

# 傾向スコア分析

## 医学統計セミナー2017 第3回目

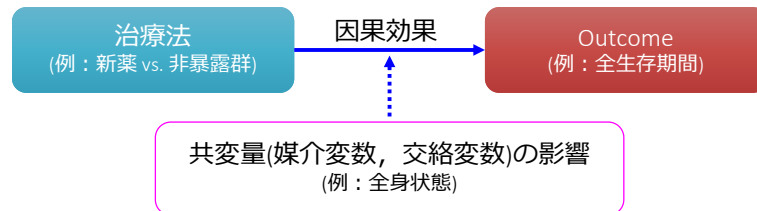
下川敏雄

医療データサイエンス学  
(附属病院臨床研究センター)

11月28日(火)	診断・検査データの解析：ROC曲線を中心に
12月22日(金)	多変量解析：重回帰・ロジスティック回帰・比例ハザードモデル
1月26日(金)	傾向スコアによる解析：マッチングによる解析を中心に
2月23日(金)	メタアナリシス：2値・量的・生存期間データに対するメタ回帰分析
3月23日(金)	臨床試験における必要症例数の計算

### 傾向スコアの動機

□ 治療法の評価 = ある医学的介入による影響(因果効果)の検証



共変量が厄介な理由：コントロールできないものが多い

検証的な側面から共変量の影響を最小限にして因果効果を評価する

- 無作為化比較試験(RCT:Randomized Controlled Trial)

観察的な側面(観察研究)において(治療選択に対する)共変量の影響を最小限にして因果効果を評価する

- 統計学的な調整法(共変量とOutcomeの線形性などの制約)

- 傾向スコア(Propensity score)の利用

### 統計的方法による共変量の調整の方法とその限界

観察研究における共変量を考慮した研究及び解析(星野・岡田, 2006)

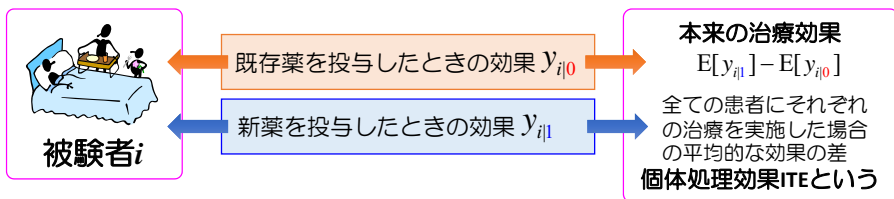
研究デザイン	(1) 均衡化：共変量の値が同じになるペアを作ることによって2つの群の被験者を構成 <ul style="list-style-type: none"> <li>- 完全に一致するペアを作るとは事実上不可能</li> <li>- 多数の共変量を含めることができない</li> <li>- ペアは観測者が決めるため、主観的で恣意性を排除できない。</li> </ul>
	(2) 恒常化・限定：共変量のある被験者のみに限定して解析を行う <ul style="list-style-type: none"> <li>- 一部の被験者に限定するため、研究の一般可能性が低くなる。</li> <li>- 多数の共変量を含めることができない</li> </ul>
統計	(3) 統計的調整(多変量解析等) <ul style="list-style-type: none"> <li>- 誤った統計的モデルを用いると誤った結果を導く可能性がある。</li> </ul>

複数の共変量の一つの変数(ものさし)におきかえ、そのもとで均衡化や層別化などを行い、誤った統計モデルを用いても「頑健」な結果をえることができる方法

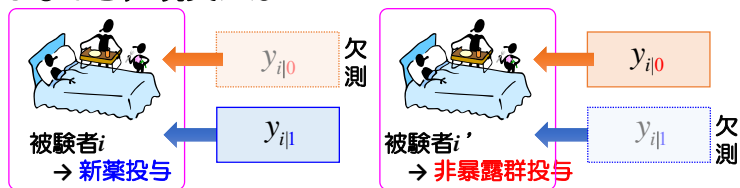


傾向スコア分析(propensity score)

## 因果効果の定義：Neyman-Rubinの反事実的モデル



しかしながら、現実には...



となり、いずれか一方は欠測であると考えられる。

## 因果効果の定義：Neyman-Rubinの反事実的モデル

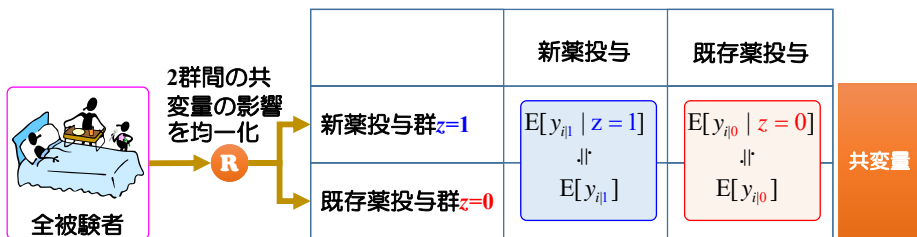
### □ 無作為化比較試験の枠組みで考える

既存薬投与群 $z=0$ ，新薬投与群 $z=1$ としたときの平均的な効果( $z$ は割付変数と呼ばれる)

	新薬投与	既存薬投与	共変量
新薬投与群 $z=1$	$E[y_{i1}   z=1]$	$E[y_{i0}   z=1]$ 欠測値	
既存薬投与群 $z=0$	$E[y_{i1}   z=0]$ 欠測値	$E[y_{i0}   z=0]$	
	$E[y_{i1}]$	$E[y_{i0}]$	

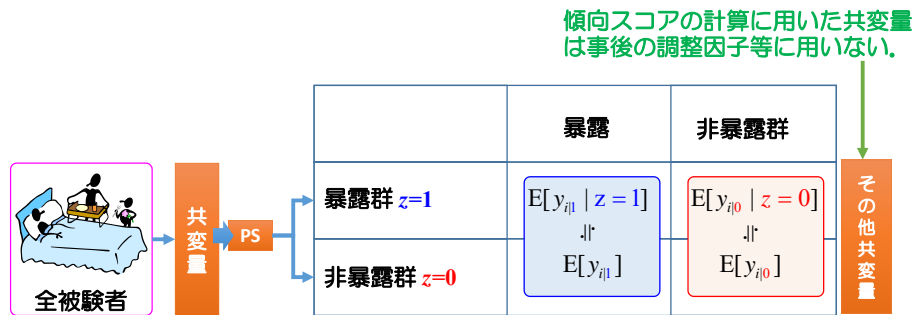
本来の因果効果  $E[y_{i1}] - E[y_{i0}]$

## 無作為化比較試験による群間の均一化



となることから、欠測と捉えられていた部分を無作為化によって補うことで、本来の因果効果を評価することができる。

## 傾向スコアで期待されること



傾向スコアを用いることで複数の共変量から1つの変数に集約できる(傾向スコアモデルに関する適切性の評価をc統計量、疑似決定係数で評価すること(評価していない論文が多いことが指摘されている(Weizen et al., 2004))).

共変量による投与群の選択性を揃えることで、処理の割付については偶発性のみに依存するようにする。これにより、群間の被験者層を揃えることができる。

すなわち、傾向スコアとは、観察研究でのインバランスな群間のバランスを行う「ものさし」となる。

## 観察研究における効果量の定義

	暴露	非暴露
暴露群 $z=1$	$E[y_{i1}   z=1]$	$E[y_{i0}   z=1]$ 欠測値
非暴露群 $z=0$	$E[y_{i1}   z=0]$ 欠測値	$E[y_{i0}   z=0]$
	$E[y_{i1}]$	$E[y_{i0}]$

ATT (Average Treatment Effect for Treated) is indicated by a blue double-headed arrow between  $E[y_{i1} | z=1]$  and  $E[y_{i0} | z=1]$ .  
ATE (Average Treatment Effect) is indicated by a pink double-headed arrow between  $E[y_{i1}]$  and  $E[y_{i0}]$ .

### ■ 平均処理効果(ATE: Average Treatment Effect) ・ 平均因果効果(ACE: Average Causal Effect)

母集団のすべての個体での暴露でのアウトカムの期待値と非暴露でのアウトカムの期待値の差

$$ATE = E[y_{i1}] - E[y_{i0}]$$

### ■ 暴露群の平均処理効果(ATT: Average Treatment effect for Treated)

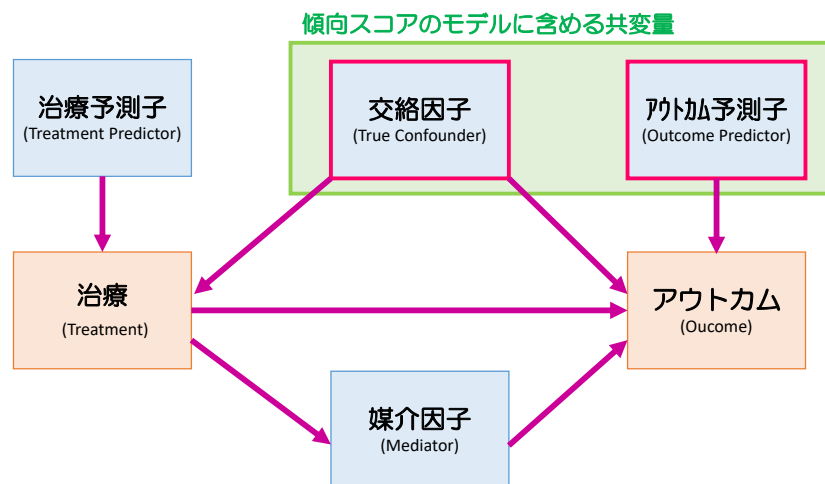
母集団のうち暴露群での暴露のアウトカムの期待値と非暴露でのアウトカムの期待値の差

$$ATT = E[y_{i1} | Z=1] - E[y_{i0} | Z=1]$$

## 傾向スコア解析の手順(Leite, 2017)

Step	目標	手順の例
1. 準備	解析可能な「完全な」データ集合を得る	<ul style="list-style-type: none"> <li>共変量選択</li> <li>欠測値に対する処理</li> </ul>
2. 傾向スコアの推定	暴露群と非暴露群に対して傾向スコアを推定する。	<ul style="list-style-type: none"> <li>ロジスティック回帰</li> <li>Random Forests</li> <li>一般化ブースティング樹木</li> </ul>
3. 傾向スコア手順の実施	暴露群と非暴露群の共変量の分布のバランスをとるための戦略を実施する。	<ul style="list-style-type: none"> <li>傾向スコア・マッチング</li> <li>傾向スコア層化</li> <li>IPW(傾向スコア重み付け)</li> </ul>
4. 共変量のバランスの評価	暴露群と非暴露群のあいだの共分散の分布のバランスが達成された程度を評価する。	<ul style="list-style-type: none"> <li>標準化された平均距離の計算</li> <li>分散比の計算</li> </ul>
5. 処置効果推定	処置効果とその標準誤差を推定する。	<ul style="list-style-type: none"> <li>重み付け平均距離</li> <li>一般化線形モデル</li> </ul>
6. 感度分析	除外された共変量が処置効果の有意性検定をどの程度変化させるかについて評価する。	<ul style="list-style-type: none"> <li>Rosenbaum(2002)の方法</li> <li>Carnegie et al.(2016)の方法</li> </ul>

## 因果関係に関する模式図(Leite, 2017)



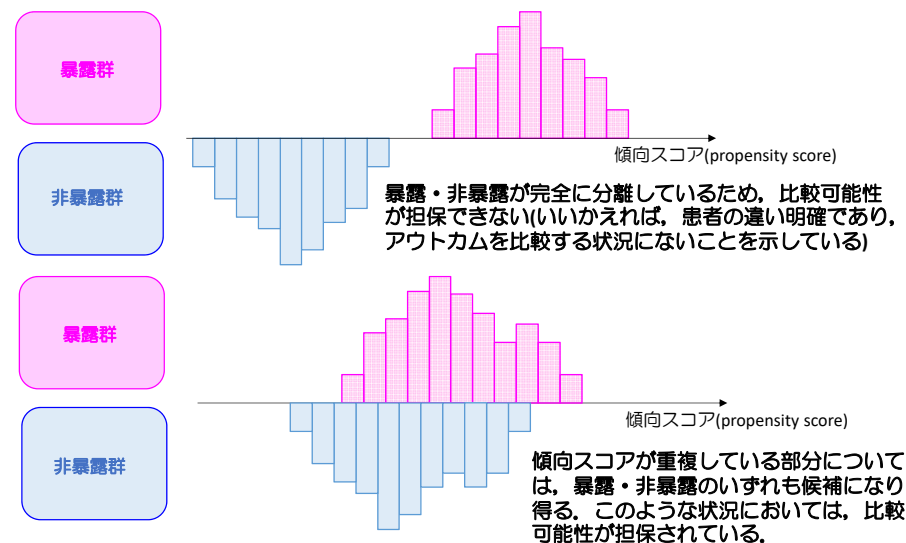
### それぞれの因子の意味(Brookhart et al., 2006)

交絡因子：処置効果の偏りとバラツキの削減に繋がる。

アウトカム予測子：処置効果のバラツキの削減に繋がる(偏りは関係ない)。

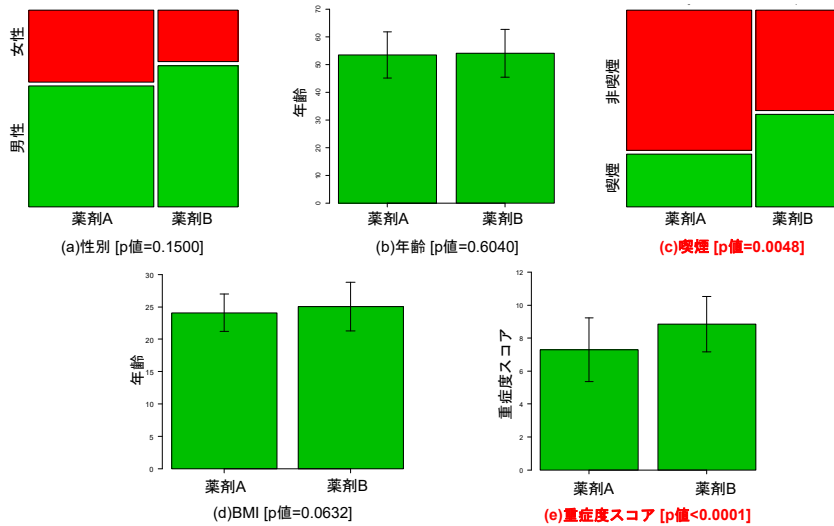
## 傾向スコアによる解析の留意点

傾向スコアとは、研究対象因子に暴露する確率で表される(つまり、傾向スコアは0~1の範囲をとる)。



## 傾向スコアの推定：疑似データ

これは、薬剤Aを投与した110例と薬剤Bを投与した71例を調査した後ろ向き研究のデータである。共変量としては、性別、年齢、喫煙の有無、BMI、重症度スコアが計測されている。



## 傾向スコアの推定

### 傾向スコアの特徴

- 条件付き独立性：傾向スコアでの割付がアウトカムに影響しない
- バランシング：同じ傾向スコアの値に対応した共変量の分布は、それぞれの群で等しい

ただし、傾向スコアを計算するのに用いた共変量の範囲でのことであり、観測されていない(あるいは、傾向スコアの計算に用いていない)共変量があり、それが交絡因子である場合には、その限りではない。したがって、傾向スコアを計算するうえで、もろいる共変量は十分に吟味しなければならない。

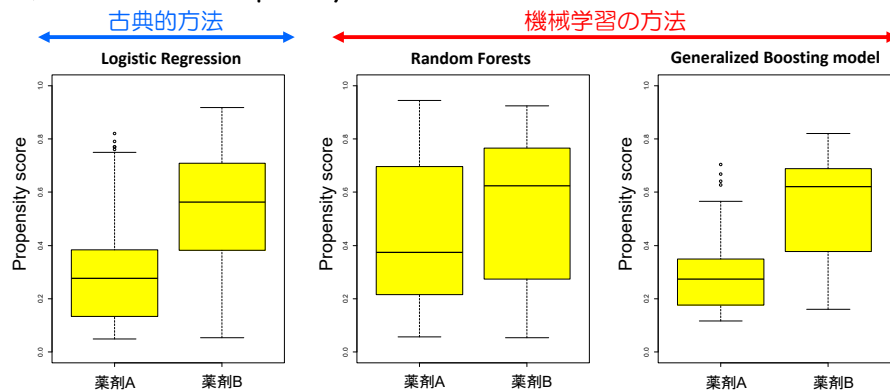
### 傾向スコア推定の方法

傾向スコア推定の方法には、ロジスティック回帰を用いる方法から、機械学習(非線形分類手法)の方法を用いる方法まで様々な方法が提案されている。

- ロジスティック回帰  
解釈が分かりやすいものの、交互作用を捉えられない可能性がある。
- 機械学習の方法(例：Random Forests, Generalized Boosting Modeling)  
交互作用を捉えられるものの、解釈ができない。

その他にも、Bayes流傾向スコア推定(An, 2010；Kaplan & Chen, 2012,2014; McCandless et al., 2009), 2段階(傾向スコア推定と実施)の同時推定(Imai & Ratkovic, 2014)などが提案されている(つまり、いっぱい提案されている)。

## 例示に対するPropensity Scoreの推定結果：3種類の方法



- Generalized Boosting model (GBM)では傾向スコア(Propensity score)があまり重複していない。

→ マッチングを行う場合には症例数が大幅に削減される可能性がある。

- 様々な方法を用いるよりも、一つのモデルを用いて、それが不適切ならば、次の方法を用いることが推奨される(Leite, 2017)。
- 適切性には、c指標(c-index)を用いて、0.8程度であればよいとされる(星野・岡田, 2006)。
- 第1選択は、ロジスティック回帰を用いることである(岩崎, 2016)。

## 傾向スコア手順の実施

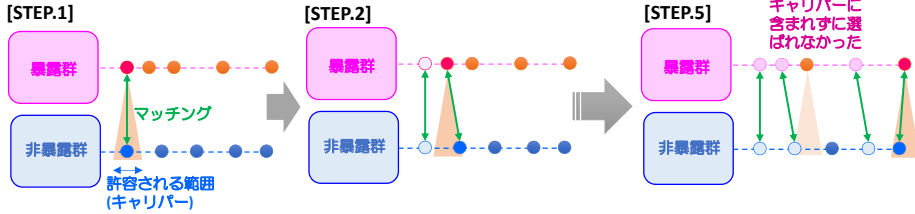
- マッチング  
傾向スコアの一致した(あるいは極めて近い)個体同士を選択する。
- 層別  
傾向スコアを用いて層別(5層程度)したうえで層別解析を行う。
- 逆確率重み付け(IPWE: Inverse Probability Weighting Estimator)  
傾向スコアを用いて層別(5層程度)したうえで層別解析を行う。
- 共分散分析  
傾向スコアを共変量にとった共分散分析を実施する。

これらの手法はいずれかを使うということではなく、組み合わせて用いることもできる。例えば、マッチング後に共分散分析を行うことが考えられる。

実際の応用分野では、マッチングあるいは逆確率重み付けを用いることが多いようである。

# マッチング

## 強欲マッチング(Greedy matching)

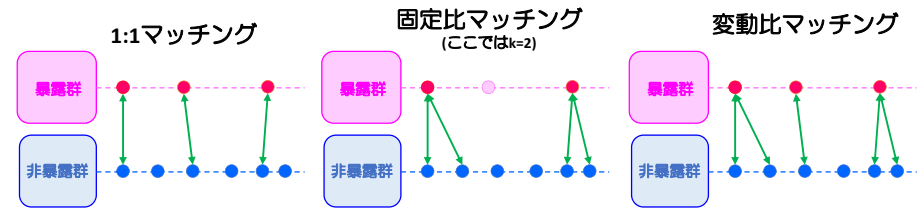


暴露群の任意の個体で、傾向スコアが最も近い非暴露群の個体を逐次に探索する方法である。強欲マッチングには様々なバリエーションがある

### ● 復元マッチング vs. 非復元マッチング

- 復元マッチングでは、暴露群の異なる個体に対して同じ非暴露群の個体を対応させることを許容し、非復元マッチングでは許容しない。
- そのため、バイアス低減の点では、非復元マッチングのほうが優れているが、復元マッチングは殆ど用いられていない。
- 非復元マッチングでは選択する個体の順番の影響を受ける(強欲マッチング以外を利用)。
- マッチング数(例数)が少ない場合には、復元マッチングが推奨される(Rosenbaum, 1989)。
- 一方で、マッチング数が多くなるにつれて、二つの差はほとんどなくなる。
- 岩崎(2015)では、1:1非復元マッチングが推奨されている。

## ● 1:1マッチング vs. 固定比マッチング(1:kマッチング) vs. 変動比マッチング



### ■ 1:1マッチング[one-to-one matching]

- 1名の暴露群と1名の非暴露群でマッチングする方法である。
- 例数の減少が最も顕著であるが、群間の例数の不均衡が起こらない。したがって、検出力の低下が最も少ない(Cohen, 1988)。

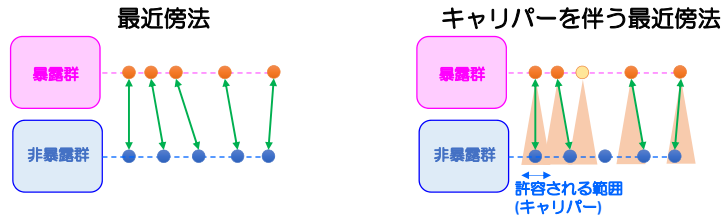
### ■ 固定比(1:k) マッチング[fixed rate matching, one-to-k matching]

- 1名の暴露群とk名の非暴露群でマッチングする方法である。
- 最も推奨されないマッチング法である(Leite, 2017)。

### ■ 変動比マッチング[variable rate matching, one-to-many matching]

- 1名の暴露群と複数(個体ごとに変動, 上限のみ設定)の非暴露群でマッチングする方法
- 暴露群の例数が非暴露群の例数よりもかなり少ない場合には有効(Leite, 2017)。
- 変動比マッチングは1:1マッチングに比べてバイアス除去に優れている(Cepeda et al, 2003; Gu & Rosenbaum, 1993; Ming & Rosenbaum, 2000)

## ● 最近傍法 vs. キャリパーを伴う最近傍法



### ■ 最近傍マッチング [Nearest neighbor]

暴露群に一番近い傾向スコアを持つ非暴露群をマッチングする方法である。

### ■ キャリパーを伴う最近傍マッチング[Nearest neighbor within caliper]

- マッチングさせる許容領域(キャリパー)を設定した最近傍マッチング(キャリパーを大きくするほどマッチングされる個体は増加する)。
- Rosenbaum & Rubin(1985)は $0.25 \times SD$ をキャリパーに設定することを推奨している( $0.20$ を推奨する報告もあるため、 $0.25 \sim 0.20$ を推奨している)。

キャリパーを設定することが一般的である。

## 最適マッチング(Optimal matching)

ID	PS	ID	PS
T1	0.46	C1	0.28
T2	0.92	C2	0.12
T3	0.89	C3	0.68
T4	0.63	C5	0.26
T5	0.72	C5	0.51
T6	0.32	C6	0.31
T7	0.28	C7	0.48

### 最適マッチング

ID	PS	ID	PS	距離
T7	0.28	C2	0.12	0.16
T6	0.32	C5	0.26	0.06
T1	0.46	C1	0.28	0.18
T4	0.63	C6	0.31	0.32
T5	0.72	C7	0.48	0.24
T3	0.89	C5	0.51	0.38
T2	0.92	C3	0.68	0.24
				合計=1.58

最適マッチング(Optimal matching)とは、PSの距離の合計値が最小になるようにペアを作る

- 暴露群と非暴露群の標本サイズに違いが大きいとき、強欲マッチングに比べて顕著に優れている。
- 標本サイズが大きい場合には、強欲マッチングとの差は小さい。

その他のマッチングの方法としては、遺伝的マッチング(Genetic matching)、フルマッチング(full matching)などがある。

## マッチングの利点と欠点(McKinlay, 1977; 岩崎, 2015)

### ■ 利点

- (1) わかりやすく, 統計の知識がなくても使える。
- (2) 独立なサンプルよりも推定効率が良い(注意: 1:1マッチングだからといって対応のある解析は行わない)。
- (3) データ解析が容易である。

### ■ 欠点

- (1) マッチングにコストがかかる。また, マッチングを行うと標本サイズが必ず少なくなる。
- (2) 標本サイズの減少に伴い, 標準誤差(標準偏差を標本サイズの平方根で割ったもの)が増加するため, 推定効率が下がる場合がある。
- (3) 他の方法(事後的な共分散分析や事後層別)に比べて必ずしも効率が良い方法ではない。

## 事例の結果

傾向スコア: ロジスティック回帰

マッチング: 1対1非復元マッチング(キャリアを伴う最近傍法: 0.25SD)

マッチング前		薬剤A (N=110)	薬剤B (N=71)	p値
性別	男性	69 (62.7%)	52 (73.2%)	0.1500
	女性	41 (37.3%)	19 (26.8%)	
年齢		53.4 (8.32)	54.1 (8.62)	0.6040
喫煙	有	30 (27.3%)	34 (47.9%)	0.0048**
	無	80 (72.7%)	37 (52.1%)	
BMI		24.1 (2.92)	25.0 (3.77)	0.0632
重症度スコア		7.3 (0.18)	8.8 (0.22)	<0.0001***

マッチング後		薬剤A (N=38)	薬剤B (N=38)	p値
性別	男性	24 (63.2%)	24 (63.2%)	1.0000
	女性	14 (36.8%)	14 (36.8%)	
年齢		55.3 (1.36)	55.3 (1.36)	0.2947
喫煙	有	14 (36.8%)	13 (34.2%)	0.811
	無	24 (63.2%)	25 (65.8%)	
BMI		24.2 (0.48)	24.3 (0.48)	0.951
重症度スコア		8.2 (0.29)	8.2 (0.29)	1.0000

## 傾向スコアの利用の有無によるアウトカムに対する結果の違い

マッチング前	効果なし	効果あり	p値
薬剤A(N=110)	21 (19.1%)	89 (80.9%)	0.0012**
薬剤B (N=71)	30 (42.25%)	41 (57.75%)	
OR = 3.10 [1.59, 6.06]			
マッチング後	効果なし	効果あり	p値
薬剤A(N=38)	9 (23.6%)	29 (76.3%)	0.0896
薬剤B (N=38)	17 (44.7%)	21 (55.3%)	
OR = 2.61 [0.98, 6.98]			

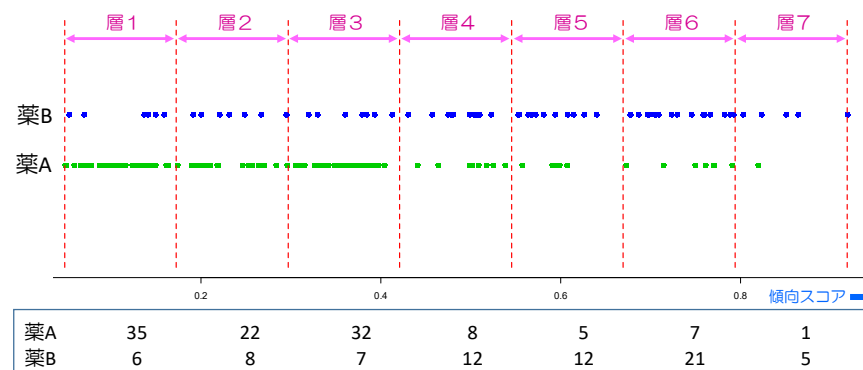
マッチングすることで, 有意差がなくなった。その理由としては,

- 症例数が半分未満になったことで検出力が減少した。
- 群間のインバランスが有意差を生じさせたため, それを解消したため。

オッズ比も減少していることから, 群間のインバランスの影響が少なからず存在したことが示唆される。

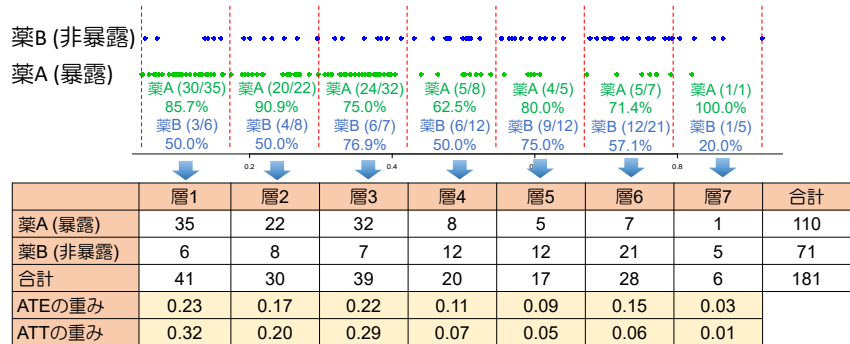
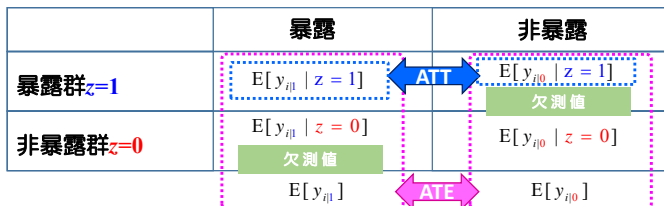
## 傾向スコア層別化

傾向スコアで層に分ける(Stratify)ことにより, 共変量の影響を除去する方法を傾向スコア層別化(propensity score stratification)という。



症例の割合で加重平均をとる方法が傾向スコア層別化である。

## ATE, ATTと傾向スコア層別化の関係



層kのATEでの重み =  $\frac{\text{層kの症例数}}{\text{全症例数}}$ , 層kのATTでの重み =  $\frac{\text{層kの暴露群での症例数}}{\text{暴露群の全症例数}}$

## 傾向スコア層別化における治療効果の推定

$$\text{推定式: } \Delta = \sum_{\text{層の個数}} (\text{層kの重み}) \times (\text{層kの治療効果})$$

	層1	層2	層3	層4	層5	層6	層7	合計
薬A (暴露)	35	22	32	8	5	7	1	110
薬B (非暴露)	6	8	7	12	12	21	5	71
合計	41	30	39	20	17	28	6	181
ATE	0.23	0.17	0.22	0.11	0.09	0.15	0.03	
ATT	0.32	0.20	0.29	0.07	0.05	0.06	0.01	

薬Aの有効率	85.7	90.9	75.0	62.5	80.0	71.4	100.0
薬Bの有効率	50.0	50.0	76.9	50.0	75.0	57.1	20.0
治療効果	35.7	40.9	-1.9	12.5	5.0	14.3	80.0

ATEでの計算	8.09	6.78	-0.41	1.38	0.47	2.21	2.65
ATTでの計算	11.36	8.18	-0.55	0.91	0.23	0.91	0.73

### 平均処理効果(ATE: Average Treatment Effect) ・ 平均因果効果(ACE: Average Causal Effect)

$$\Delta_{ATE} = 8.09 + 6.78 - 0.41 + 1.38 + 0.47 + 2.21 + 2.65 = 21.17(\%)$$

### 暴露群の平均処理効果(ATT: Average Treatment effect for Treated)

$$\Delta_{ATT} = 11.36 + 8.18 - 0.55 + 0.91 + 0.23 + 0.91 + 0.73 = 21.76(\%)$$

調整なし: 28.65%, マッチング: 21.00(%)

## 層別化解析における周辺平均重み付け

MMWS (Marginal Mean Weighting through Stratification; Hong, 2012)

先ほどの方法では、層に対して群に関係なく同じ重みを設定していた。これに対して、群毎に異なる重みを設定する方法がMMWSである。

	層1	層2	...	層S
暴露 ( $z=1$ )	$n_{11}$	$n_{21}$	...	$n_{s1}$
非暴露 ( $z=0$ )	$n_{10}$	$n_{20}$	...	$n_{s0}$
合計	$n_1$	$n_2$	...	$n_s$

	割合
暴露 ( $z=1$ )	$p_1$
非暴露 ( $z=0$ )	$p_0$

ATE: 層sの群zでの重み  $w_{sz} = \frac{n_s \times p_z}{n_{sz}}$     ATT: 層sの群zでの重み  $w_{sz} = z + (1-z) \frac{n_{s1} \times p_0}{n_{s0} \times p_1}$

	層1	層2	層3	層4	層5	層6	層7	合計	割合
薬A (暴露)	35	22	32	8	5	7	1	110	0.608
薬B (非暴露)	6	8	7	12	12	21	5	71	0.392
合計	41	30	39	20	17	28	6	181	
ATE	暴露群での重み	0.71	0.83	0.74	1.52	2.07	2.43	3.65	
	非暴露群での重み	2.68	1.47	2.19	0.65	0.56	0.52	0.47	
ATT	暴露群での重み	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
	非暴露群での重み	1.39	0.66	1.09	0.16	0.10	0.08	0.05	

## 層別化における推定量

### Horovitz-Thompson型推定量

$$\Delta = \frac{\sum_{s=1}^{\text{層の個数}} (\text{層sでの } z=1 \text{ の重み}) \times (\text{層1での } z=1 \text{ のアウトカム})}{z=1 \text{ の重みの総和}} - \frac{\sum_{s=1}^{\text{層の個数}} (\text{層sでの } z=0 \text{ の重み}) \times (\text{層1での } z=0 \text{ のアウトカム})}{z=0 \text{ の重みの総和}}$$

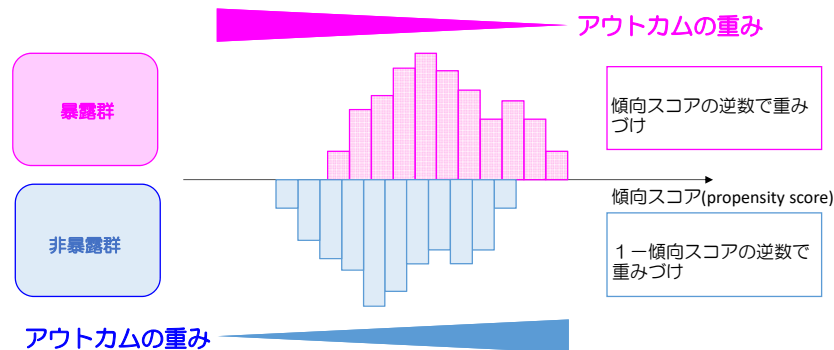
### 平均処理効果(ATE: Average Treatment Effect) : $\Delta_{ATE} = 25.62(\%)$

### 暴露群の平均処理効果(ATT: Average Treatment effect for Treated) : $\Delta_{ATT} = 22.00(\%)$

調整なし: 28.65%, マッチング: 21.00(%) , 通常ATE: 21.17(%) , 通常ATT: 21.76(%)

- MMWSは、各層に対する重みを小さくする傾向にある。
- MMWSは、IPWの代替として用いることができる(Linden, 2017)。

## 逆数重み付け(IPW: Inverse Probability Weighting)法



逆数重み付け法の場合とは、傾向スコアにおいて新薬・既存薬が選択された必然性が小さい個体の影響を強く、必然性が大きい個体の影響を弱くするように重みを付ける方法である(その極端な状況がマッチングと言える)。

## IPWの重み

### ■ 平均処理効果(ATE: Average Treatment Effect)

$$\text{被験者 } i \text{ の ATE での重み} = \frac{z_i}{e_i} + \frac{1-z_i}{1-e_i}$$

$z_i$  : 1のとき暴露群, 0のとき非暴露群  
 $e_i$  : 傾向スコア

暴露群では1/(傾向スコア)の重み

非暴露群では1/(1-傾向スコア)の重み

### ■ 暴露群の平均処理効果(ATT: Average Treatment effect for Treated)

$$\text{被験者 } i \text{ の ATT での重み} = z_i + (1-z_i) \frac{e_i}{1-e_i}$$

$z_i$  : 1のとき暴露群, 0のとき非暴露群  
 $e_i$  : 傾向スコア

暴露群の重みは1

非暴露群の重みはオッズ比

数値例

傾向スコア $e$	ATE		ATT	
	暴露群	非暴露群	暴露群	非暴露群
0.125	8.000	1.143	1.000	0.143
0.250	4.000	1.333	1.000	0.333
0.375	2.667	1.600	1.000	0.600
0.500	2.000	2.000	1.000	1.000
0.626	1.597	2.674	1.000	1.674
0.750	1.333	4.000	1.000	3.000
0.875	1.143	8.000	1.000	7.000

## IPWの難点と対応

IPWでは傾向スコアの逆数で重み付けを行う。そのため、傾向スコアが極端に小さい(あるいは大きい)場合には、重みが非常に大きくなる。たとえば、

$$(\text{非暴露群に属する被験者のPS}) = 0.999 \rightarrow (\text{ATEにおける重み}) = 1000$$

Leite (2017)は、この問題に対応する方策として、次を挙げている：

- (1) 傾向スコアを計算するモデルの誤特定に起因していないか点検する。
- (2) 傾向スコアの方法を変更する(例えば、層別化法など)
- (3) 重みの打ち切り(weight truncation)を行う

Lee et al.(2011)は、任意のパーセント点をそれ以上の重みをもつ個体に割り当てることを提案しており、Gurel & Leite(2012)はその基準として99%点の適用を推奨している。

- (4) 安定化重み(stabilized weights)を用いる

Harder et al.(2010)は、群毎に安定化を行うための定数を掛け合わせることで、上記の問題に対処している。

## 治療効果の推定

### Horovitz-Thompson型推定量

$$\Delta = \frac{\sum_{i=1}^{\text{暴露群の被験者数}} (\text{被験者 } i \text{ の重み}) \times (\text{被験者 } i \text{ のアウトカム})}{\text{暴露群の重みの総和}} - \frac{\sum_{j=1}^{\text{非暴露群の個数}} (\text{被験者 } j \text{ の重み}) \times (\text{被験者 } j \text{ のアウトカム})}{\text{非暴露群の重みの総和}}$$

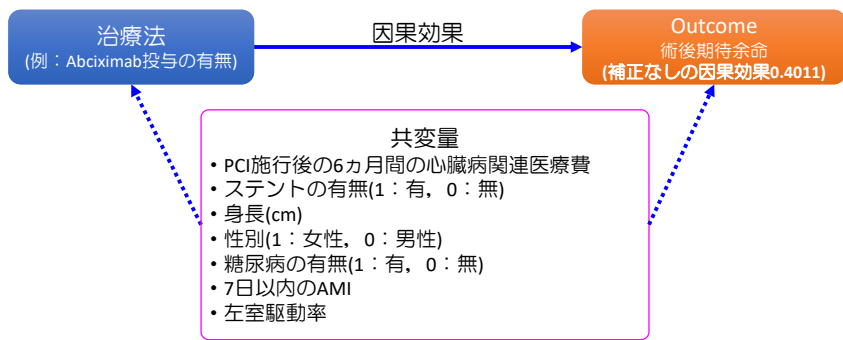
傾向スコア $e$	ATE		アウトカム		重み付きアウトカム		
	暴露群	非暴露群	暴露群	非暴露群	暴露群	非暴露群	
0.125	8.000	1.143	35	10	13.50	0.55	
0.250	4.000	1.333	40	15	7.71	0.96	
0.375	2.667	1.600	45	20	5.79	1.54	
0.500	2.000	2.000	50	25	4.82	2.41	
0.626	1.597	2.674	55	30	4.24	3.87	
0.750	1.333	4.000	60	35	3.86	6.75	
0.875	1.143	8.000	65	40	3.58	15.42	
					合計	43.5	31.5
						治療効果	12.0

傾向スコアを用いない場合が50.0-25.0=25.0であることから、IPWでの重み付け推定量の場合のほうが、減少している。

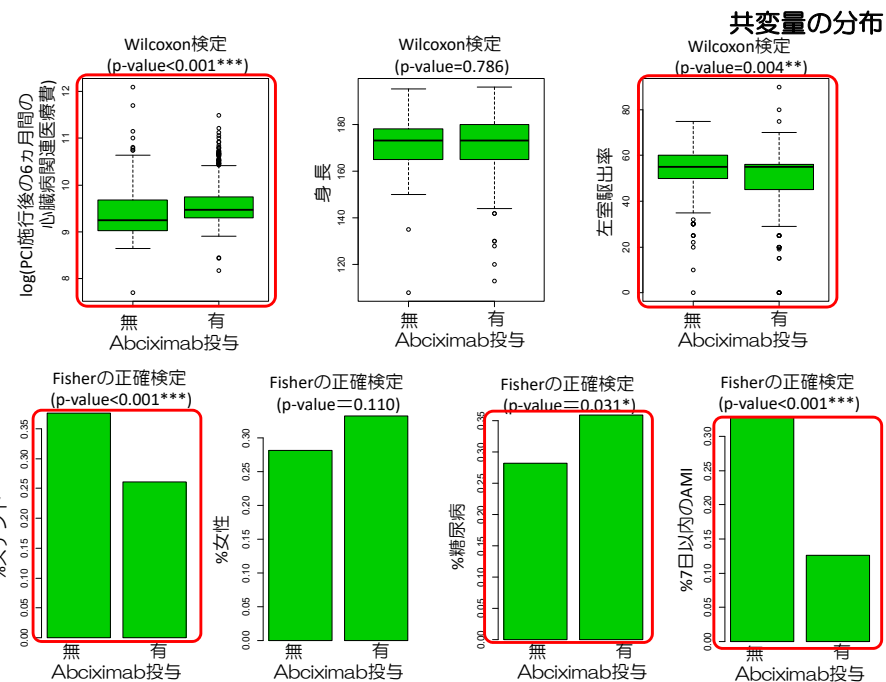


## 傾向スコア分析の実際

米国において、院経皮的冠動脈形成術を施行された996名の患者に対して、術後期待余命が調査された。

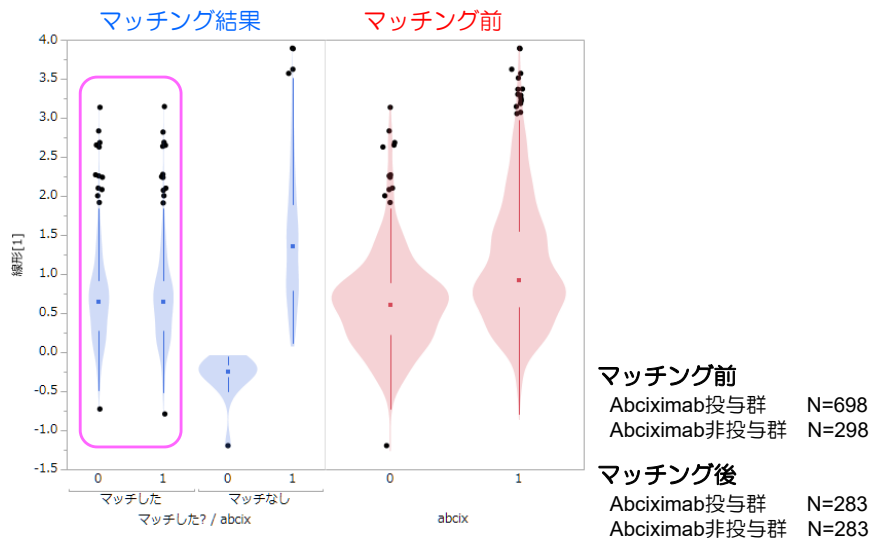


Kereiakes DJ, Obenchain RL, Barber BL, et al. Abciximab provides cost effective survival advantage in high volume interventional practice. *Am Heart J* 2000; 140: 603-610.



## マッチング後の分布

- PSの計算: ロジスティック回帰
- マッチングの方法: 1:1最近傍マッチング(キャリパー = 0.2×SD)



## 治療効果の推定

### マッチングなし

	症例数 [生存/死亡]	6カ月生存割合 [95% C.I.]	OR [95% C.I.]	Fisherの 正確検定
Abciximab投与群	698 [678/11]	98.4% [97.2%, 99.1%]	3.31 [1.50, 7.30]	0.004
Abciximab非投与群	298 [283/15]	95.0% [91.8%, 96.9%]		

### マッチングあり

	症例数 [生存/死亡]	6カ月生存割合 [95% C.I.]	OR [95% C.I.]	Fisherの 正確検定
Abciximab投与群	283 [279/4]	98.3% [96.4%, 99.4%]	3.36 [1.08, 10.43]	0.046
Abciximab非投与群	283 [270/13]	95.4% [91.8%, 97.3%]		

## JMPを用いた解析

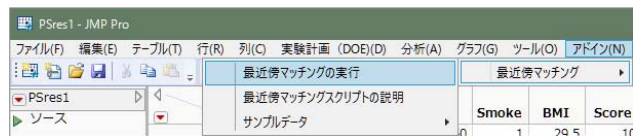
JMPでは、キャリパーを伴う1:1最近傍マッチングの方法をアドインとして公開している。

<http://www.jmp.com/content/dam/jmp/documents/jp/support/propensityaddin.zip>

### [STEP.1] Propensity scoreを計算 (ここではロジスティック回帰)

- (1) 「分析」→「モデルのあてはめ」を選択
- (2) 「役割変数の選択」に『グループ変数(名義尺度)』, 「モデル効果の構成」に共変量を選択する。
- (3) 「▼」から「確率の計算式の保存」を選択する(もとのデータセットに予測値が表示される)。

### [STEP.2] 事前に入手したアドイン(上記参照)をインストールした場合, 以下のメニューが追加される。



### [STEP.3] マッチングを行う。

- (1) 「アドイン」→「最近傍マッチングの実行」を選択
- (2) 「Y, 応答変数」に『アウトカム』, 「T, 処理変数」に『グループ変数』, 「マッチングに使うスコア」に『線形[1]』を選択。
- (3) 「ロジットを変換するか」のチェックを外す。
- (4) 「元データにペアの列を追加するか?」にチェックを入れる。
- (5) キャリパーを任意で設定(0~1であり, 推奨は0.2~0.25)

### [STEP.4] 治療効果を推定する。

- (1) 「分析」→「二変量の関係」を選択
- (2) 「Y, 目的変数」に『アウトカム』, 「X, 説明変数」に『グループ変数』, 「By」に『マッチングした?』を選択する。

出力は, 「マッチングした」のみを通常の解析方法と同様に解釈する。

**Thank you for your kind attention**

