

医学統計セミナー アドバンスコース  
**統計的因果推論と傾向スコア**

下川 敏雄

和歌山県立医科大学 臨床研究センター

2016年度 医学統計セミナー

■ ベーシック・コース

- ▶ 基礎統計学 (6月15日・住金棟5F 大研修室)
- ▶ 量的データの解析 (7月27日・住金棟5F 大研修室)
- ▶ 質的データの解析 (8月24日・住金棟5F 大研修室)
- ▶ 共変量調整を伴う解析 (11月2日・病院棟4F 臨床講堂1)
- ▶ 生存時間・臨床検査データの解析(11月16日・住金棟5F 大研修室)

■ アドバンス・コース

- ▶ 多群・経時データの解析と多重比較  
(11月30日・病院棟4F 臨床講堂1)
- ▶ 臨床試験における症例数設定とガイドライン  
(12月28日・住金棟5F 大研修室)
- ▶ アンケート調査データの解析 (2月1日・病院棟4F 臨床講堂1)
- ▶ 統計的因果推論と傾向スコア (2月22日・住金棟5F 大研修室)
- ▶ メタアナリシス (3月22日・病院棟4F 臨床講堂1)

関係の種類

薬剤 A を投与した群と薬剤 B を投与した群での有効率を評価した結果、薬剤 A の有効率のほうが薬剤 B の有効率よりも高かった。

薬剤 A 投与群と薬剤 B 投与群で患者背景が異なる可能性がある。薬剤投与と有効率の関係は**連関関係(association)**という。



薬剤 A の投与と薬剤 B の投与をランダムに振り分けた場合、患者背景が均一になる。この場合、薬剤投与以外に有効率の原因になり得ない。この場合、原因(薬剤の違い)と結果(有効率)が明確である。この場合には、薬剤投与と有効率の間を**因果関係(causality)**という。



因果推論を行う上での4種類の変数のタイプ

- 処理の違いを表す変数 Z (例：新治療/既存治療)
- 結果変数 Y (例：アウトカム)
- 観測される共変量 X (例：年齢, 疾患の重症度など)
- 観測されない共変量 U

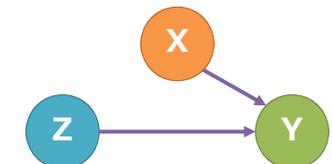
Scenario A

薬剤の投与 Z のみがアウトカム Y に影響を及ぼす場合



Scenario B

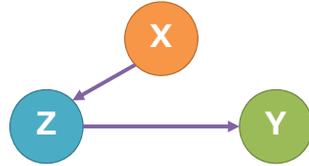
共変量 X がアウトカム Y に影響を及ぼす状況において、薬剤の投与 Z のアウトカム Y への影響を評価する場合



## 因果推論を行う上での4種類の変数のタイプ

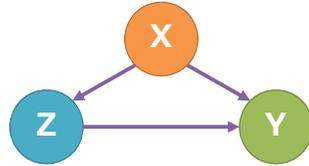
### Scenario C

共変量 X が薬剤の投与 Z に影響を及ぼす状況において、薬剤の投与のアウトカム Y への影響を評価する場合



### Scenario D

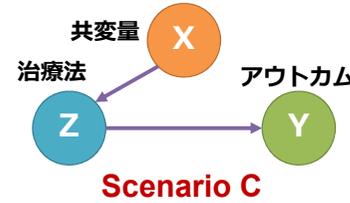
共変量 X が薬剤の投与 Z とアウトカム Y に影響を及ぼす状況において、薬剤の投与のアウトカム Y への影響を評価する場合



無作為化比較試験の場合、Scenario A or Bになる。Bの場合にはXの影響を統計学的に処理することでXの影響を排除できる。

観察研究(医師が自由に治療法を選択できる)場合、Scenario C or (D)になる。このような状況下でZのYへの因果関係を考えることが今回の目的である。

## 観察研究において起きうる状況：Scenario C,Dの再考



### Scenario Cの例

例えば、自由診療の治療法Aと保険診療の治療法Bを選択することを考えるとき、医療費 X (共変量)は治療法の選択に影響を及ぼすが、二つの治療法によるアウトカムに影響を及ぼさない(治療法とアウトカムに共変量の影響は存在しない)。

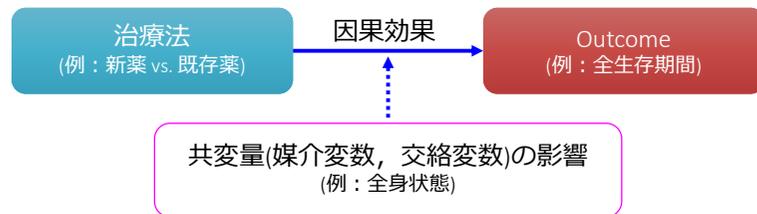
### Scenario Dの例

例えば、現在の治療Aと以前の治療法Bを選択することを考えるとき、年代 X (共変量)は治療法の選択とアウトカム(例えば、検査方法の違い等が含まれるため)に影響を及ぼす。

観察研究の多くは、Scenario Dの状況である。この状況下において、共変量Xの治療法Zへの影響を打ち切ることで、Scenario Bの状況下で評価する必要がある(Scenario Bであれば統計学的な調整が可能になる)。

## 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

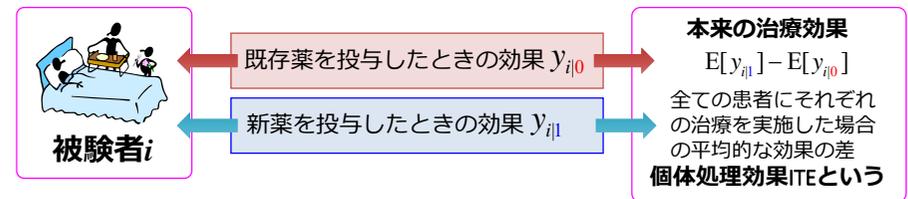
検証的な側面から共変量の影響を最小限にして因果効果を評価する

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

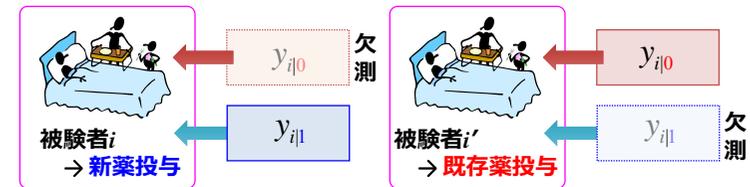
観察的な側面(観察研究)において(治療選択に対する)共変量の影響を最小限にして因果効果を評価する

- 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

## 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...

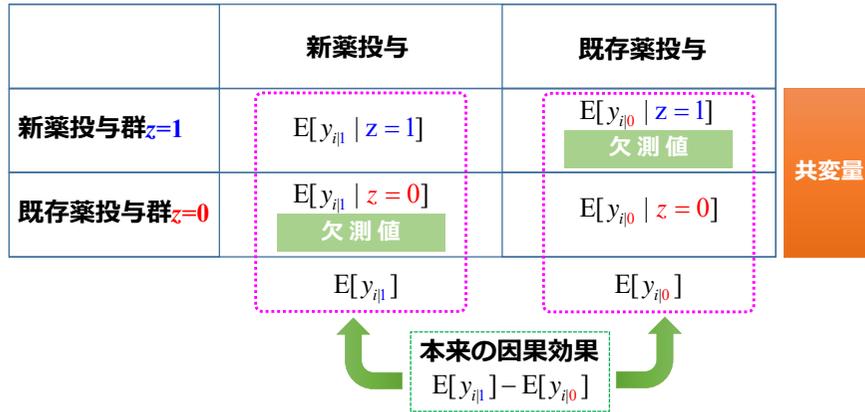


となり、いずれか一方は欠測であると考えられる。

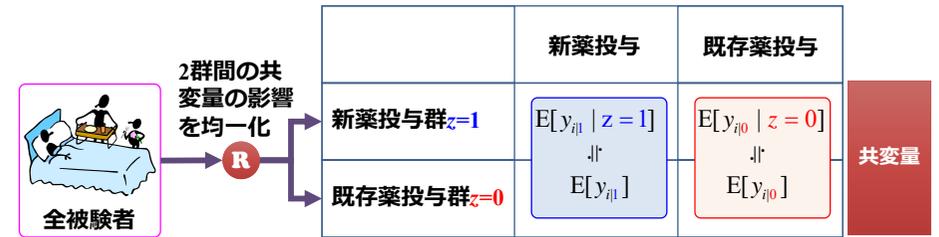
## Neyman-Rubinの反事実的モデル

### □ 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)



## 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



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。

## 傾向スコア(propensity score)による統計解析

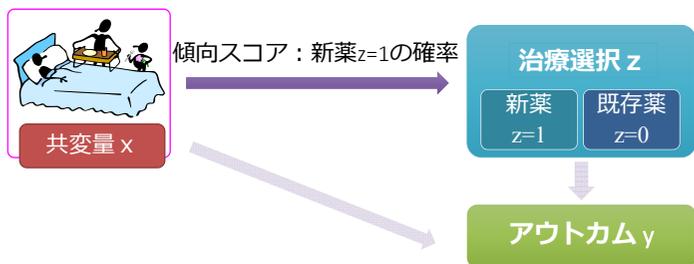
### □ 傾向スコアの推定

いずれかの群を1としたもとで、共変量(交絡因子)を用いてモデル(例：ロジスティック回帰)を構築し、傾向スコアを推定する。

### □ 傾向スコアを用いた調整

- 傾向スコアを類似性の測度として観測値のバランスをとる
  - マッチング, 層別解析の利用
- アウトカムの調整に用いる
  - 共分散分析の調整変数に利用する.
  - 傾向スコアを用いた重み付けによって因果効果の推定量を計算 (IPW(Inverse Probability Weighting)推定量, 2重口バスト推定量etc...)

## 傾向スコアとは？



共変量  $X$  が与えられたときに、新薬  $z=1$  が与えられる確率を **傾向スコア** という。

傾向スコアの計算方法は、治療選択  $z$  を応答変数とした回帰モデルを用いて計算される。傾向スコアを精度よく推定することが回帰モデルの目的であることから、統計手法・機械学習手法など、様々な方法が提案されている。そのため、傾向スコア計算に用いた方法を論文に記載する必要がある：

- ・ロジスティック回帰分析
- ・プロビットモデル
- ・ニューラルネットワーク
- ・一般化加法モデル
- ・Boostingモデル etc...

## 傾向スコアのための共変量 $X$ の選定

### 治療選択の根拠となる共変量を選ぶ

傾向スコア計算のための共変量  $X$  は、アウトカムへの影響を与える要因として考えるのではなく、**治療選択  $Z$  の根拠となるものを選択しなければならない**(治療選択に交絡がない状況(条件付独立性)が担保されなければならない)。

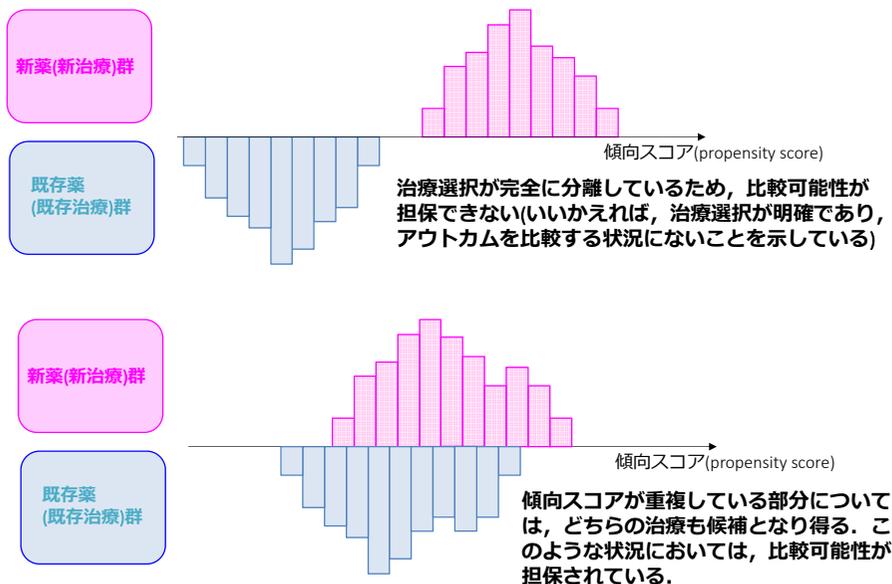
### 必要な変数はなるべく入れる

予後因子の評価で問題となる多重共線性は、共変量の解釈上で問題になるが、傾向スコア計算では、予測精度に主眼がおかれているため、あまり問題にならない。同様に、変数選択の方法によっては用いなくてもよい。

### 何でもかんでも入れてはいけない

治療に影響を及ぼす共変量をなるべく含めるべきであるが、治療の影響を受ける共変量は含めるべきでない(つまり、治療前の要因のみが選択されるべきである)。

## 傾向スコアによる解析の留意点



## 傾向スコアの利用

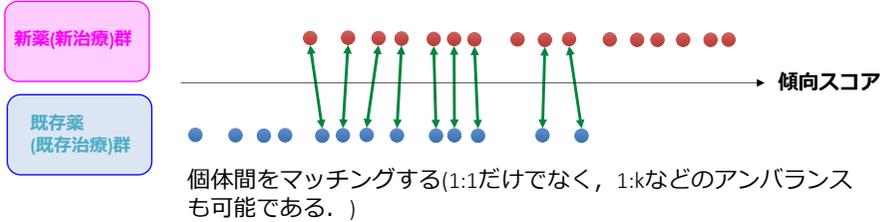
- **マッチング**  
傾向スコアの一致した(あるいは極めて近い)個体同士を選択する。
- **層別**  
傾向スコアを用いて層別(5層程度)したうえで層別解析を行う。
- **逆確率重み付け(IPWE: Inverse Probability Weighting Estimator)**  
傾向スコアを用いて層別(5層程度)したうえで層別解析を行う。
- **共分散分析**  
傾向スコアを共変量にとった共分散分析を実施する。

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。

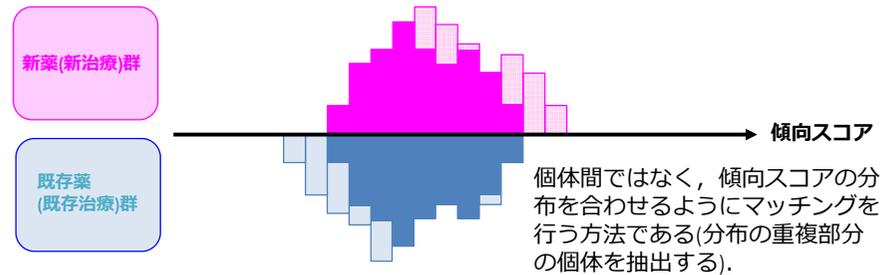
実際の応用分野では、マッチングあるいは逆確率重み付けを用いることが多いようである。

## マッチング

### ■ ペアマッチング



### ■ 度数マッチング



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

### ■ 利点

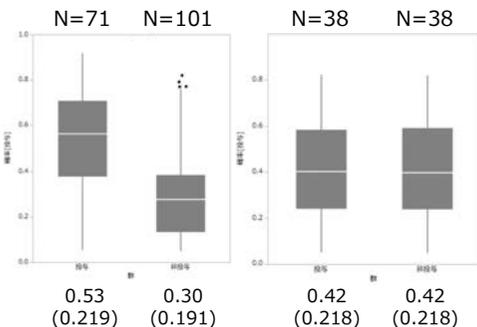
- (1) わかりやすく、統計の知識がなくても使える。
- (2) 独立なサンプルよりも推定効率が良い(注意：1:1マッチングだからといって対応のある解析は行わない)。
- (3) データ解析が容易である。

### ■ 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) 必ずしも効率が良い方法ではない。

## マッチング結果の評価

マッチングの効果は、傾向スコアの分布のバランスのチェックに統計的検定(有意でないからマッチングがうまくいっていると解釈する)を**用いてはならない**。(1) 両群での平均値および分散, (2) グラフによる分布の重複性のチェック, にとどめるべきである。



### JMPの仮想データ

共変量：年齢、性別、喫煙歴、BMI、重症度スコア

群：薬剤の投与、非投与

アウトカム：効果あり/効果なし

**傾向スコア：ロジスティック回帰**  
**マッチング：1対1マッチング**

## 先ほどのJMPの事例におけるマッチング後の共変量

マッチング前

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

マッチング後

	投与群 (N=38)	非投与群 (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	13 (34.2%)	14 (36.8%)	0.811
無	25 (65.8%)	24 (63.2%)	
BMI	24.3 (0.48)	24.2 (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	効果あり	効果なし	p値
投与群 (N=71)	30 (42.25%)	41 (52.25%)	0.0012**
非投与群 (N=110)	21 (19.1%)	89 (80.9%)	

OR = 3.10 [1.59, 6.06]

マッチング後	効果あり	効果なし	p値
投与群 (N=38)	17 (44.7%)	21 (55.3%)	0.0896
非投与群 (N=38)	9 (23.6%)	29 (76.3%)	

OR = 2.61 [0.98, 6.98]

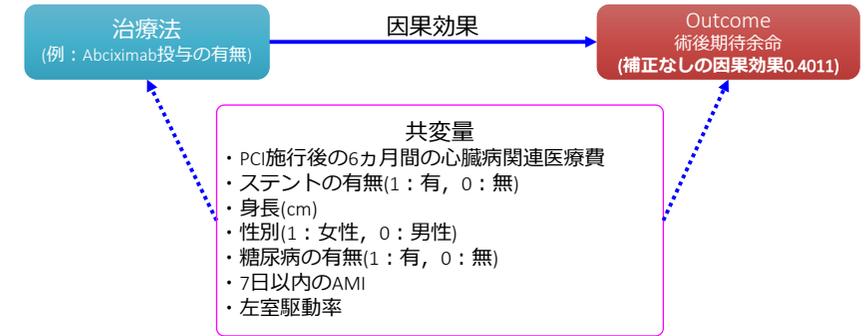
マッチングすることで、有意差がなくなった。その理由としては、

- ・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。

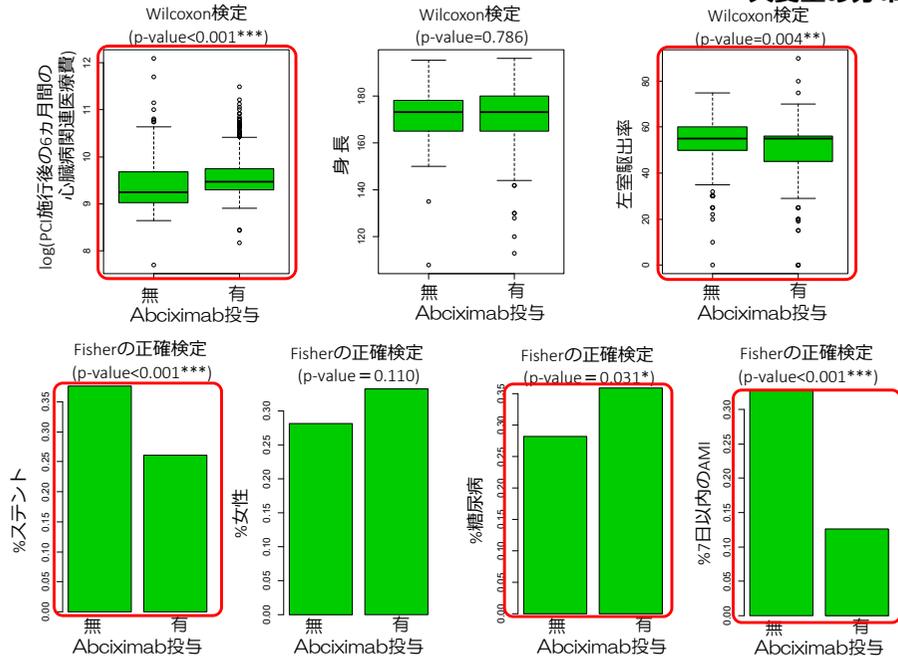
## 事例検討

米国において、996名の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。

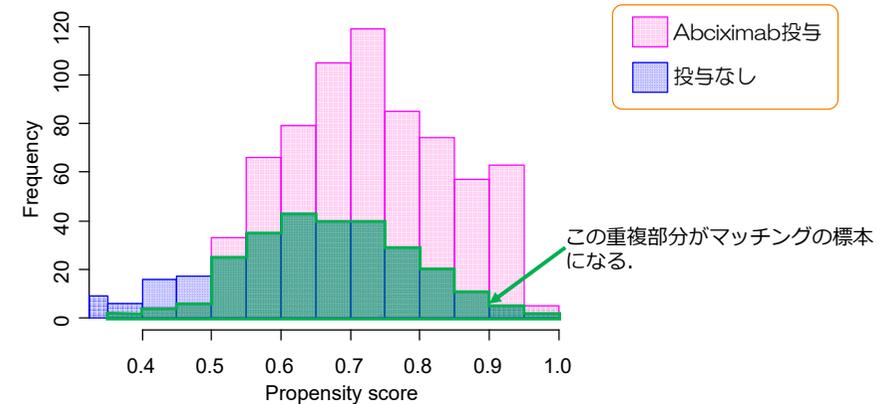


Kereiakes DJ, Obenchain RL, Barber BL, et al. Abciximab provides cost effective survival advantage in high volume interventional practice. *Am Heart J* 2000; 140: 603-610.

### 共変量の分布



### 群毎の傾向スコアの分布とマッチング



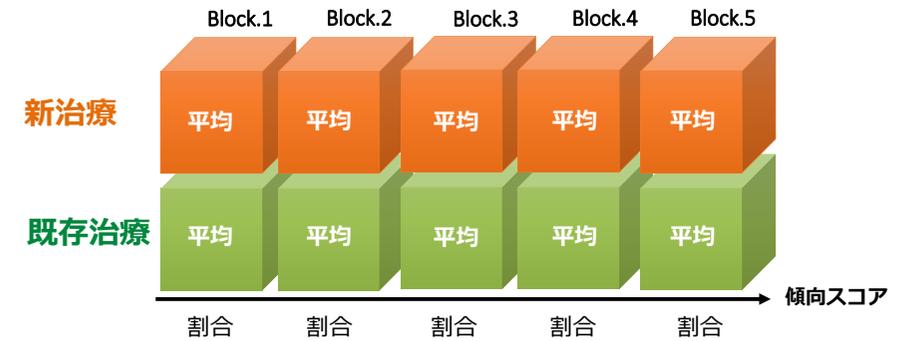
## マッチング後の結果と因果効果

マッチング前			マッチング後				
	投与群	非投与群	差		投与群	非投与群	差
log(医療費)	9.5816	9.3982	0.1834	log(医療費)	9.5816	9.6401	0.0585
ステント	0.7049	0.5839	0.1210	ステント	0.7049	0.6999	0.0050
身長	171.44	171.45	0.0100	身長	171.44	171.61	0.1700
女性	0.3310	0.3859	0.0549	女性	0.3310	0.3237	0.0073
糖尿病	0.2049	0.2685	0.0636	糖尿病	0.2049	0.1863	0.0186
7日以内のAMI	0.1791	0.0604	0.1187	7日以内のAMI	0.1791	0.1354	0.0437
左室駆出率	50.403	52.289	1.8860	左室駆出率	50.403	50.700	0.2970

### マッチング後の例数

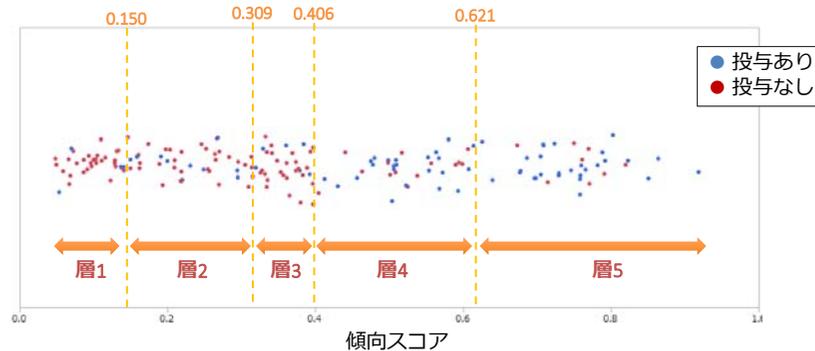
Abciximab投与 : 698                      投与なし : 298 → 190  
**マッチング後の因果効果 : 0.4011 → 0.8033**

## 層化解析法



新治療・既存治療の平均値  
 = (各ブロックでの平均値) × (各ブロックでの割合) の合計

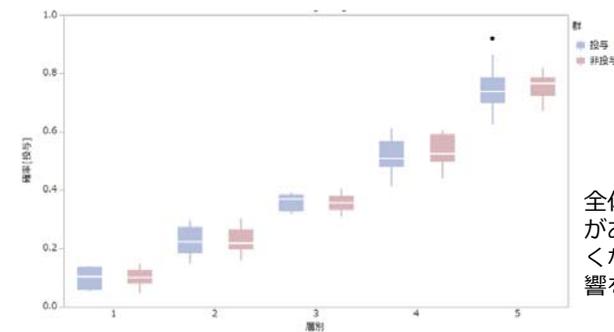
## 層化解析の例示



### アウトカムによる層化解析の違い

- 連続変数の場合 : 層を共変量とした分散分析
- 2値変数の場合 : 層を共変量としたMantel-Haenszel検定
- 生存時間の場合 : 層を共変量にした層別ログランク検定

## 層化解析の結果



全体で見ると傾向スコアに違いがあるが、層毎では違いが小さくなる。層別解析では、層の影響を考慮したうえで評価を行う。

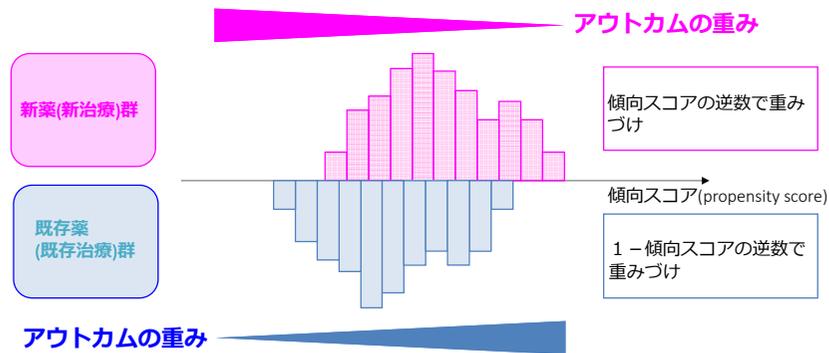
### ■ Mantel-Haenszel検定の結果

p値は0.0190になる。したがって、薬剤投与による影響が示唆される。また、このときの調整オッズ比は、

$$OR = 2.58 [1.17, 5.82]$$

である。

## 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)。

## JMPによる傾向スコアの解析

### ■ JMPを用いてマッチングを行うときのアドイン

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

このアドインでは、K近傍法という方法を用いた1:1マッチングの方法を提供している。ただし、K近傍法では、乱数によって結果(マッチングされる個体)が変化する可能性があるため、繰り返す場合には乱数シードを任意の数で一定にしたほうが良い。また、キャリパー(Caliper)とはマッチングの厳しさを表す数値であり、数字を大きくするほどマッチングの条件が甘くなる(マッチングされる個体数が増える)。

# ご清聴ありがとうございました

