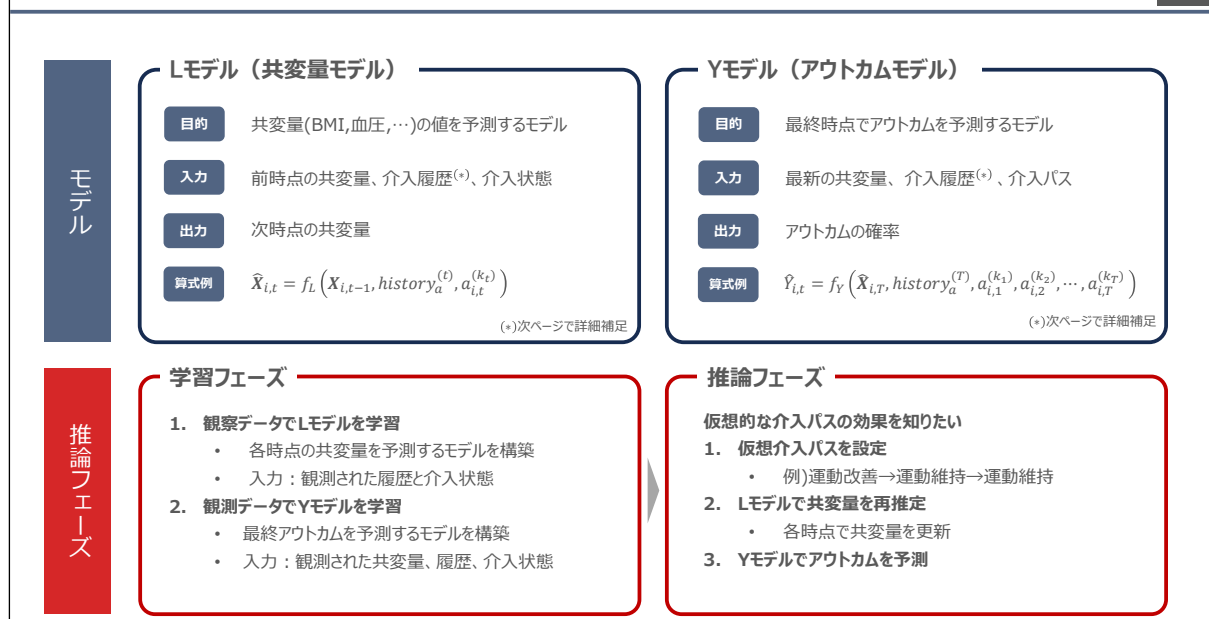
マッチング後の共変量分が  
均等か確認

では、続いて、具体的な各手法の紹介に移りたいと思います。はじめに、傾向スコアを用いた平均効果(ATE)の推定方法についてご説明します。傾向スコアとは、左上の算式にございますとおり、ある個人が介入を受ける確率を共変量に基づいて推定した値です。条件付き確率になっています。確率なので、値は0から1の範囲を取りえます。

ATEの算出フローですが、真ん中の算式を飛ばして、下の推定フローをご覧ください。まず、ロジスティック回帰や機械学習モデルを用いて、個人ごとに傾向スコアを算出します。次に、傾向スコアが近い介入群と対照群をペアリングすることで、共変量の分布を均一化させてバイアスを除去します。今回はこの方法を「マッチング」と呼んでいます。ただし、全く同じ値を持つ人を探すことは難しいので、「キャリパー」と呼ばれる、微小な範囲内であれば同一のスコアと見なしてペアリングを行うことが一般的です。最後に、このペアリングさせたペアにおけるアウトカムの差を平均化することで、ATEを推定していくという流れになっています。それらの算出過程を算式で表現したものが、真ん中の数式になります。

また、マッチングの妥当性を確認するために、「事後検証」とある、傾向スコアの分布や、SMD(標準化平均差)を用いて、調整前後でのバランスも評価します。Appendixに、今回の事例で傾向スコアの調整前後のグラフを載せていますので、そちらは後ほどご確認いただければと思います。このようにして観察データから反事実を推定して、因果効果を推定することができます。





続いて、タイトルにある二つめのテーマである、複数期間にわたる介入効果を評価するための手法として、「g-formula」と呼ばれる手法をご紹介します。繰り返しになりますが、予防や健康改善のような介入は、短期的な効果だけではなく、継続的な介入によって徐々に体況が変化して、最終的な疾患リスクに影響を与える。このため、単一期間での介入の有無だけでは、実態に即した評価が難しいです。加えて、各時点の状態は、前の状態に依存します。例えば前の状態の共変量や、それまでの介入履歴などに応じて変化する可能性がありますので、それを踏まえてアウトカムを予測する必要があります。

このような背景から、g-formula では二つのモデルを構築します。一つめが、左側にある「Lモデル」と呼ばれるもので、前の時点の共変量や介入履歴を基に、次の時点の共変量を予測するモデルです。右側にある「Yモデル」は、最新の共変量と介入履歴を基に、最終的なアウトカムを予測するモデルになっています。この二つのモデルを用いて、どのように、何を推定するのかということ、推論フェーズとしてご説明いたします。次のページに算式も含めて記載がございますので、そちらでご説明します。

## 学習フェーズ

1. 観察データでLモデルを学習  $\mathbf{a}_{i,t} = (a_{i,t}^{(1)}, a_{i,t}^{(2)}, \dots, a_{i,t}^{(K)})$

$$\hat{X}_{i,t} = f_L(X_{i,t-1}, \text{history}_a^{(t)}, \mathbf{a}_{i,t}) \quad (t \geq 1)$$

$$\begin{cases} X_{i,t-1} : \text{前時点の共変量(観察値)} \\ \text{history}_a^{(t)} = \{\mathbf{a}_{i,t-1}, \text{cum}_a(t), \text{ever}_a(t)\} \\ \mathbf{a}_{i,t} : \text{時点tの介入状態(観察値)} \end{cases}$$

$f_L$ : 線形回帰モデルを利用

2. 観測データでYモデルを学習

$$\hat{Y}_{i,t} = f_Y(\hat{X}_{i,T}, \text{history}_a^{(T)}, \mathbf{a}_{i,1}, \mathbf{a}_{i,2}, \dots, \mathbf{a}_{i,T})$$

$f_Y$ : 勾配ブースティング分類器を利用

## 推論フェーズ

1. 仮想介入パスを設定

$$\bar{\mathbf{a}} = (a_1, a_2, \dots, a_T)$$

2. 仮想介入パス適用

$$\mathbf{a}_{i,t} \leftarrow \bar{\mathbf{a}} \quad (\forall i, t = 0, 1, \dots, T)$$

3. Lモデルで共変量を再推定

$$\hat{X}_{i,t}^{(\bar{a})} = f_L(\hat{X}_{i,t-1}^{(\bar{a})}, \text{history}_{\bar{a}}^{(t)}, \bar{\mathbf{a}}_t)$$

4. Yモデルでアウトカムを予測

$$\hat{Y}_i^{(\bar{a})} = f_Y(\hat{X}_{i,T}^{(\bar{a})}, \text{history}_{\bar{a}}^{(T)}, \bar{\mathbf{a}}_1, \bar{\mathbf{a}}_2, \dots, \bar{\mathbf{a}}_T)$$

■ 介入履歴  $\text{history}_a^{(t)} = \{\mathbf{a}_{i,t-1}, \text{cum}_a(t), \text{ever}_a(t)\}$

## ✓ ラグ変数:

前時点t-1における介入状態(介入有無)

$$\mathbf{a}_{i,t-1} = (a_{i,t-1}^{(1)}, a_{i,t-1}^{(2)}, \dots, a_{i,t-1}^{(K)})$$

## ✓ 累計介入回数:

現時点tまでの介入kの累計回数(現在を除く)

$$\text{cum}_a(t) = \sum_{j=1}^K \sum_{s=0}^{t-1} a_{i,s}^{(k)}$$

## ✓ 過去に一度でも介入があったか:

過去に介入kを受けたことがあるか(0/1)

$$\text{ever}_a(t) = \begin{cases} 1 & (\text{cum}_a(t) > 0) \\ 0 & (\text{otherwise}) \end{cases}$$

冒頭に恐縮ですが、誤植がございます。学習フェーズの2番のfの添え字は、ここでは修正されたものですが、事前に公開されている資料ですとLになっていると思いますが、ここはYです。もう一つ、推論フェーズの3番のところは、逆にYになっていますが、Lになります。大変失礼いたしました。

改めて推論の流れですが、g-formulaでは2段階で構成されています。まず左側の学習フェーズにおいて、観察データを用いてYモデルとLモデルを構築します。これは、観察データを用いてモデルの学習をさせているフェーズです。モデルの入力データには、共変量だけではなく、「history」と記載させていただいた介入の履歴を含めました。これによって、介入の蓄積の効果や過去の影響を反映させることにしています。どのような内容かは下に記載がございまして、ラグ変数は1時点前の介入の有無ですね。そして、累計の介入の回数と、過去に一度でも介入があったかということを取り入れています。

重ね重ね申し訳ありませんが、先ほどの記号でセットアップした定義ですと、このa自体が、改善から改善の悪化の継続など、1、2、3、4の値を取ると記載していましたので、その定義に基づく、累計回数が期間を超えてしまうような状態になってしまっていて、この定義に従うと累計回数が正しく定義できていませんでした。やりたいこととしては、次の右側の推論フェーズでご説明する、注目したい介入パスにおける介入回数の累計や過去の介入有無を判別するものになっています。

この学習フェーズで構築したモデルを使って、次に推論フェーズにおいて、今回、効果を推定したい仮想的な介入パスをまず設定して、そのアウトカムを推定していきます。例えば、「運動改善、継続、継続」といったような介入の順序の効果を推定したい場合は、まずそれを介入のパスとして設定して、その介入前提で共変量を逐次的に更新していき、最終的なアウトカムを推定していくという流れです。従って、3番にある「Lモデルで共変量を再推定」の中のX共変量に関しては、推定した値を使っています。観察値ではなく、推定した値を使っている点をご確認いただければと思います。

この仕組みによって、個々に対して、実際には観測されていない介入のシナリオに対して、反事実のアウトカムを定量的に評価することが可能になっています。従って、このg-formulaの導入意義としては、単一の期間での評価では見えない介入の順番や継続性を定量化できる点にあります。保険会社においては、

どの施策をどの順で行っていくべきかを判断するための、強力なツールであると考えています。

## ユースケース

# Section 3

### DeSCデータ：全体像（保険者データの網羅性と代表性）

15

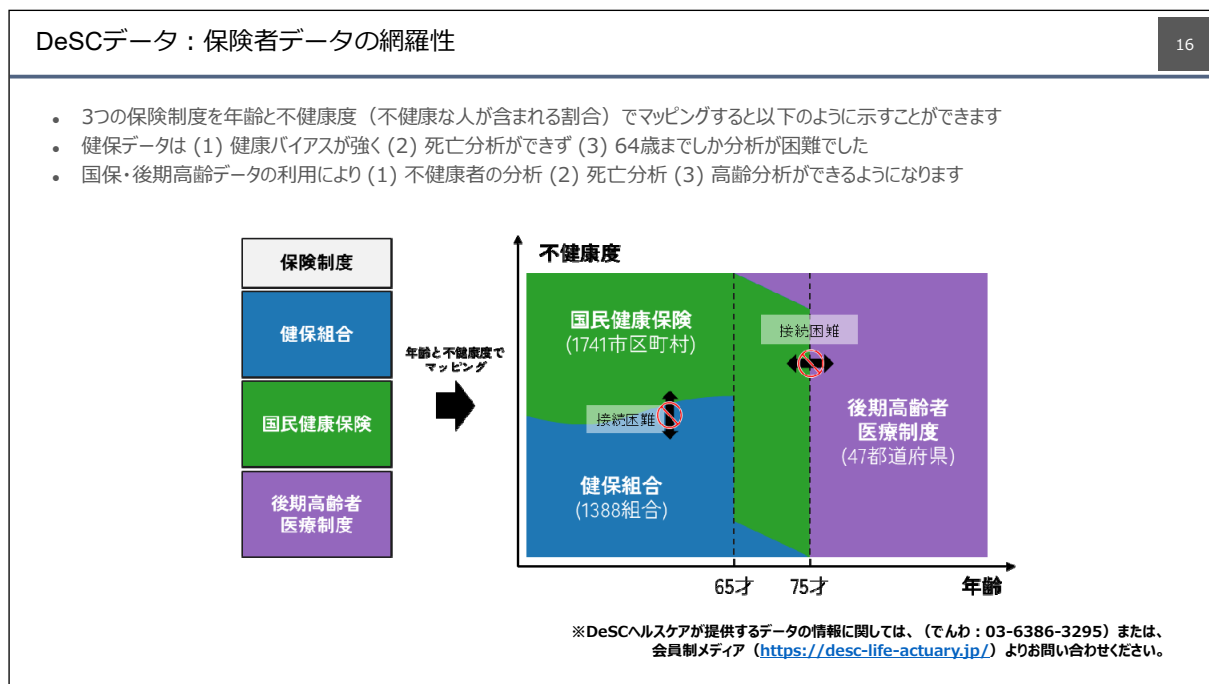
DeSCヘルスケアは以下3種類の保険制度にかかるレセプトデータを匿名加工情報として提供しています

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 ただし基本的に64歳以下
国民健康保険	市区町村	自営業者等、 企業に属さない者ただし基本的に74歳以下
後期高齢者 医療制度	都道府県 (広域連合)	基本的に75歳以上 〔 一定の障がいを持つ人は 65歳から加入 〕

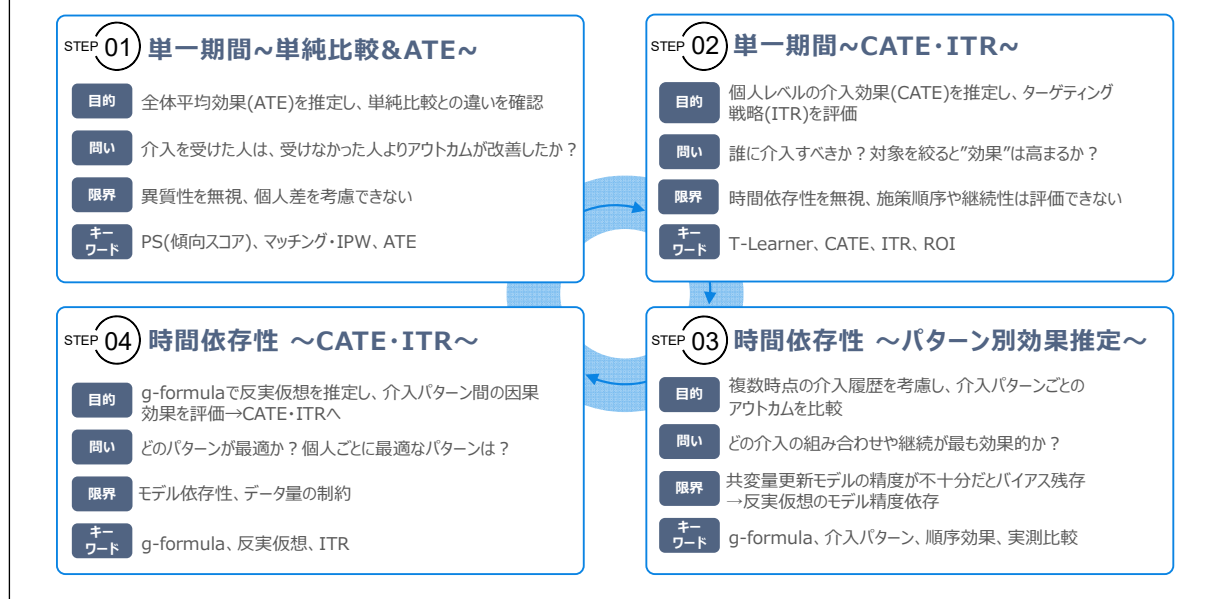
※DeSCヘルスケアが提供するデータの情報は、（でんわ：03-6386-3295）または、  
会員制メディア（<https://desc-life-actuary.jp/>）よりお問い合わせください。

では、ここまで見てきた手法を用いて、ユースケースの紹介に移りたいと思います。今回の分析は、繰り返しになりますが、DeSCヘルスケア様にご提供いただいています。ここでは、簡単に2枚ほどではありますが、今回の推定に用いたデータについてご紹介させていただきます。対象となる保険制度ですが、健保、国保、後期高齢の三つになります。仮個人IDでひもづけることによって、疾病や各種レセプト、健診・問診データをひもづけて分析することができます。当然ながら、今回の分析のように複数期間にわたっての追跡の分析も可能になっています。データの詳細は、ページ下部にございます URL からご

確認いただくことが可能ですので、後ほどそちらもご確認ください。



もう1点、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、死亡の情報は、院内死亡についてのレセプトの転記欄に記載されている程度で、病院外死亡を含めたより網羅的な死亡の情報を用いた分析ができないことと、後期高齢部分の分析はできないという課題点があります。国保と後期高齢のデータを用いることで、死亡の分析や高齢部分の分析が可能ということが、このデータセットの特徴です。今回は、健保と国保のみを使って分析しています。それ以外に含まれるデータセットの特徴としては、Appendixに記載していますので、そちらもご確認くださいと思います。



では、前置きが長くなりましたが、ここから本分析のユースケースに入っていきたいと思います。はじめに、全体図を示すロードマップがこちらになります。大きく4ステップで構成していきまして、ステップ1において、単一期間における平均効果（ATE）を推定して、単純比較との違いを確認します。次いでステップ2において、異質性をCATEを用いて推定し、ITRにおいて「誰に介入すべきか」というルールを決めて、個別化戦略を検討します。ここまでの、単一期間における分析です。

その後、ステップ3で、時間依存性を考慮したうえで、複数期間にわたる介入パターン別の効果を推定していきます。最後にステップ4において、単一期間と同様にCATEとITRを見ていき、検証していくという流れになっています。各ステップの目的や検証課題については、スライドに記載してございます。特に「限界」と書いてあるところが次のステップの前提になっているところですので、併せてこちらもご確認いただければ幸いです。

- 対象集団：
  - ・ 健保・国保に所属する30代～60代の男女
- アウトカム：
  - ・ 死亡or3CI罹患(がん、脳卒中、心筋梗塞)・・・脳卒中、心筋梗塞は状態継続日数を考慮せず、レポート発行を罹患の定義とする
- 共変量：
  - ・ 保険者属性：性別・年齢・保険者種別(健保・国保)
  - ・ 健診項目、問診項目 (※)

■ 介入：

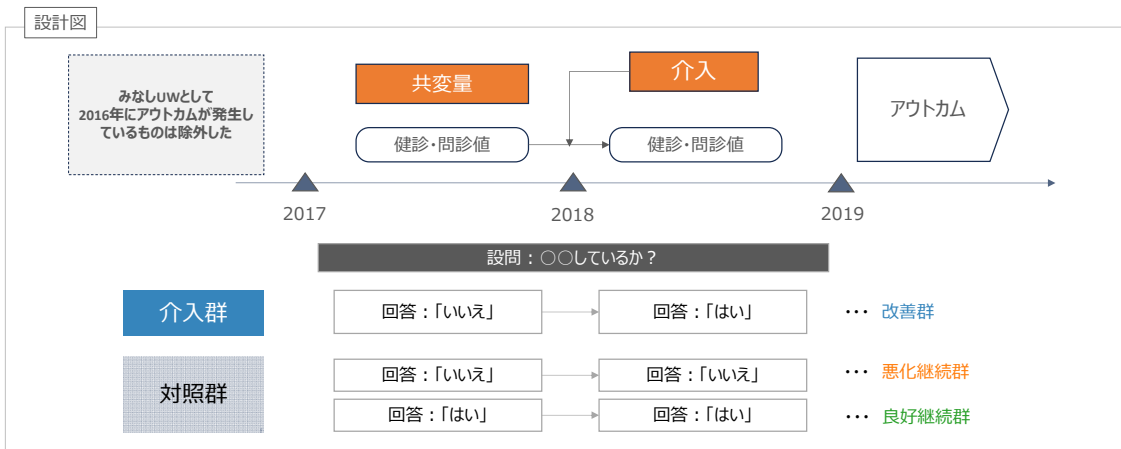
介入種類	使用項目	判定									
運動習慣 (exercise)	設問：1回30分以上の軽く汗をかく運動を週2日以上、1年以上実施 回答：はい/いいえ	<table border="1"> <thead> <tr> <th>前年\翌年</th> <th>はい</th> <th>いいえ</th> </tr> </thead> <tbody> <tr> <th>はい</th> <td>良好継続群</td> <td>悪化群</td> </tr> <tr> <th>いいえ</th> <td>改善群</td> <td>悪化継続群</td> </tr> </tbody> </table> <ul style="list-style-type: none"> <li>・ 飲酒の場合：「毎日・時々」→「ほとんど飲まない」を改善としている</li> <li>・ 喫煙の場合：「はい」→「いいえ」が改善</li> </ul>	前年\翌年	はい	いいえ	はい	良好継続群	悪化群	いいえ	改善群	悪化継続群
前年\翌年	はい		いいえ								
はい	良好継続群		悪化群								
いいえ	改善群		悪化継続群								
歩行習慣 (walking)	設問：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：はい/いいえ										
睡眠 (sleep)	設問：睡眠で休養が十分とれている 回答：はい/いいえ										
飲酒 (dring)	設問：お酒(清酒、焼酎、ビール、洋酒など)を飲む頻度 回答：毎日/時々/ほとんど飲まない										
喫煙 (smoking)	設問：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：はい/いいえ										

(※) 詳細はAppendixに記載

では、ステップ1の単一期間の分析から入っていきたいと思います。その前に、分析の条件ですが、こちらは単一期間・複数期間共通で、対象集団は健保・国保に所属する30代から60代の男女。アウトカムは、死亡もしくは3CIの罹患としました。介入は五つ設定してまして、運動習慣、歩行習慣、睡眠、飲酒、喫煙の五つ。判定方法としては、前年と翌年の問診データ2年間を使って、例えば歩行習慣の場合ですと、「1日1時間以上、実施していますか」という問いに対して、前年が「いいえ」、翌年が「はい」になっている人を「改善・介入あり」と判断しています。

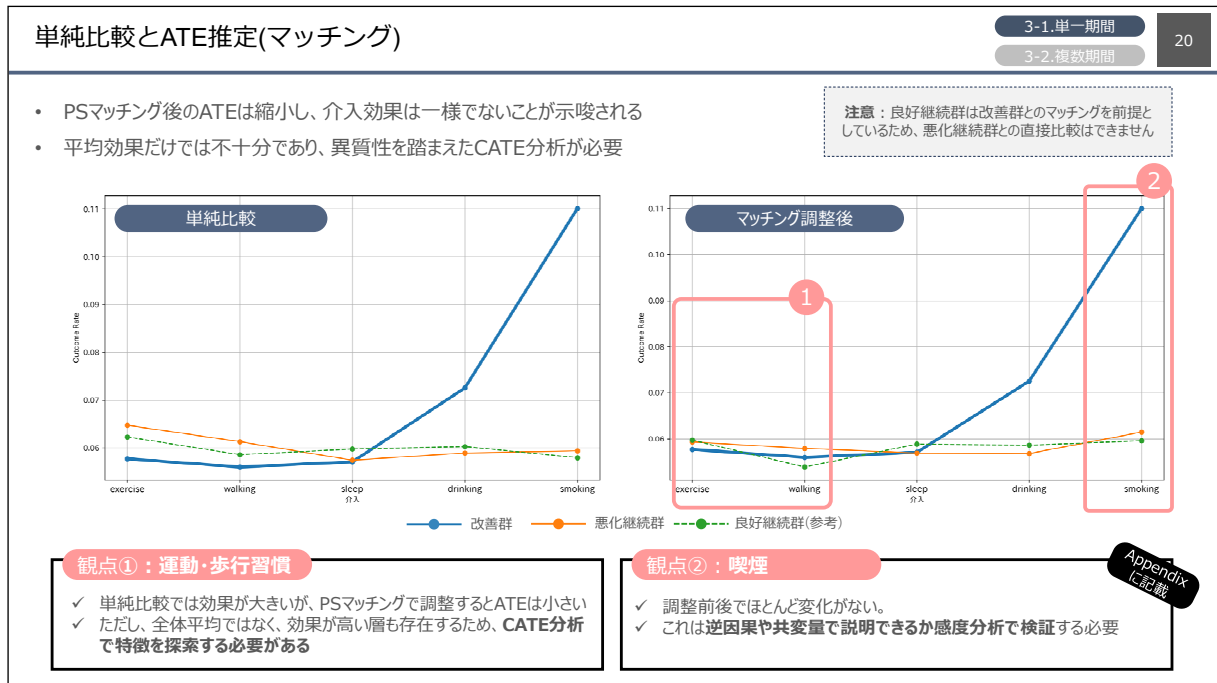
単一期間における分析フレーム

- ・ 共変量：2017年に観察された健診・問診値
- ・ アウトカム：2019年～2020年で観察
- ・ 介入：2017年→2018年の検診値を用いて設定



単一期間における分析のフレームワークですが、観察開始は2017年度として、共変量は2017年度のも

の。介入は、2017～2018年にかけての間診データで設定しています。介入群自体は、問診の問いに対して「いいえ」から「はい」になった人。改善群を今回は介入群と設定して、対照群は、「いいえ」から「いいえ」のままの悪化の継続の人と、参考に「はい」から「はい」の良好な継続の人も加えています。



この設計のもとで、単純比較と傾向スコアによるATEの推定の結果をお見せします。各介入をX軸にして、単純比較と、傾向スコアにマッチング後のATEの推定結果をプロットしたものです。ここでは三つほど、論点として述べたいと思います。

まず一つめが、運動と歩行習慣の結果です。単純比較では効果が大きいように見えますが、調整後はATEが縮小しています。これは、共変量の調整によってバイアスが除去された結果であって、調整後の平均効果は小さいものの、元々がそれなりにあったので、効果が高い層が存在する可能性がある。ここを次のCATE分析で深掘りしていこうと思います。

二つめが、飲酒や喫煙です。特に喫煙が、調整後も効果が高いままになっています。これは逆因果、つまり健康状態が悪化したため禁煙したという影響や、健診・問診データだけでは背景因子を十分に調整できていない可能性、未観測の交絡の影響が考えられます。後者に関しては、感度分析という手法を用いて未測定の影響を見てみました。結果としては、その影響は限定的であるということで、逆因果の可能性が高いのではないかと考えられますが、今回はこの深掘りは行っていません。

三つめが、先ほど「ATE自体は大きい」という表現をしましたが、そもそもATE自体、オーダーとしては0.01程度なので、かなり小さいです。これは、単一期間では付帯サービスとしての介入効果がほとんど見えないという結果になります。加えて、問診項目で介入を判定すること自体が定性的で、個人の主観に依存していることも影響しているのではないかと考えています。この点は、実際の実務の設計において、条件を明確に定量化することで改善可能であると考えていますが、今回の分析では問診データで代用していますので、このまま進めていきます。

各介入ごとの単純比較及び調整後のアウトカム比較

介入種類	単純比較		調整後-matching		精度(AUC)	マッチング割合		(参考)調整後-IPW	
	1(改善)	3(悪化継続)	1(改善)	3(悪化継続)		1(改善)	3(悪化継続)	1(改善)	3(悪化継続)
運動習慣(exercise)	0.0577	0.0548	0.0577	0.0593	0.67	100%	15%	0.0571	0.0556
歩行習慣(walking)	0.0560	0.0565	0.0560	0.0579	0.65	100%	27%	0.0554	0.0568
睡眠(sleep)	0.0571	0.0529	0.0571	0.0569	0.60	100%	43%	0.0561	0.0535
飲酒(drinking)	0.0725	0.0565	0.0725	0.0568	0.74	100%	7%	0.0685	0.0564
喫煙(smoking)	0.1101	0.0542	0.1100	0.0615	0.61	100%	10%	0.1090	0.0549

介入種類	単純比較		調整後-matching		精度(AUC)	マッチング割合		(参考)調整後-IPW	
	1(改善)	4(良好)	1(改善)	4(良好)		1(改善)	4(良好)	1(改善)	4(良好)
運動習慣(exercise)	0.0577	0.0623	0.0577	0.0597	0.68	100%	31%	0.0579	0.0614
歩行習慣(walking)	0.0560	0.0586	0.0560	0.0539	0.68	100%	34%	0.0562	0.0571
睡眠(sleep)	0.0571	0.0597	0.0571	0.0589	0.63	100%	16%	0.0579	0.0593
飲酒(drinking)	0.0725	0.0603	0.0718	0.0586	0.83	99%	7%	0.0668	0.0602
喫煙(smoking)	0.1101	0.0580	0.1101	0.0596	0.74	100%	2%	0.1170	0.0580

こちらは、先ほどのプロットの数値を表形式でまとめたものです。時間の都合上、口頭では説明はいたしません。次に、誰に聞くのかを明らかにするための CATE 分析に移っていきます。

背景

- ATE(平均効果)では、集団での効果は推定できるが、「誰に効くか」が分からない
- 介入の費用対効果を最大化するために「誰に介入すべきか」を明らかにしたい

アプローチ1：CATE分析

個人ごとの介入効果を推定 = 異質性

Why CATE?

- ✓ ATEでは「誰に効くか」までは分からない
- ✓ CATEにより背景因子(性年代・体況など)に応じた効果の異質性を把握

What Goal ?

介入効果の高い層の抽出と特徴把握

アプローチ2：ITR分析

p.25~

効果の高い人を優先する  
ターゲティング戦略の設計・評価

Why ITR?

- ✓ CATEでは「誰に効くか」は分かるが、実際の施策では「誰に介入すべきか」を決めるルールが必要

What Goal ?

CATEに基づいた「介入ルールの選定」とその「効果」の評価

CATE 分析の結果に入る前に、その必要性を整理しておこうと思います。ATE では、集団全体の平均差は分かりますが、誰に聞くのか、どの層で効果的なのかまでは分かりません。今回のモチベーションとしては、誰に介入すべきかを明らかにすることですので、このあとのページでは、まず CATE によって効果が高い層を特定して、その特徴を把握していきます。ただ、CATE だけでは誰に聞くのかまでしか分からず、実際の施策では、誰に介入すべきかというルールを決める必要があります。そこで、アプローチ2に書いて

ある ITR を用いて幾つかルールを設定して、その効果を評価していくという流れになります。

CATE分布とLow CATE層の特徴

3-1.単一期間
23

- Low CATE層はアウトカム率が高く、対象として重要
- ただし標準介入では効果が限定的なため、**一律ではなくカスタム化したメッセージングが必要**

算式：各個人 $i$ のCATE

$$CATE_i = \hat{f}_1(X_i) - \hat{f}_0(X_i)$$

- 介入=1, 対照=0
- $\hat{f}_1(X_i), \hat{f}_0(X_i)$ …ランダムフォレストによる回帰

▶ 解釈

同じ特徴量を持つ個人の介入有無によるアウトカムの差

- CATE>0(正)：介入を行うとアウトカムが**増加**する (逆効果)
- CATE<0(負)：介入を行うとアウトカムが**減少**する (望む効果)

介入種類	CATE_group	アウトカム	介入率
歩行習慣 (walking)	Low CATE (Risk ↑)	0.0811	0.223
	High CATE (Risk ↓)	0.0527	0.190
運動習慣 (exercise)	Low CATE (Risk ↑)	0.0729	0.136
	High CATE (Risk ↓)	0.0671	0.142

▼

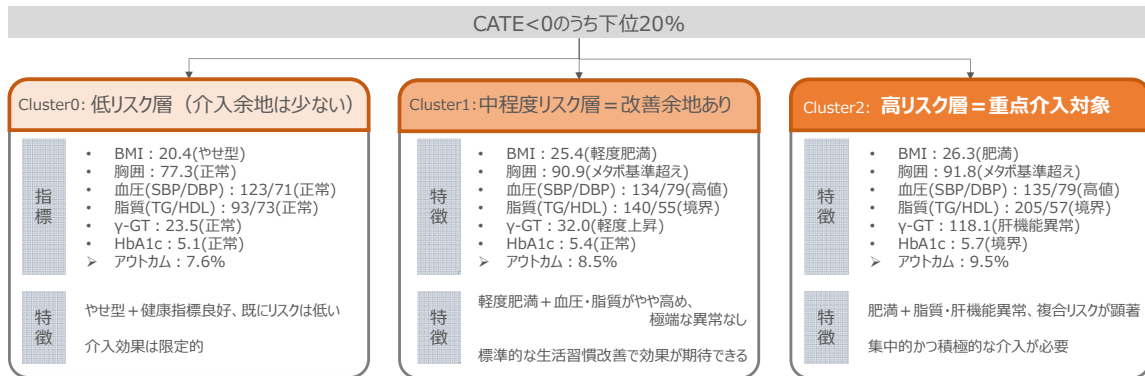
Low CATE層の特徴をクラスタリングで明らかにし、その結果を基にカスタム化されたITRルールを設計・評価する

CATE 分析の結果です。算式は前のページでご紹介したとおりですが、ここでは CATE の定義として、「介入ー対照群」としています。ここで差を取っているので、CATE の値が負の場合が、アウトカムが減少する方向。つまり、望ましい効果があるということを意味しています。左下のグラフをごらんいただければと思いますが、CATE が 0 付近に集中しているため、全体として異質性を見たうえでも、効果が小さいことが分かりました。ただ、「効果が小さいです」で終わらせてしまうと、ここで単一期間の分析が終わってしまうので、僅かではありますが負の値を持つ層が存在しているため、この特徴を深掘りすることで、施策の対象者、ターゲティングの可能性を探っていきます。

右の表では、歩行と運動習慣について、CATE の値が負の層から 20%。これを「Low CATE」と表現しています。逆に、CATE が正の値から上位 20%の「High CATE」を取って、アウトカムと介入率を比較しています。結果を見ますと、いずれも Low CATE の方がアウトカム率が高く、運動習慣においては介入率が低い結果が得られました。これに関しては、リスクが高い層へのアプローチが不足している可能性があるかと解釈できるかと思います。

このあとの特徴を探る上では、アウトカム率が高い歩行習慣に注目し、CATE 下位 20%の集団がどのような特徴を持つのかを次のページで見っていきます。

改善効果大きい層（CATE下位20%）をさらに3つの特徴的な集団に分類し、ターゲティング精度向上のためのプロファイル可視化



- ATEでは効果なし⇒CATE&クラスタ別でリスク特性を把握し「平均的な施策では効果が見えなかったが、ターゲティングを行うことにより効果を最大化できる」ことが示唆される
- NEXT : クラスタ別にターゲティングルールを決め、介入効果を確認する

特徴を探る上ではどのような手法を使ってもいいとは思いますが、ここではクラスタリングによって集団を分解して、解釈しています。BMI や血圧などの個別の項目ごとに、CATE の下位 20%とそれ以外の集団で比較をしていたのですが、あまり個別の項目単位だと差が得られず、特徴として解釈することが難しかったので、クラスタリングを用いて解釈しています。

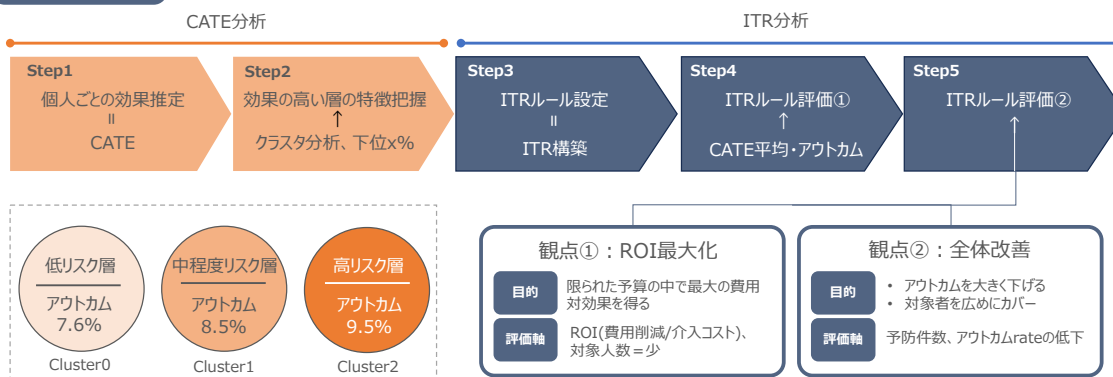
結果としてクラスタ数は三つで、記載のようにグルーピングしています。左から、低リスク層。こちらは介入余地は少ない。真ん中が中程度のリスク層ということで、軽度の肥満や、血圧・脂質がやや高めな層ということで、改善余地がある層。一番右が肥満や脂質異常の集団で、重点的に介入していく層というようにクラスタリングして解釈しました。このように、平均効果では見られなかった特徴や、改善の可能性を、CATE を基点にしてクラスタ別に把握することができました。

ITR（Individualized Treatment Rule）の目的・・・効果の高い人を優先するターゲティング戦略の設計・評価

Why ITR?

✓ CATEでは「誰に効くか」は分かるが、実際の施策では「誰に介入すべきか」を決めるルールが必要

ITR 推定ステップ



次のページは、この特徴を踏まえた介入ルール設計を、ITRを用いて行っていきます。ここでも改めて、ITRの流れをはじめに記載しています。先ほどから「効果进行评估する」と表現していますが、①、②に記載している二つの観点で効果进行评估していこうと思います。一つめが「ROI 最大化」という観点で、ここで言うROIは、非常にシンプルに介入コストと疾患時のコストで見えています。二つめは、全体としてということで、ROIや改善の効果、または対象者数で、トータルで見えて評価をする。詳細は後ほど触れたいと思います。ここで重要なことは、全員に介入することは非効率であって、対象者を絞ることによって大幅にROIや効果を改善できるのではないかという仮説を持ってルール設計・評価を行うということです。

- CATEにより抽出したルールごとに、①CATE平均②アウトカムを比較
- 全員介入は効果が薄く非効率であることが理解できる

Rule	平均CATE	アウトカム	対象人数
Rule1 : CATE下位20%	-0.018	0.081	中程度
Rule2 : クラスタ 2	-0.019	0.080	少ない
Rule3 : クラスタ 1 or 2	-0.018	0.087	少ない
Rule4 : 全員介入	-0.002	0.056	多い(基準)

介入効果

- ✓ 全員介入は非効率
- ✓ CATEを基にしたルールでは、効果が確認できる (全体介入と比べると約9倍)

アウトカム

- ✓ クラスタを含めるとリスクが高い層を重点的にカバーできる
- ✓ 全体平均は低くなる(良好な人が含まれるため)

総合評価

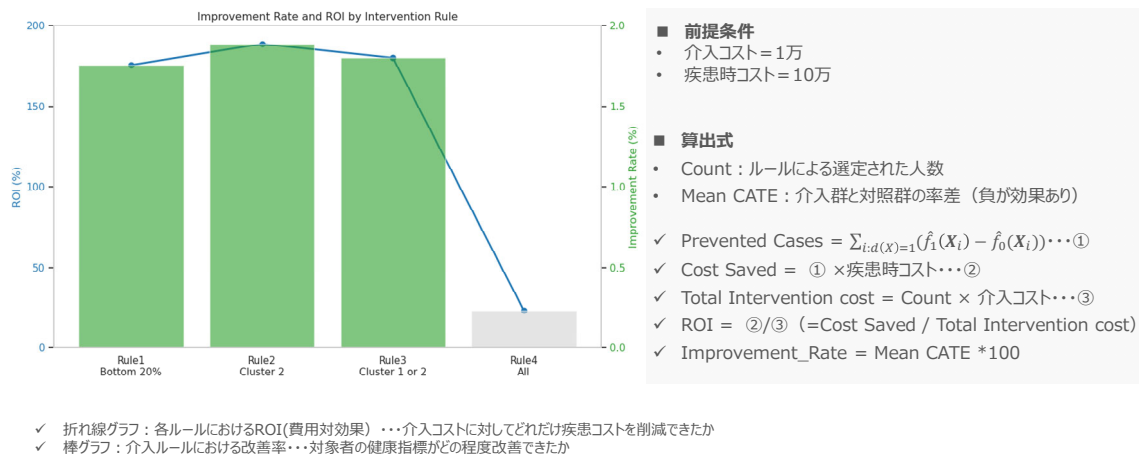
- **ROI最大化視点** : Rule2・Rule3は効果が高く、対象が絞られるためROI向上が期待できる
- **全体改善視点** : Rule1は広範囲をカバーし、効果も一定 → バランス型戦略
- **非効率回避** : Rule4は効果が低く、コストが膨大 → 推奨されない

ここまでの分析で、各ルールの効果とリスク構造を明らかに。  
これらの特徴を踏まえて、実際にどの戦略が費用対効果(ROI)で優れるのかを評価する

では、具体的に介入ルールの結果をごらんいただきます。このページでは、四つほどルールを設定してみました。一つめは、CATE の下位 20%全員に対して行うもの。二つめが、クラスタ 2、重点的に介入すべきだという層。三つめが、真ん中の中程度のリスクも含めた層。最後に全員介入です。比較の効果としては、平均の CATE とアウトカム、対象人数で見えています。対象の人数を今回は掲載することを避けていまして、全員介入を基準にして、「多い」、「少ない」、「中程度」という形で表現させていただいています。

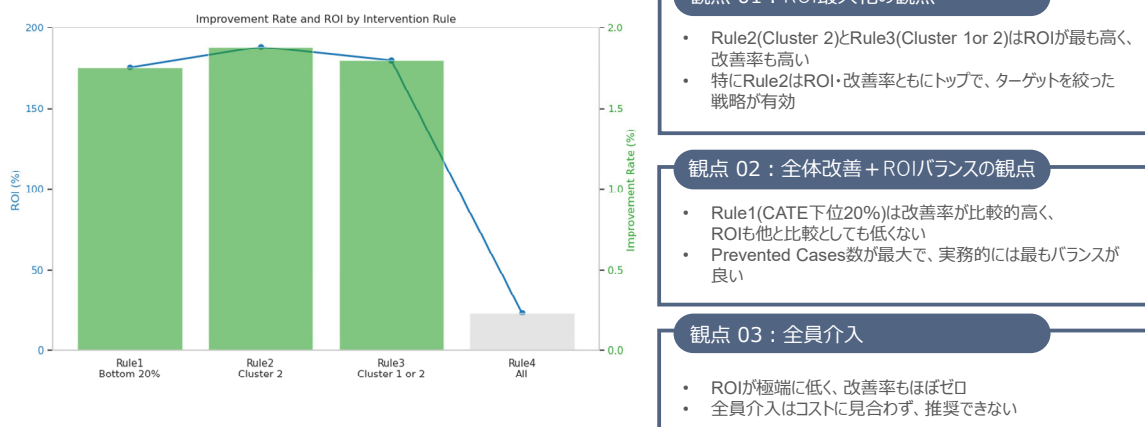
結果を見ますと、全員介入は、平均の CATE がほぼゼロですので、効果が低く、非効率であることが分かります。一方、CATE を基にしたルールでは、特にルール 2、ルール 3 では、対象人数を絞ることによって ROI 向上が期待できます。アウトカム・ベースで見た際には、全員介入が 0.056 と一番低いので、一見、全員介入がいいのではないかと思いがちですが、アウトカムの低さが必ずしも介入効果の高さを意味しておらず、そもそも全員介入ですので、体況がいい人もここには含まれている。改善による効果ではなくて、元々体況が良好な人も含んだ平均値ですので、ここでは平均の CATE の値で評価いただければと思います。この結果から、二つの観点それぞれで、どのルールを採用すべきかということ結論づけたいと思います。

- 介入対象の選定ルールによる「ROI」と「Improvement\_Rate」を可視化・比較を行った  
→ITRは最も低いリスクを達成し、ROI効率でも他戦略を上回る



こちらは、前提条件を記載しているページになります。1点だけ、ROI ですが、介入コストと疾患時のコストの二つで見えていくと申し上げました。ここでは、疾患の発生を防ぐことで削減できるコストという意味合いで記載しています。つまり、給付金や医療費など、疾患時に発生するコストの減少効果という意味で用いています。「1万」、「10万」とありますが、これも私が仮置きしたものですので、実際に何かを参照して設定したわけではございません。グラフは次のページにも同様のものを載せていますが、ROI を折れ線で、改善率を棒グラフで示しています。


- ROIを最大化したい場合 : 高リスク層 (Cluster 2やCluster 1 or 2) を狙う戦略が有効
- 改善効果とROIのバランスを重視する場合 : CATE下位20%を対象とするRule1が最適
- 全員介入 : 非効率であり、リソース配分の観点からも避けるべき



はじめに ROI 観点ですが、ルール2がクラスタに重点介入すべきだと言った層に介入することが一番有効だと、この結果から判断できると思います。二つめの全体改善の観点で言うと、ROIこそルール2や3に劣るものの、対象者数が一定多いルール1に介入していくことが、ターゲティングする上では有効だと考えました。最後に全員介入ですが、先ほどから申しているとおり、ROIが極端に低いですし、改善率も低いことから、コスト見合いでリジェクトされるべきだという直感がデータでも確認できたと思います。このように、目的に応じて戦略を選択する一つの方法として、CATEを用いて異質性を考慮したうえで、ITRで適用していくという流れで、今回はご説明させていただきました。


まとめ

3-1.単一期間
3-2.複数期間
29



分かったこと

- **“バイアス”調整の重要性**  
✓ 単純比較では“バイアス”が残り、介入効果を過大評価するリスクがある
- **平均効果の限界**  
✓ ATEだけでは異質性を捉えられず、全員介入は非効率
- **ターゲティングの有効性**  
✓ CATE分析→クラスタ分析により改善効果が大きい層を特定可能
- **ROI改善の可能性**  
✓ ITRを用いたターゲティング戦略はROIを大幅に改善できる



課題点

- **介入達成条件の曖昧さ**  
✓ 「運動習慣あり」などの定義が定性的で、個人差が大きい → **定量的な達成基準の設定**が望ましい
- **短期間での効果発現の難しさ**  
✓ 健康改善は時間を要するため、単一期間では効果が見えにくい → **継続的な介入効果の評価**が必要
- **施策への反応データの不足**  
✓ メール開封やクリックなど、実際の行動反応を加味できていない → **行動特性を反映したモデル構築**が課題
- **実務とのギャップ**  
✓ コストや施策強度、実施頻度など、現実の施策条件を十分に反映できていない

POINT

平均では見えない効果を、個人ごとの異質性を捉えて可視化し、戦略に活かす

では、ここまでのセクション3の1を一旦まとめた後に、複数期間の分析に行きたいと思います。分かったこととしては、平均効果、ATEだけでは、どの施策を推進していくべきかという意思決定に必要な情報が不十分であるということです。全体の効果を推定するだけではなく、どの層に介入すべきか、その層が効果的なのかというところまで評価して初めて、検討の土台に乗せることができると考えています。

課題点ですが、スライドの赤字に記載しているとおり、実務との乖離がございます。例えば介入の強度や頻度、コスト構造などは、今回のデータセットでは反映できていない。これは、分析の限界というよりは、実務で考慮すべきポイントだと思っています。最後にポイントとして記載していますが、平均では見えない効果を異質性を捉えて評価していくことで、戦略の検討に生かすことが可能だということです。では、複数期間に入る前に、一度ここで質疑応答の時間を取れればと思います。

司会 質疑応答に入る前に、冒頭、オンラインで参加の方は音声聞こえなかったということで、大会委員よりお詫び申し上げます。

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。質問がないようですので、Slidoから質問を読み上げます。「スライド18に関連する質問です。検診問診項目を見ると、(死亡+)脳卒中・心筋梗塞に絞った方が、より介入効果がよりクリアに分析できる気もするのですが、いかがでしょうか。」

という質問が来ています。

尾上 おっしゃるとおりでして、問診項目、健診項目で対象の項目を絞ることで、より効果としてはクリアになると思いますが、今回は得られている健診・問診項目全て使って介入していて、その深掘りや調整は行っていないのが現状です。

司会 ありがとうございます。もう1点、質問が来ております。「介入を実際に行うためのインセンティブには、例えば保険料の引下げがあると思いますが、これをクラス1・2に限定するサービスとして行うことは、健康体との公平性はどうか整理すればよいのでしょうか。所属会社でも「健康チャレンジ制度」をお持ちのようですが、尾上さん個人のコメントで結構ですのでお伺いできればと思います。」というご質問が来ております。

尾上 保険料の引き下げなど、インセンティブを考慮して、どのように効果を推定したらよいかという質問ですね。

司会 健康体との公平性を、どのように整理すればいいかという質問です。

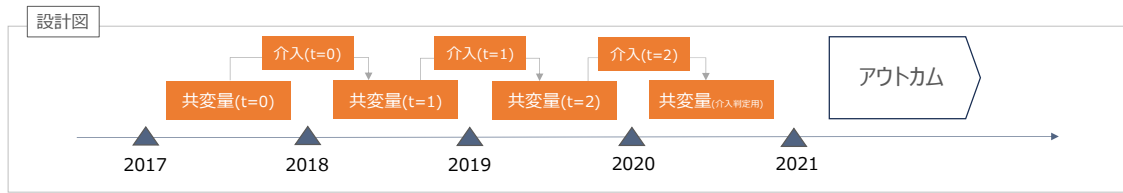
尾上 公平性という観点ですか。

司会 はい。

尾上 特定の層だけにのみ、公平性を欠くようなサービスの提供は、やはり難しいと考えています。その点、インセンティブを特に与えるようなものに関しては、公平性はしっかり考慮する必要があると思うのですが、ここですぐに「こういった形で考慮すべき」という解が出ず、私自身、ここも今の所属で検討しているところですので、引き続きその点は考えていきたいと思います。この場で回答できず、申し訳ないです。

司会 ありがとうございます。では、続きをお願いいたします。

- 共変量、介入：2017年～で毎年観察
- アウトカム：2021年～2022年で観察



変数		t=0	t=1	t=2	t=T
②共変量	①介入変数	観察値	設定値（指定）		-
	検診・問診値	観察値	予測値（L-model）		-
	属性変数 <sup>(※)</sup>	観察値			-
	履歴変数	観察値から計算	更新後の値		-
③アウトカム			-		予測値（Y-model）

各時点における変数の取り扱い

(※)観察開始時の年齢や性別、保険者種別など時点に依存しない変数

尾上 では、ここからは、時間移動性を考慮した、すなわち複数期間の分析に移っていきたく思います。分析前提は単一期間と同様で、共変量、介入の取り方はスライド記載のとおりです。下の表は推論フェーズにおける変数の扱いを表したもので、基本的に観察値ではなくて推定した値を使います。属性の変数、いわゆる年齢や性別、保険種別などは、時点に依存しない観察値を使うということで、ここで整理させていただきます。

単一期間と異なる考え方に関して補足

疑問①	単一期間のように傾向スコアを用いた共変量の調整とは考え方が違う？
疑問②	観察された介入パターンを持つ人同士の比較だと何がダメなの？モデルを用いて推定する必要性は？
疑問③	単一期間では調整前後でアウトカムを比較していたが、複数期間でも調整前後の比較を行うか？

実際の分析結果に入る前に、単一期間との違いを、素朴な疑問という形で整理させていただきました。私自身、この理解に時間を要したところでもあるので、疑問①、②、③という形で記載しています。

疑問①

単一期間のように傾向スコアを用いた共変量の調整とは考え方が違う？

回答：異なる。指定した各介入パターンに対して、反実仮想アウトカムを推定している

アプローチ	単一期間：PS調整	複数期間：g-formula
調整対象	介入群・対照群の共変量データ	介入によって変化する共変量の時系列データ
調整方法	傾向スコアによる調整（マッチング/IPW）	逐次予測モデル（L-model）による共変量生成
介入の扱い	一時点の介入有無	複数時点の介入履歴(順序や継続性など)
アウトカム推定	観察値(の加重平均)	反実仮想アウトカムのモデル予測値
比較方法	介入群 vs 対照群	パターンごとのリスク比較

まず一つめの疑問ですが、傾向スコアを用いた推定とは調整方法が異なるのかということで、答えとしては異なります。単一期間では、傾向スコアを用いて、あくまで介入群と対照群でペアを作って、そのペアにおいて共変量が一致するように、IPW という手法や今回のマッチングという手法によって調整しています。一方で g-formula は、逐次的に前の時点の共変量や介入の履歴を用いて共変量を予測していくという流れになっていますので、介入群や対照群ではなくて、評価をしたいパターンごとのリスクの比較になっている点が、単一期間との違いになっています。

疑問②

観察された介入パターンを持つ人同士の比較だと何がダメなの？モデル用いて推定する必要性は？

回答：反実仮想を推定するため

ある人が「運動→歩行→禁煙」という改善介入を受けた場合、「運動→睡眠→禁煙」という状態は観察値では答えられない

観察値でその介入パターンの人がいればその集団と比較すればよいのでは・・・？

交絡問題

✓ 仮に同一介入パターンの人がいっても、共変量が異なる可能性  
→アウトカムの差異が介入パターンの違いによるものなのか、共変量の違いによるものなのか判別できない

件数問題

✓ 特定の介入パターンの人が少ないと、安定性に欠ける  
→観察データのみでは、網羅的かつ安定的な検証が困難

推定フロー

1. g-formulaにより、「介入を変更した場合に共変量はどう変化し得るか」をモデルで予測・・・L-model
2. 予測された共変量で、アウトカムを予測し、評価する・・・Y-model

続いて疑問②として、実際に観察されたパターンどうしの比較だと何か都合が悪いのかということですが、記載の2点、交絡と件数の問題が特にネックとなりえます。一つめが、観察された介入パターンを持つ人どうしで比較しても、そのアウトカムの差が、介入のパターンによるものなのか、背景因子、共変量の違いによるものなのかが分からない点。

もう一つが、観察ベースだと特定の介入パターン的人数が少ないと比較の際に安定性に欠けるという問題があります。もちろん g-formula でも、モデルを用いますので、極端に件数が少ない、例えばある介入が1人しかいないという場合には、学習データがその人だけになるので、精度としては落ち込むことになるかと思えます。そのような場合、基礎集計など、そもそもその介入パターン自体を見ていきたいのかという点で、考慮するか否かを検討する必要があると思っています。

素朴な疑問③

3-1.単一期間
3-2.複数期間
34

疑問③

単一期間では調整前後でアウトカムを比較していたが、複数期間でも調整前後の比較を行うか？

回答：目的に応じて実施有無は異なる

アプローチ	単一期間：PS調整	複数期間：g-formula
比較目的	バイアスの影響を評価	実際の介入パターンと理想的な介入パターンの差分から改善余地を評価
主張内容	<ul style="list-style-type: none"> <li>• バイアスの主要因とその影響度合い</li> <li>• 調整による因果効果の推定</li> </ul>	<ul style="list-style-type: none"> <li>• 介入の順序や継続性の効果</li> <li>• 最適な介入パターンの設計</li> <li>• 実施中の施策の改善余地の定量化</li> </ul>
実務観点	介入有無による効果検証	介入戦略の設計・最適化

最後の疑問ですが、単一期間では調整前後でアウトカムを比較していましたが、複数期間では調整前後の比較を行うかということで、単一期間の場合、バイアスの影響を見えています。調整前後で ATE が広がるのか、縮小するのか。縮小するということは、バイアスによって過大評価されていたことになります。一方で g-formula の場合ですと、調整前後の比較というよりは、注目したい介入の効果を見たいわけですので、介入パターン間の比較になります。調整前後の比較という観点で言うと、実際に観察された介入パターンの効果がどれだけ改善余地があるのかを見る上では、調整前後で見る必要があると思います。意味合いが異なる点をご理解いただければと思います。

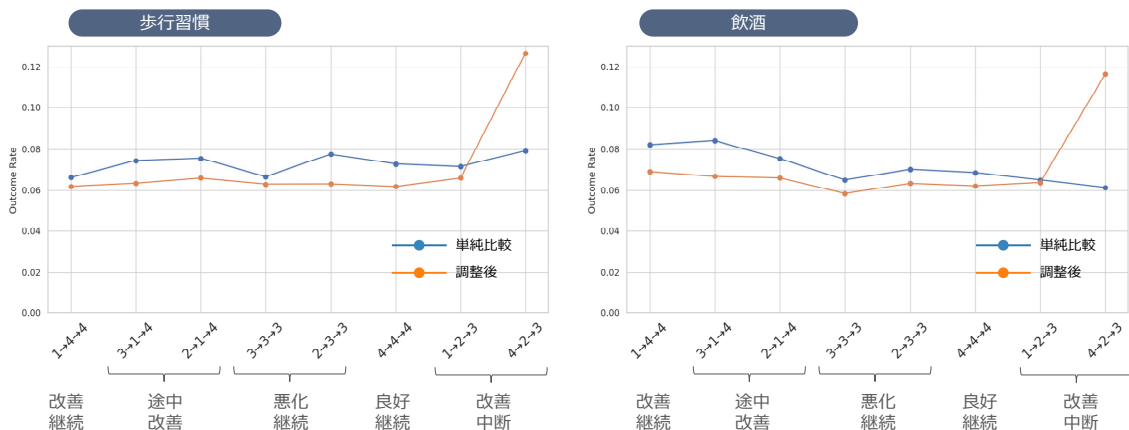
同一の介入パターンの効果を推定する。介入項目とフラグの意味を再度確認する

- 介入(再掲) : 運動習慣、歩行習慣、睡眠、飲酒、喫煙の5つ
- 介入フラグの意味 : 改善 = 1 , 悪化 = 2 , 悪化継続 = 3 , 良好継続 = 4

介入タイプ	パターン表記
改善継続者	<ul style="list-style-type: none"> <li>1 → 4 → 4 (改善→良好継続→良好継続)</li> </ul>
途中改善	<ul style="list-style-type: none"> <li>3 → 1 → 4 (悪化継続→改善→良好継続)</li> <li>2 → 1 → 4 (悪化→改善→良好継続)</li> </ul>
悪化継続者	<ul style="list-style-type: none"> <li>3 → 3 → 3 (悪化継続→悪化継続→悪化継続)</li> <li>2 → 3 → 3 (悪化→悪化継続→悪化継続)</li> </ul>
良好継続者	<ul style="list-style-type: none"> <li>4 → 4 → 4 (良好継続→良好継続→良好継続)</li> </ul>
改善中断者	<ul style="list-style-type: none"> <li>1 → 2 → 3 (改善→悪化→悪化継続)</li> <li>4 → 2 → 3 (良好継続→悪化→悪化継続)</li> </ul>

では、実際に、ここから結果に移ります。まず、それぞれの統一の介入パターンを評価しました。運動であれば、「運動の改善、継続、継続」なのか、途中で離脱したようなパターンなのか。組み合わせとしては、取りうる8パターンを五つのタイプとして分類しています。

- g-formulaによる時間依存性調整で、単純比較と異なる傾向が確認され、バイアス補正の重要性が示唆された
- 歩行・飲酒ともに、改善継続群はリスクが低く、中断群は最も高い傾向。単純比較では見えにくい継続性の影響を定量的に確認



結果に関しては、歩行と飲酒のみを、紙面の都合上、掲載しています。単純比較と異なって、改善を途中でやめてしまっているパターンが調整後で最も高い率を示していて、改善の継続や良好の継続群は、同程度に低い結果になっています。これは、バイアスや時間依存性を考慮しない場合には見えにくい影響で、今回の手法によるバイアス補正の重要性を表す結果になったと思っています。

ただ、今回の分析は、実際に介入を行ったわけではないので、問診データを用いて「改善した」という仮定に基づいて分析を行っていますので、その点を踏まえますと、今回のこの分析結果自体を基に「改善層が最もよい」と判断するのではなく、改善の持続や中断の影響を定量的に示したところに価値があると思っています。

複数介入組み合わせパターンの効果推定

3-1.単一期間
3-2.複数期間
37

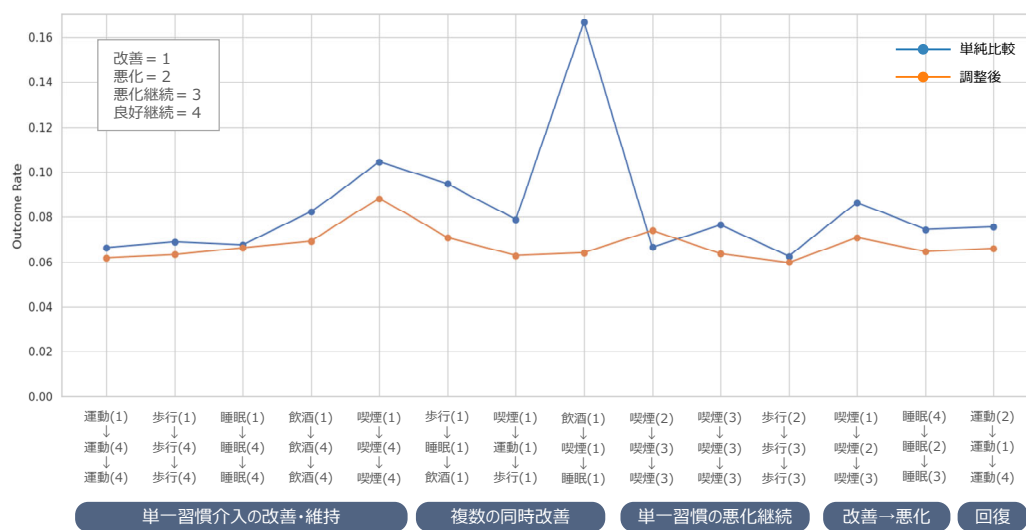
複数の介入パターンの組み合わせの効果を推定する

- 介入(再掲)：運動習慣、歩行習慣、睡眠、飲酒、喫煙の5つ
- 介入フラグの意味：改善 = 1，悪化 = 2，悪化継続 = 3，良好継続 = 4

介入タイプ		パターン表記
1. 健康行動の改善系	単一習慣の改善と継続	• 各介入ごとに 1 → 4 → 4
	複数の同時改善	• 歩行(1)→睡眠(1)→飲酒(1) • 睡眠(1)→運動(1)→歩行(1) • 飲酒(1)→喫煙(1)→睡眠(1)
2. 健康行動の悪化系	単一習慣の悪化継続	• 喫煙(2)→喫煙(3)→喫煙(3) • 喫煙(3)→喫煙(3)→喫煙(3) • 歩行(2)→歩行(3)→歩行(3)
3. 行動の変動・リバウンド系	改善 → 悪化 (リバウンド)	• 喫煙(1)→喫煙(2)→喫煙(3) • 睡眠(4)→睡眠(2)→睡眠(3)
	悪化 → 改善 (回復)	• 運動(2)→運動(1)→運動(4)

次に、複数の介入の組み合わせを見たものが、このページです。全ての介入の組み合わせをプロットするわけにもいかないので、幾つか代表的なものをピックアップして掲載しています。上から「改善系」と「悪化系」、「途中で悪化したものと回復したもの」の三つに区分して、そのパターンを掲載しています。

単純比較ではばらつきが大きいが、共変量を加味した推定と仮想シナリオ分析により、介入パターン間の公平な比較が可能になる



こちらが結果です。単純比較だとばらつきが多くて、どのパターンがよいか、判断がなかなかしにくいのですが、調整後では構造が明確になったと思います。特にグラフのオレンジ線、調整後ですね。複数の習慣を同時に改善するパターンが、単一習慣の改善と同様に低リスクであって、複数改善の効果が示唆されると思います。一方で、改善後に悪化するリバウンド型は、調整後でもリスクはやや高くて、行動の持続が重要であることが示せました。

今回、解釈が難しいと思って掲載するか悩んだものとして、左から五つめの喫煙と、右から三つめの喫煙のところを見ていただくと、直感的には禁煙を継続した方がいいはずなのですが、調整後だと、禁煙後に悪化した方がアウトカムが低い結果になっています。これは、ベースラインの健康状態や、同時期における他の習慣の影響などを見ていく必要があるのではないかと考えていますが、今回の分析においては、この違いが何によるものなのか。何を調整できていないがゆえにこの結果になってしまったのかといったところの深掘りができていません。もう一つ、右から三つめのこのパターンは、件数が結構少なく、モデルの精度も低かったことも影響していると思います。

- g-formulaを用いた時間依存性CATEにより、特定パターンとその他のパターン間での効果のばらつきを可視化
- 単一期間では見えない、**複数期間にわたる介入パターンの影響**を定量的に把握できる

算式：各パターンのCATE

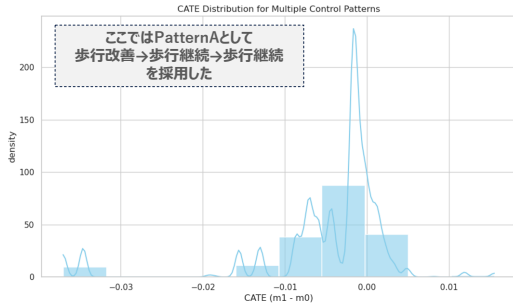
$$CATE_i = \hat{Y}_i(Z = 1) - \hat{Y}_i(Z = 0)$$

- $Z$ : 介入パターン (1: Pattern A, 0:Other)
- $\hat{Y}_i(Z) = \hat{f}_i(X)$  ...g-formulaを用いたモデルによる予測値

解釈

同じ特徴量を持つ個人が、パターンAを選択した場合と、  
その他のパターンを選択した場合のアウトカムの差

- $CATE > 0$ : Pattern Aでアウトカムが増加 (望ましくない)
- $CATE < 0$ : Pattern Aでアウトカムが減少 (望ましい)



- (1) **単一期間との違い**
  - 単一期間：ある時点での介入 vs 非介入の効果
  - 時間依存性：複数期間にわたる介入パターンの影響を考慮し、より現実的な因果推定が可能
- (2) **パターンA vs その他の意味合い**
  - 「特定の改善パターン(例：継続改善)を選択した場合、他のパターンと比べてどれだけリスクが変わるか」を個別に評価できる
  - 実務的には「どのパターンを推奨すべきか」の判断材料になる
- (3) **ITRとの関係性**
  - ITRは「誰にどの介入を割り当てるか」を決めるルール
  - CATEは「特定パターンの効果の異質性」を示すため、ITR設計の背景情報として有用
  - 直接的なアルゴリズム連携はないが、**ITRのターゲティング精度を高めるための知見**になる

続いて、CATE と ITR に移っていきたいと思います。算式は単一期間と同じですが、複数期間の場合は、特定のパターンを選んだ場合とその他のパターンを選んだ場合のアウトカムの差を表しています。特定のパターンとして今回は、「歩行の改善、継続、継続」というパターンを、注目したい介入パスとして設定しています。CATE が負の場合が、このパターンでアウトカムが減少するという意味になります。ヒストグラムを見ますと、単一期間と違ってばらつきがあって、個人ごとに効果が異なることが分かります。繰り返しになりますが、ここでの CATE 分析は、特定の改善パターンを選んだ場合、他のパターンと比べてどれだけリスクが変わるかということ、個別に、個人ごとに見ているものです。これは実務的に非常に重要だと思っていて、個人ごとにどのパターンを推奨すべきかを判断する材料になりえると思います。

- ITRルールを適用することで、ランダム割り付けよりも平均アウトカムを低減できることを確認
- 個別最適化された介入パターンの選択が、全体のリスク低減に寄与する可能性を示した

方法	平均アウトカム
ITR	0.0609
ランダム割り付け	0.0679
参考：単一介入（歩行改善）	0.0633

■ アウトカムの関係性

ITR < 単一介入 < ランダム割り付け  
0.0609 < 0.0633 < 0.0679

■ 解釈・示唆

- ITRはランダム割り付けよりも約10%リスク低減、単一介入に比べわずかに優れている
- 個別化ルールに基づく介入は、標準的な一律介入やランダム割り付けよりも効果的
- 【実務観点】**ITR導入により、ROI改善やリスク低減の両立が期待できる

ITRパターン選択率	選択率
運動改善継続：改善→継続→継続	61.9%
睡眠改善→運動改善→歩行改善	11.9%
運動改善→歩行改善→睡眠改善	9.9%
睡眠改善継続：改善→継続→継続	9.8%
その他	5%未満

■ 選択率とは？

- 各個人に対して「予測リスクが最も低いパターン」を選び、そのパターンが全体でどれくらい選ばれたかを示す割合

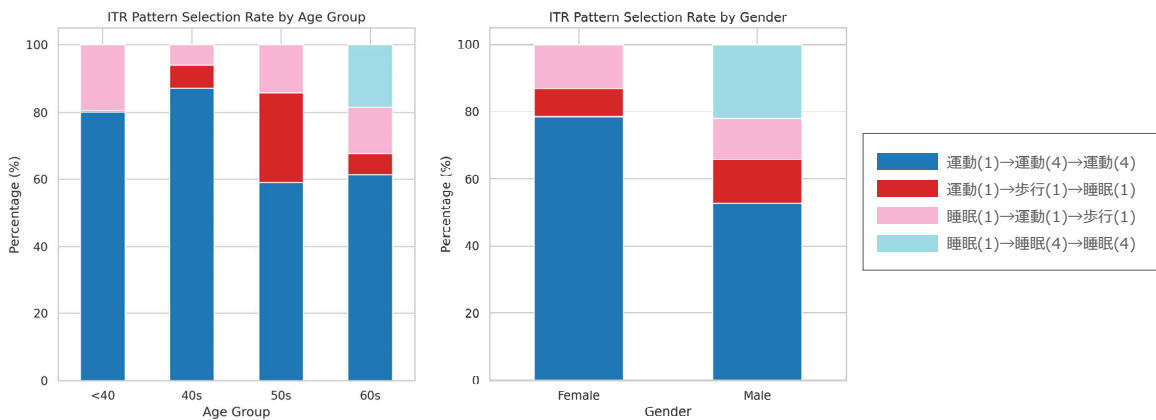
■ 解釈・示唆

- 「運動改善継続」が61.9%で圧倒的に多く、最も多くの人にとって最適なパターン。
- その他は5%未満（パターン数としては10個）→ 特定のサブグループにしか適さない
- 多くの人に共通する「標準施策候補」と、特定層向けの「オプション施策」を分けた設計を考慮する必要  
→ ITR選択率は、どの介入を標準化し、どれを個別対応にするかの意思決定材料

次に、この結果を用いて ITR を設定していきたいと思います。このページで言う ITR は何かというと、CATE によって推定された個別の介入効果を基に、各個人にとって最もアウトカムのリスクが低下する介入パターンを選択した場合のシナリオを意味しています。それが下にある「ITR パターン選択率」というもので、6割程度が、運動改善継続が最もアウトカムが下がる介入のパターンになっています。要するに、各個人にとってアウトカムが最も低下するパターンを採択させたものが ITR になります。結果としては、介入をランダムに割り付けるものに比べて、平均アウトカムを約 10%程度低減するという結果が得られました。参考として、先ほど見ていた歩行の改善と比べても、ITR の方が優れています。これは、個別に最適化した介入パターンの選択が、全体のリスクの低下に寄与する可能性を示していると思います。

最後に実務への接続ですが、事前にどのように ITR のパターンを割り当てるのかという点ですが、体況や生活習慣、いわゆる共変量を基にして、既存のモデルから予測することができます。「既存のモデル」と言っているのは、g-formula によるモデルの推定を活用することで、実務においても個別化介入の実装が現実的に可能だと思っています。

- ITRの選択傾向は年齢層・性別で異なり、介入パターンの最適解は一律ではない(施策は一律ではなく属性別に最適化が必要)
- 若年層は運動改善中心、高齢層や男性は睡眠改善や複合改善の重要性が高い



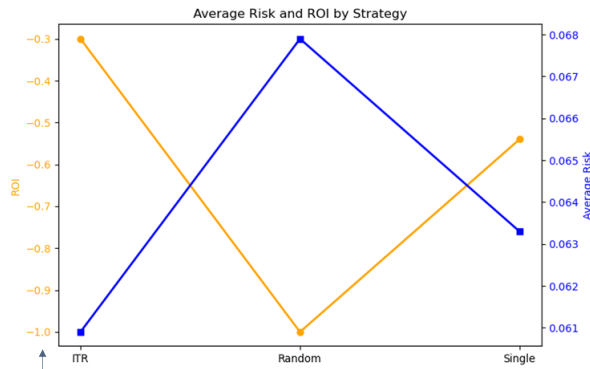
※ 本分析では、実務上アクション可能な改善・継続パターンのみを対象とするため、悪化や悪化継続のパターンは除外しています。

ここでは、先ほど個別ごとに、個人ごとにどの介入パターンが最適かということでポートフォリオを組みましたが、もう少し粗く属性別に、年齢と性別でどの介入を推進していくのがよいのかを見たものになっています。

左側の年代別で見ますと、40代では運動改善継続が中心ですが、高年齢になるにつれて、睡眠改善や複合改善の組み合わせの割合が高まってきている。性別で見ますと、男性は睡眠改善・複合改善の選択率が高い一方で、女性は運動改善の継続が効果的であると。これらは、あくまでアウトカムを最小化するという観点で、この条件において個別にどの介入パターンが最も効果的かということで選択しているものすぎないので、実際の保険会社で言うと満足度や継続率などは考慮していないので、その点は、分析の際に含める必要があると思います。

この結果から、施策の設計において、一律に同じ介入を推奨していくのではなく、例えば今回のように性別・年齢別であれば、介入を全体に打つにしても、まずはこのような介入を打ってみることが効果的であるという、介入の標準化を検討する際の題材としても、このような分析が効果的だと思っています。

個別最適化（ITR）による介入は、費用対効果とリスク低減の両面で最も優れている



**注意：**  
 今回のROI試算は、仮定で設定した介入コストおよび疾患時コストに基づく結果であり、ROIがマイナスとなっています。  
 これはITRの設計がROI向上に寄与しなかったことを意味するものではありません。

**結果**

ROIはITRが最も高く、ランダム割り付けが最も低い  
 (リスクは前項で示した通り)

**今後の検討内容**

- ROIは仮定条件下でマイナスだが、ITRは他戦略より効率性が高い
  - 介入コストの調整により、ROIをプラスに転換できる可能性がある
- ↓
- 実務的には「ITR+コスト最適化」が最も合理的な戦略  
 → ROI改善のための介入コストシナリオ分析(介入コストを変動させ、ROIがプラスとなる閾値を試算)の実施

では、セクション3の最後になります。ROI 観点で ITR を評価しています。この結果から、ITR による個別最適の介入は、リスク低減の面だけではなく、ROI の観点でも最も高い効率性を示すことが確認できます。グラフを見ますと、ITR が平均リスクが最も低いことと、ランダムに介入するとリスクが高いということは先ほどのページでもご説明したとおりです。

今回、ROI を算出する基になっている数字は私が仮置きしていたもので、単一期間ではプラスだったのですが、複数期間でマイナスになってしまって、このあたりをプラスにするためにごちゃごちゃいじるのもと思って、このまま掲載させていただきましたが、あくまで ITR が他の戦略よりも高い水準にあって、効率性の高さを示唆している点をご確認いただければと思います。繰り返しになりますが、仮のコストを置いているので、実際の施策では ROI がマイナスになることを表した結果ではございません。コストの前提やコスト構造というものを取り入れることで、もちろんプラスに転換できる可能性もありますし、各社の実際の条件に応じて分析可能だと思っています。

単一期間の前提ではありますが、Appendix に介入コストを動かしたときの ROI の関係性をシミュレーションした結果も併せて載せていますので、後ほどこちらも見ていただければと思います。事前にもそうですし、既に打っている施策自体の評価として、今後どこまでコストをかけて回収していくべきなのか、その改善余地があるのかということを試算する上でも、今回の手法は適用可能だと思っています。

分かったこと

- **短期では見えない効果が経過を考慮することで顕在化**  
→ 継続的な介入により、リスク低下が累積的に進む
- **時間依存性を考慮することで推定の精度が向上**  
→ g-formulaにより、現実的なシナリオを再現可能
- **複合介入の効果を評価できる枠組みを構築**  
→ 実務に近い意思決定に活用可能

課題点

- **介入の継続性と施策強度の定義が不十分**  
→ 実際の施策では頻度・強度が異なるため、モデル精緻化が必要
- **施策コスト・便益の時間的変動を反映できていない**  
→ 実務では年ごとにコスト構造が変化

POINT

複数期間での分析により、短期では見えない効果と戦略の持続性を定量化できる

最後に、複数期間におけるまとめになります。g-formulaを導入することで、ある時点の前の状態を加味したうえで注目したい介入パスの効果を推定できたことが、今回の営みのポイントだと思います。これによって、注目したい介入一つだけではなく、複数の注目したい介入のパターン間での比較を行うことができるので、どの施策を、どの順で推進していくべきかを判断する上での材料になりえると思っています。一方で課題は、単一期間と同様ですが、実務上での前提条件やコストの構造の反映は、今回は一切行っていないというところがありますので、それらをモデルに組み込んでいったり、解釈をどのようにしていくかなどが、今後の課題になっています。

## まとめ

# Section 4

### まとめ

45

因果推論を活用した分析は、付帯サービス評価以外にも、保険会社のバリューチェーン全体で戦略判断の質を高めるフレームワークと考える

#### 今回の分析から

##### ■ 因果推論手法の価値

- ✓ 単純比較ではなく、共変量調整+シナリオ分析により、介入効果を公平に評価可能
- ✓ このアプローチは、付帯サービス評価に留まらず、UW(引受)やBE推定など保険の他領域にも展開可能

##### ■ データ活用の進化

- ✓ 疾患率だけでなく、複合的な指標(医療費,継続率,満足度など)で評価する必要性
- ✓ 社内データ+レセプト連携により、精緻かつ実態に即した分析が可能

##### ■ バイアス調整の重要性

- ✓ 生体データやウェアラブルデータの普及により、個人レベルで詳細な情報が得られる時代において、保険会社も健康・医療データを活用し、背景因子の調整を行うことが重要

#### 今後の課題・検討テーマ

##### ■ ROIの多次元評価

- ✓ 介入コストによるROIだけでなく、疾患リスク×顧客満足度や継続率など統合した総合評価指標の設計

##### ■ ITRの高度化

- ✓ 本分析ではリスクベースの最適化を行った。今後はコスト・行動特性・実行可能性を加味した「現実的ITR」の検討が必要

##### ■ シナリオ分析の拡張

- ✓ 仮想介入だけでなく、複数戦略の比較や「介入順序の最適化」など、より複雑な意思決定支援

では、最後に、全体のまとめになります。単一期間と複数期間、それぞれの得られた示唆としては、それぞれのセクションの最後のまとめのページに記載していますので、ここではそもそもの因果推論の実務への応用という観点で、幾つか記載させていただきました。特に最後のバイアス調整の重要性ですが、個人的に注目というか、必要性を強く感じているところでして、個人で端末などを含めてデータの取得が容易になってきていると思う一方で、得られているデータ自体を用いて個人がアクションを起こす、管理するという事は、まだまだ一定のハードルがあると思っています。

一方で、これは単に保険会社だけではなく、例えばヘルスケア業界などもこのあたりは関わっているところで、ゆくゆくは個人がデータを基にしてアクションを取りえる時代が、それほど遠くない時代に来る

だろうと考えると、より一層のオーダーメイドやカスタマイズ化されたサービスの提供も、保険会社としては検討を行っていく必要があると思ひまして、こちらに記載させていただきました。

課題としては、これまで記載した ROI や ITR のコストや行動の特性。特に行動の特性に関しては、先ほどご質問でもいただきましたが、例えば当社で言えばアプリを展開していますが、そのアプリ上のログを取り入れたり、メールのクリックやオープンなど、施策への顧客の反応を取り入れることで、より実態に即した ITR の設定や評価が可能になると思ひます。

最後になりますが、私自身、因果推論自体が非常に好きというか、注目して見ている手法でして、理由としては、機械学習モデルと違って効果が推定できる点が、このような施策を検討する、決めていく上での意思決定に非常に重要な材料となりえるということで、非常に注目して見ているものになります。皆さんの中にも私より詳しい方がいらっしゃると思ひますので、興味を持っていただいた方は、ぜひお声がけいただければ幸いです。スライドがかさんでいて、かなり早口で話してしまつて、お聞き苦しい点があったと思ひますが、発表としては以上になります。ご清聴いただき、ありがとうございました。

#### 参考文献

46

[1]Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *\*Journal of the American Statistical Association\**, 113(523), 1228–1242.

[2]Qian, M., & Murphy, S. A. (2011). Performance guarantees for individualized treatment rules. *\*The Annals of Statistics\**, 39(2), 1180–1210.

[3]Robins, J. M. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—Application to control of the healthy worker survivor effect. *\*Mathematical Modelling\**, 7(9–12), 1393–1512.

[4] 板倉兼介,尾上辰徳.(2023)『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』,2023年度アクチュアリー-年次大会

[5]金本 拓(2024)『因果推論: 基礎から機械学習・時系列解析・因果探索を用いた意思決定のアプローチ』,オーム社

司会 ありがとうございます。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。では、お願いいたします。

勝野 技術的な話を全くしなくて、この先の話なのですが、これはパーソナライゼーションの話だと思うのですね。今、例えば YouTube でも Spotify でも、僕が何かを見ると、「あなた、これ好きですよ」と薦めてくれる。それは、僕が動画を見たい、音楽を聴きたいということがあるので、お薦めどおりにやるのですが、運動や禁煙は、どちらかというやりたいものではないではないですか。パーソナライゼーションの中で、どのように介入していくのか、どのようにモチベーションをつけていくのかということ自体が、一つの重要なポイントになってくると思ひます。因果推論の効果を出すために。それはどのようにすればいいのか、何かお考えがあれば教えていただけないかということが一つです。

もう一つは、先ほどのスライドの第2部のところで、調整後にすることによって、ぴよんと飛び出ていたものが消えたということについては、これは非常にうまくいっていると思うのですね。この前のところでは調整後にすることで飛び出ていますが、なぜきれいなものが、このように飛び出るということが起こるのか。逆にこれは、元データに入っていない変なことが起きていませんか。例えば、非常に健康状態が悪くて改善しなければいけない人が、頑張ってみたけれどもうまくいかずに、悪い結果が出てしまったというような、変な群団が入っている影響がないかどうか、解釈のところについて教えてもらえればと思います。

尾上 ありがとうございます。一つめのレコメンという観点で言うと、今回は、あくまでアウトカムが下がるということで、設計のところにある ITR のようなものは選択しています。レコメンという観点では、保険会社として与えられるものとしてはインセンティブですかね。例えばアプリ上でランキングやステージなどを付与したときに、顧客がどのように反応したかというようなログがあれば、それを組み込むことでレコメンとしてのパターンの選択ということで、一定、カスタマイズ化できるのではないかと考えています。実際の行動ログなど、会社として行うインセンティブの付与などの効果を共変量という形で入れ込むことによって、推定ができるのではないかとということが、1点目に対しての回答になります。

2点目は、これは逆に跳ねていて、こちらは下がっていてということですが、飲酒の全て改善というパターンに関しては、「このパターンを見たいです」として、得られている対象者全員に飲酒の改善、喫煙の改善、睡眠の改善というものを与えています。ですから、元々飲酒をしていない、たばこを吸っていない、睡眠もぐっすり眠れているという人に関しては、効果がゼロとして出てきます。逆に言うと、こちらの良好な人から悪化してというパターンは、元々良好でずっといる人に対して「悪化しなさい」というパスを与えているので、通常の悪化の人に比べて、より影響としては跳ねているところがあります。

そこは、載せていませんが、推定区間を併せて表示すると、特に改善の中断に関しては、他に比べると幅としては結構あって、元々体況がいい人に悪化を促すと推定自体がブレるということで、バイアスというのですが、先ほどのこのようなものが含まれているのではないかとこの観点で言うと、良好な人に無理にたばこを吸わせて、飲酒させてという効果に変に出てしまった結果であると考えています。

勝野 ありがとうございます。

司会 会場で他に質問のある方は、挙手をお願いいたします。では、あと2分しか残っていませんが、最後にスライドから一つリクエストがあります。「g-formula というモデルがどのように構築されているか、補足説明をお願いできないでしょうか」というご質問が来ております。

尾上 繰り返になってしまうかもしれませんが、推論フェーズでは、今回は「運動の改善、継続、継続」を見たいのだという、それを対象者全員に一律に与えます。それをどのように推定するかということ、学習フェーズで、まず観察データから共変量の分布やアウトカムをモデルとして学習させているフェーズになっています。ですので、2段階の推定になっています。この学習フェーズにおける、特に「観察データでLモデルを学習」ですが、共変量が増えれば増えるほど、対象集団も散らばりがないほど精度としては下がってしまうので、その分、推論フェーズでは、精度の観点から、先ほど言った跳ねや、あまり効果が得られない、想定とは逆の結果が起きるとは思っています。

質問に対して話がそれてしまいましたが、回答としては、学習フェーズで観察データからまずモデルの学習をさせたうえで、推論フェーズで、このモデルを使って実際にわれわれが見たい介入パスの推定を行っていくという、2段階の推定モデルになっています。すみません、冗長に余計なことを話してしまった気がするのですがすけれども。

司会 ありがとうございます。では、もう少し質問はあったのですが、時間もちょうど来ましたので、これで今回のセッションを終わります。有益なお話をありがとうございました。

尾上 ありがとうございます。

---

Appendix

---

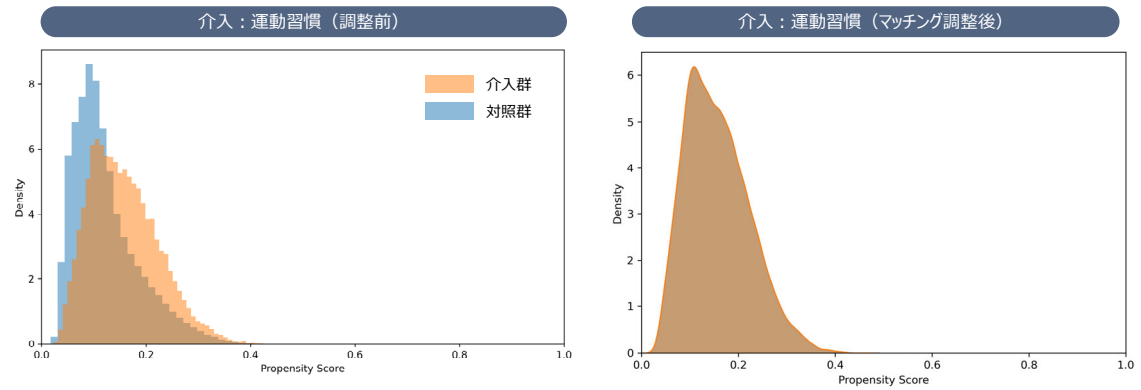
Section 5

- 傾向スコア分布(調整前後)
- SMD
- 問診項目一覧
- 禁煙介入の感度分析
- CATE特徴量～walking～
- 費用対効果のシミュレーション～walking～
- DeSCデータ補足

## 傾向スコア：単一期間

## 傾向スコア分布

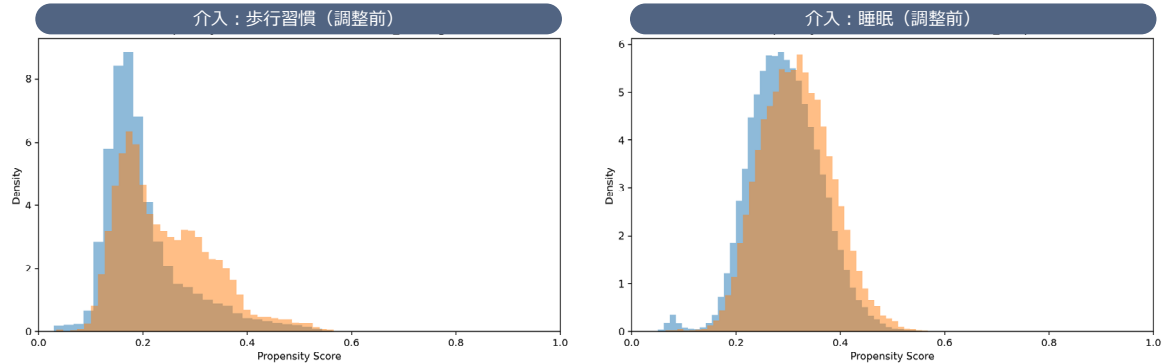
- 単一期間：運動習慣
- 介入群 = 改善群、対照群 = 悪化継続群



以降、他の介入におけるマッチング後の傾向スコア分布は調整前のlapしている分布となるため記載を割愛

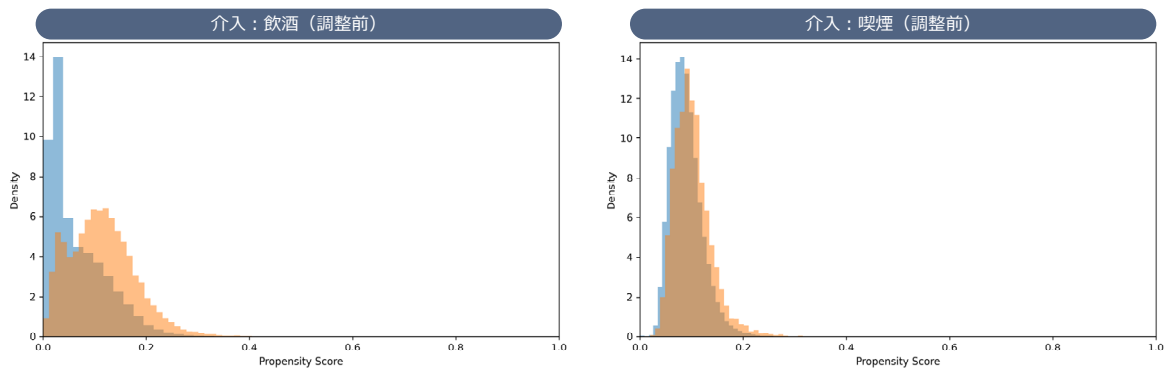
傾向スコア分布

- 単一期間：歩行習慣（左図）、睡眠（右図）
- 介入群 = 改善群、対照群 = 悪化継続群



傾向スコア分布

- 単一期間：飲酒（左図）、喫煙（右図）
- 介入群 = 改善群、対照群 = 悪化継続群



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

項目	単純比較			マッチング調整後		
	介入群平均	対照群平均	SMD	介入群平均	対照群平均	SMD
age_group	52.89	51.37	0.165	52.89	52.81	0.009
sex_code(※1)	1.24	1.19	0.116	1.24	1.24	0.004
payer_type(※2)	0.34	0.42	0.175	0.34	0.34	0.004
Bmi	23.12	23.35	0.057	23.12	23.09	0.008
abdominal_circumference	84.29	84.75	0.038	84.29	84.20	0.008
systolic_blood_pressure	127.97	128.11	0.008	127.97	127.92	0.003
diastolic_blood_pressure	77.90	78.29	0.033	77.90	77.95	0.004
fasting_triglyceride	141.03	150.19	0.079	141.03	142.15	0.011
hdl_c	58.02	56.66	0.083	58.03	58.24	0.012
ldl_c	120.18	118.96	0.038	120.18	120.42	0.008
Got	24.17	24.66	0.035	24.09	24.40	0.023
Gpt	24.49	25.61	0.055	24.32	24.64	0.018
gamma_gt	48.90	55.11	0.093	48.90	49.55	0.011
Hba1c	5.08	5.11	0.020	5.08	5.11	0.018
urinary_glucose_code(※3)	1.17	1.21	0.047	1.17	1.16	0.010
urinary_protein_code(※3)	1.27	1.27	0.003	1.27	1.26	0.014

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

※1)男性=1、女性=2 ※2)国保=0、健保=1 ※3)1 (-)、2 (±)、3 (+)、4 (++)、5 (+++)

項目	単純比較			マッチング調整後		
	介入群平均	対照群平均	SMD	介入群平均	対照群平均	SMD
age_group	53.76	52.56	0.133	53.76	53.74	0.003
sex_code(※1)	1.54	1.51	0.058	1.54	1.54	0.008
payer_type(※2)	0.31	0.38	0.156	0.31	0.31	0.000
Bmi	23.00	23.18	0.047	23.00	22.98	0.006
abdominal_circumference	83.34	83.95	0.053	83.34	83.29	0.004
systolic_blood_pressure	127.68	127.30	0.022	127.68	127.65	0.002
diastolic_blood_pressure	76.79	77.00	0.019	76.79	76.75	0.003
fasting_triglyceride	122.90	124.55	0.017	122.90	123.02	0.001
hdl_c	62.63	61.62	0.058	62.63	62.64	0.001
ldl_c	122.78	122.17	0.020	122.78	123.13	0.011
Got	23.95	23.81	0.010	23.91	23.88	0.003
Gpt	23.33	23.86	0.029	23.31	23.39	0.004
gamma_gt	38.41	40.17	0.034	38.40	38.59	0.004
Hba1c	5.17	5.14	0.023	5.17	5.17	0.002
urinary_glucose_code(※3)	1.12	1.13	0.010	1.12	1.12	0.001
urinary_protein_code(※3)	1.20	1.22	0.022	1.20	1.20	0.002

※1)男性=1、女性=2 ※2)国保=0、健保=1 ※3)1 (-)、2 (±)、3 (+)、4 (++)、5 (+++)

問診項目①

54

問診項目一覧(項目名、質問・回答)

項目	質問内容・回答
既往歴 1 脳血管コード	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
既往歴 2 心血管コード	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
既往歴 3 腎不全HDコード	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
貧血コード	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
喫煙コード	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
体重変化 2 0歳からコード	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
運動習慣 3 0分以上コード	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
歩行又は身体活動コード	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
歩行速度コード	問診の質問：ほぼ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
体重変化 1 年間コード	問診の質問：この 1 年間で体重の増減が±3 kg 以上あった 回答：1:はい 2:いいえ

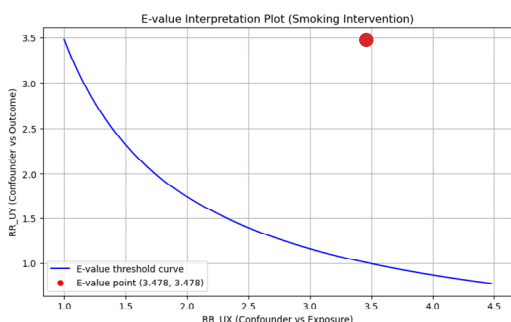
問診項目②

55

問診項目一覧(項目名、質問・回答)

項目	質問内容・回答
食べ方 1 早食い等コード	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 3:遅い
食べ方 2 就寝前コード	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
食べ方 3 夜食間食コード	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
食べ方 3 間食コード	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
咀嚼コード	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
食習慣コード	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
飲酒コード	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
飲酒量コード	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(約 500ml)、焼酎 3 5 度(80ml)、ウイスキーダブル一杯(60ml)、ワイン 2 杯(240ml) 回答：1:1 合未満 2:1~2 合未満 3:2~3 合未満 4:3 合以上
睡眠コード	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
生活習慣の改善コード	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:既に改善に取り組んでいる(6ヶ月未満) 5:既に改善に取り組んでいる(6ヶ月以上)

観点②：喫煙の感度分析・・・本紙p.20



- $RR_{UX}$  : 未測定交絡(U)が曝露X(禁煙施策を受けたか否か)に与える影響の強さ  
→もし「禁煙意欲の高さ」などという未測定因子があった場合、それが施策を受ける確率をどの程度高めうるか
- $RR_{UY}$  : 未測定交絡(U)がアウトカムY(3 CI)に与える影響の強さ  
→「禁煙意欲の高さ」が3CI発生をどの程度高めうるか

■ E-valueとは？  
未測定交絡がどの程度強ければ観察された効果を説明できるかを示す指標

$$E-Value = RR + \sqrt{RR \times (RR - 1)}$$

ただし、RR：効果推定値

■ どう解釈した良いのか？  
この値が大きいほど、未測定交絡が強くなければ観察結果を覆せないことを意味する

- E-value ≈ 1.0 → 結果は脆弱（弱い交絡で覆る）
- E-value ≥ 2.0 → 中程度以上の交絡が必要（ある程度頑健）
- E-value ≥ 3.0 → 非常に強い交絡が必要（かなり頑健）

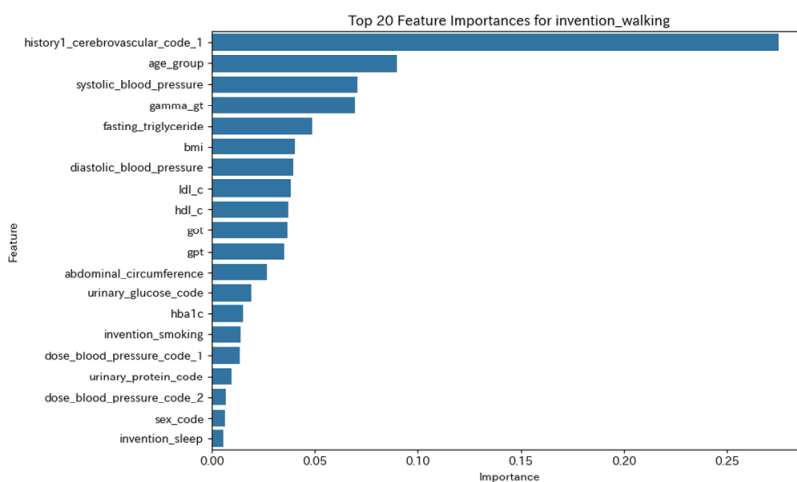
■ 今回の結果

- 未測定交絡が曝露・アウトカム双方に3.5倍以上の影響を与えない限り、結果は覆らない。
- これは実務的には、これほど強い交絡は稀＝結果はかなり頑健と解釈
- つまり、**喫煙の調整前後で変化がない理由は他にある（逆因果の可能性）**  
※本分析上ではその影響まで深堀は行っておりません

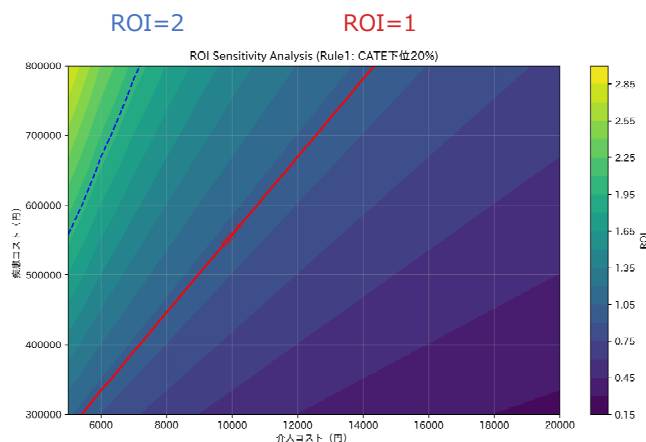
VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. Ann Intern Med. 2017;167(4):268-274.

CATE特徴量～walking～

- 歩行習慣改善に関してCATEの推定を行い、「どのような特徴を持つ人に介入効果が大きいか」を評価。
- その後、モデルの特徴量重要度（Feature Importance）を算出し、「介入効果に強く影響を与える因子」を抽出



- 疾患コストに対して、どの水準まで介入コストを設定できるかをシミュレーションで検証・・・費用対効果のシミュレーション
- 対象としたITR：Rule1=CATE下位20%(歩行習慣改善)



DeSCデータ：保険者データの提供範囲

保険制度ごとのデータ量・期間は以下のとおりです（2025年10月時点）

	会社	DeSC		
		健康保険組合 0~74歳	国民健康保険 0~74歳	後期高齢者医療制度 65歳~
	保険者 (データソース)	2012/02~2025/03	2013/04~2022/12	2014/04~2022/11
	観察可能期間	125	783	390
	ユニーク加入者数 (万人) ※2025年1月時点	158	117	93
	最大観察可能月数 (月)	○	○	○
母集団	加入者情報	×	○	○
	脱退事由 (死亡情報)	○	○	○
医療情報	医療レセプト (入院/外来、傷病、診療行為等)	○	○	○
健康診断	健康診断	○	○	○
生活習慣など	kencomライフログ	○	▲	×

※DeSCヘルスケアが提供するデータの情報に関しては、(でんわ：03-6386-3295) または、  
会員制メディア (<https://desc-life-actuary.jp/>) よりお問い合わせください。

レセプト/健診/kencom/介護データを、加入者台帳上の同一個人と紐付けて分析可能



\* kencomデータは健保・国保・契約当該自治体の住民が対象 ただしレセプト・健診・台帳と結合できるのは保健事業の契約保険者に限る  
\*\* 介護データは国保・後期高齢者のうち、利用許諾が得られた自治体が対象

※DeSCヘルスケアが提供するデータの情報に関しては、(でんわ：03-6386-3295) または、  
会員制メディア (<https://desc-life-actuary.jp/>) よりお問い合わせください。

生命保険業界における DeSCデータ活用の意義と可能性

課題定義 (生命保険会社が抱える共通課題)	DeSCデータが提供できる価値 (DeSCデータによる課題解決アプローチ)
<b>基礎率の精緻化の限界</b> 既存の契約データや健保データでは、年齢層や疾患領域に偏りがある。特に高齢層・慢性疾患の発症率推定が困難。	<b>全年齢の代表性と長期追跡性</b> 健保・国保・後期高齢を 気通貫でカバーし、全年齢層の疾病発生率を観察可能。
<b>引受査定モデルの説明変数設計</b> 健診情報は取得できても、医療・介護の実態が分からないため、リスク因子の正確な重みづけが難しい。	<b>医療 × 介護 × 健診 × ライフログの統合</b> 受診行動・介護移行・生活習慣を横断的に分析でき、商品設計・査定モデルに必要な「リスクの全体像」を把握できる。
<b>新商品の検討に必要な疫学的エビデンス不足</b> 健康増進型保険や医療・介護連携型商品の開発にあたり、発症率や介護移行率を裏付ける公的データが限定的。	<b>迅速な仮説検証と再現性</b> ACTUAを活用することで、非分析者でも短時間で発生率(基礎率)・介護移行率を再集計可能。商品開発や引受査定の現場で“アータドリブンな意思決定”を実現。

※本スライドは、生命保険業界での活用例を示す目的で作成しています。  
※DeSCヘルスケアが提供するデータの情報に関しては、(でんわ：03-6386-3295) または、  
会員制メディア (<https://desc-life-actuary.jp/>) よりお問い合わせください。