

Propensity Score法とは

古川敏仁

株式会社バイオスタティスティカル リサーチ

応用数理研究会

2010/10/25



本日はまず、共変量を調整する際、
変数の数に気をつけること

変数が多くなったらどう対処すればよいか
を知っていただくのがテーマです



共変量の調整

- Y: 応答変数 (治療効果)
- T: 治療群割付 (1=実薬、0=対照)
- Z1 ~ Zn: 治療に影響を与える変数 例: 年齢、性別、重症度、.....

- $Y = b_0 + b_1 \cdot T$
- 一般的な予測モデル 線形
- b_1 が治療の効果 b_1 の推定が問題

共変量の調整 (Z1、Zn) が偏ったら

- $Y = b_0 + b_1 \cdot T + b_2 \cdot Z_1$
- 年齢Z1がYに影響するならモデルにZ1を入れる

- 7つの因子が有意に影響あり多少偏ったら
- $Y = b_0 + b_1 \cdot T + b_2 \cdot Z_1 + b_3 \cdot Z_2 + \dots + b_8 \cdot Z_7$

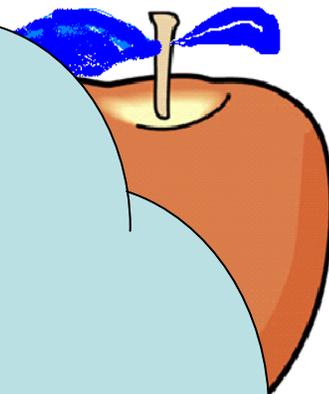
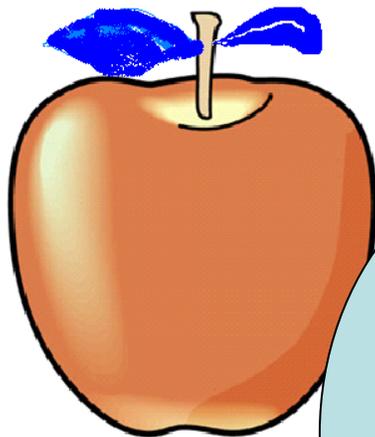
- これで、正しく、 b_1 は推定できるでしょうか

回答 no

- 絵をかいて、回答を説明してください
- 白板に絵を描く

回答

- 共変量調整による治療効果 b_1 の推定するときに、使用できる共変量は3~4まで。
- 推奨は1個、あるいは2個
- 多項目性の問題
- 共変量が増えると、モデルの線形性の仮定に影響を与える要因が増え、 b_1 は正しく推定できなくなる
- 多重共線性の問題も発生する。



なぜ、無作為化比較試験でしか、薬剤の効果を正しく推定できないのかを説明します

それは、交絡です

シン普森・パラドックスという言葉を知っていますか？

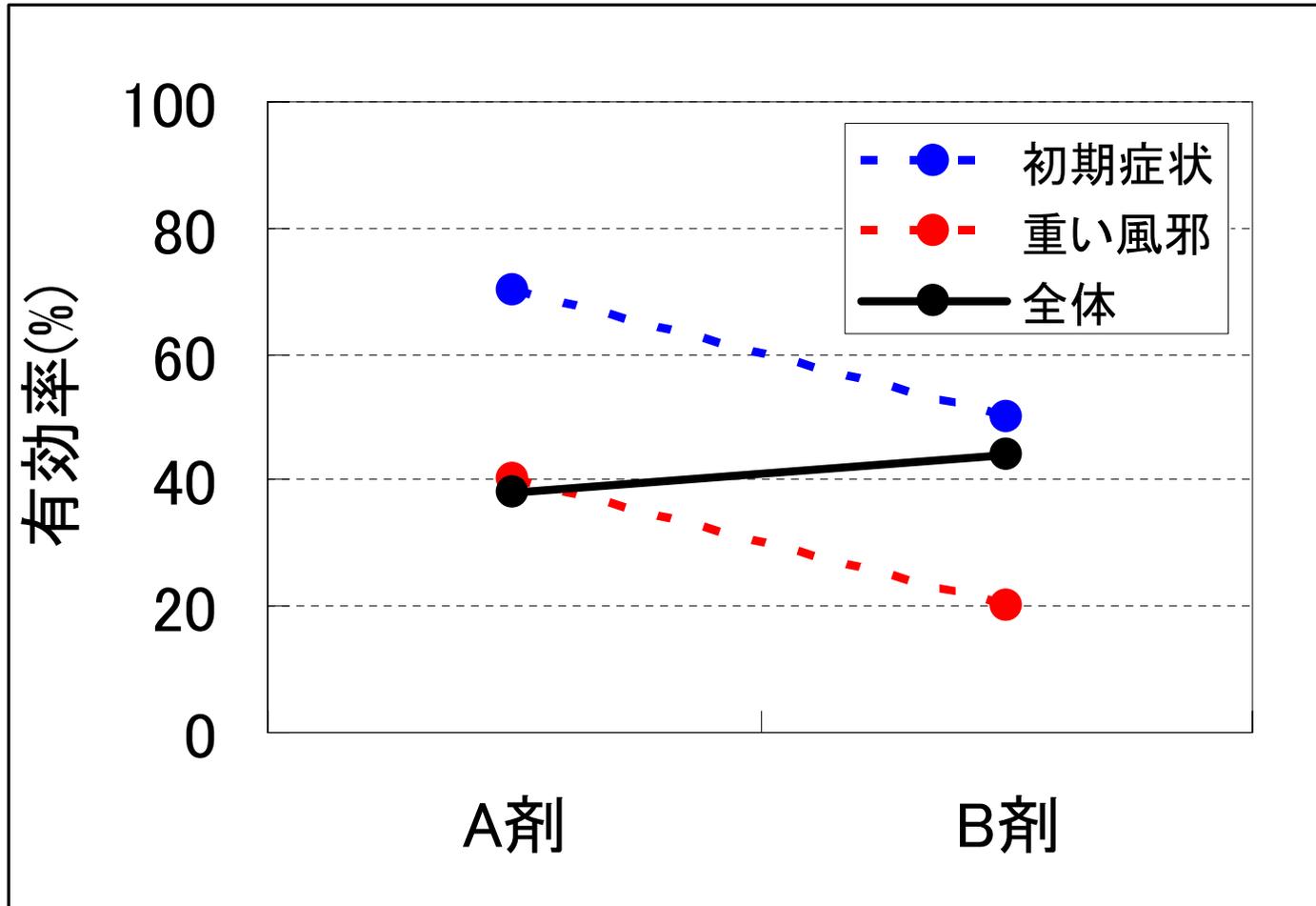
シンプソン・パラドックス (Simpson's Paradox)

- 交絡の1種
- 風邪薬は、風邪の初期症状には良く効くが、重たい風邪にはあまり効果がないことが知られている。
- ある製薬メーカーが従来のB剤にOTC成分を加えたA剤を開発した。
- 風邪薬投与2日後の風邪薬有効率を臨床試験で確認したところ、A剤は38% (38/100)、B剤は42% (42/100)で、新たに開発したA剤はB剤よりも有効率が低い結果となり、大問題となった・・・。

シンプソン・パラドックス (Simpson's Paradox)

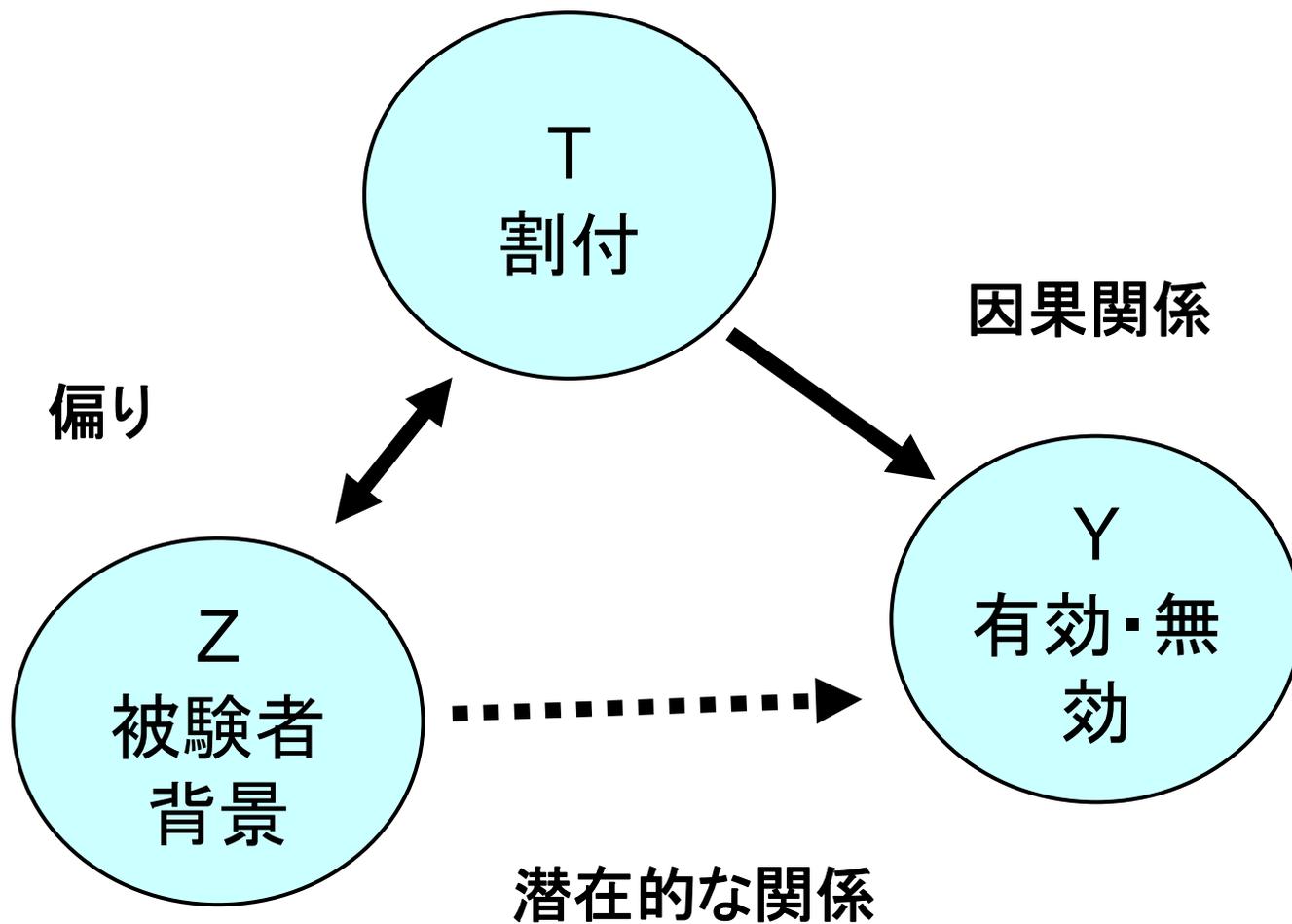
| | 風邪薬の有効率 (%) | | | |
|------|-------------|----------|-----|----------|
| | A剤 | | B剤 | |
| 初期症状 | 70% | (14/20) | 50% | (40/80) |
| 重い風邪 | 30% | (24/80) | 10% | (2/20) |
| 全体 | 38% | (38/100) | 42% | (42/100) |

シンプソン・パラドックス (Simpson's Paradox)



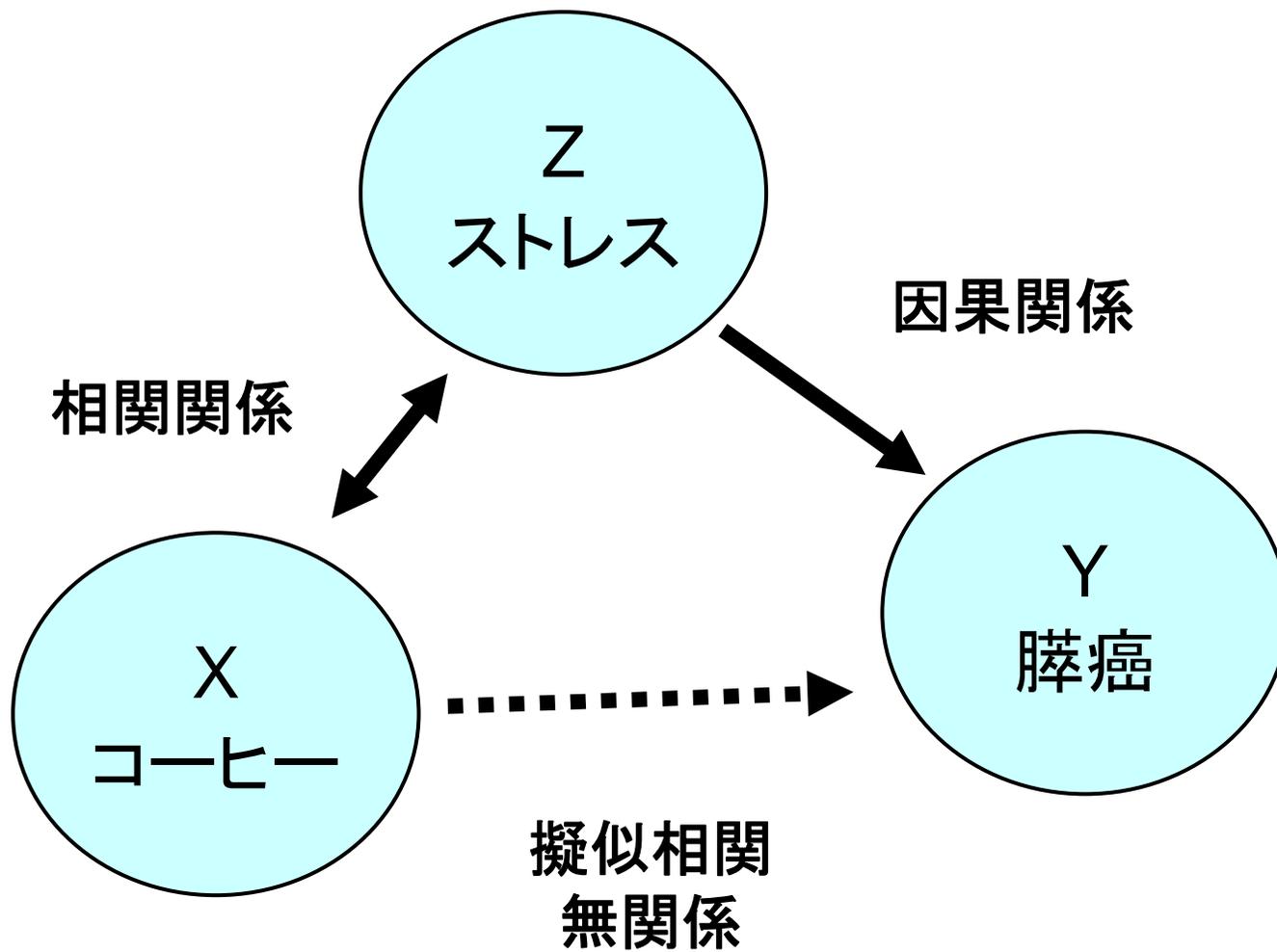
シンプソン・パラドックス (Simpson's Paradox)

- なぜだ その回答は
- 風邪の初期症状の症例でA剤とB剤の有効率を比較してみると、70%(14/20) vs 50%(40/80)、
- 重い風邪の症例では30%(24/80) vs 10%(2/20)と
- いずれも新たに開発したA剤の方が有効率が優れていたの
である。
- なぜ、全体で見るとB剤が優れているかのような結果となっ
たのであろうか。それは、A剤の評価では、初期症状-重い風
邪の比率が20:80であるのに対し、B剤では逆に80:20と
なっている。B剤の評価では直りやすい初期症状の症例が
多く、その結果、薬の効果とは関係なく風邪の症状が改善し
た症例が多かったのである(表4.1、図4.2)。まさに、パラドッ
クスである。

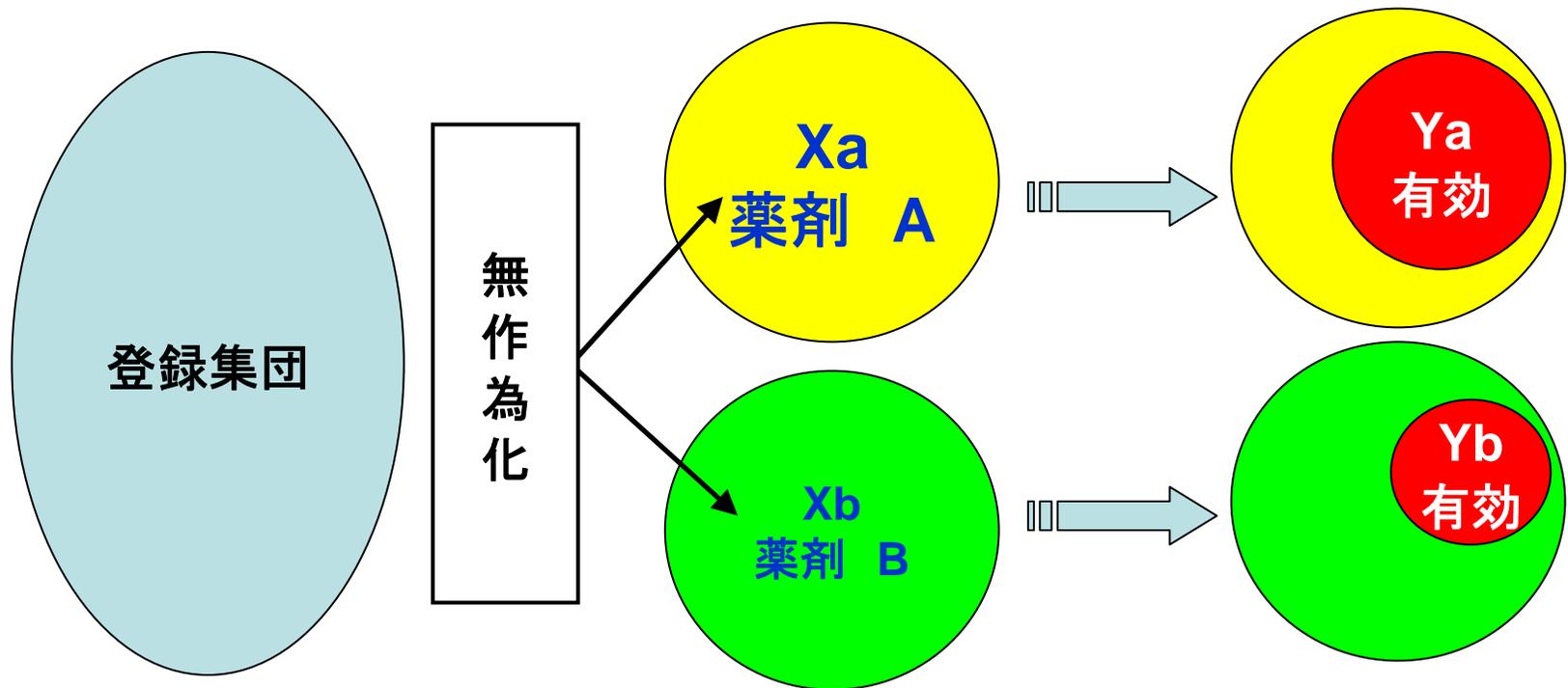


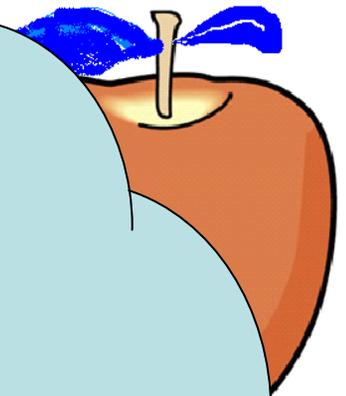
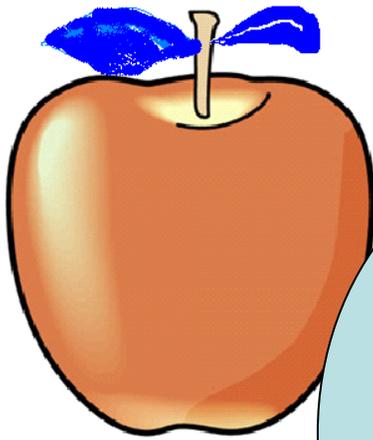
交絡の例 コーヒーと膵癌

- 1981年、医学会で最も権威のある雑誌といわれるThe New England Journal of Medicine (NEJM)に特集で膵癌とコーヒーには因果関係があり、コーヒーを多飲すると膵癌になりやすいというショッキングな論文が発表された¹⁾。
- 天下のNEJM の論文である。当時、コーヒー愛好家の私としては大問題であった。
- その後、「実はコーヒーと膵癌には因果関係はなく、コーヒーを良く飲む人はストレスが強く、ストレスと膵癌には因果関係があるため、あたかもコーヒーと膵癌に因果関係があるかのような現象が観察された」との報告もあり、コーヒーと膵癌の因果関係は不明であるというのが現在の説となっている(図4.1)。
- 1) B MacMahon, S Yen, D Trichopoulos, K Warren, and G Nardi 'Coffee and cancer of the pancreas', The New England Journal of Medicine, 304:630-633, 1981



無作為化比較臨床試驗 (RCT: randomized controlled trial)





ここから、因果推論の基本的な考え方を説明します。

下記を参考にしています。

(1)「統計的因果推論」 宮川雅巳 朝倉書店 2004

(2)「不完全データ解析の基礎と統計的因果推論」 狩野裕
2010:10-13 統計数理研究所 夏期講座

(3) Rubin, Donald. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies, Journal of Educational Psychology, Vol. 66, No.5, (1974), pp. 689

因果推論の基礎 (Causal Inference)

- Y: 反応変数 (response variable)

Y1: 治療群の反応

Y0: 対照群の反応

- T: 治療 (treatment)

T1: 治療群

T0: 対照群

注意: これは実際に行われた治療ではなく、割り付けられた状態

- われわれが、知りたいもの、治療群と対照群の治療効果の差 **平均因果効果**と呼ぶ
- $E[Y1 - Y0] = E[Y1] - E[Y0]$
- 実際に観察されるもの
 - 治療群 T1に割り付けられた治療効果 $E[Y1|T1]$
 - 対照群 T0に割り付けられた治療効果 $E[Y0|T0]$
- 観察研究 T1とT0とでは重症度のような背景が違ふ
よって、 $E[Y1] - E[Y0] \neq E[Y1|T1] - E[Y0|T0]$

- 無作為化比較試験の理論的背景
- 被験者への割付は無作為に起こる、T1、T0の割付は被験者背景とは関係なく、T1、T0とY1、Y0は独立

$$E[Y1|T1] = E[Y1|T0] = E[Y1]$$

$$E[Y0|T0] = E[Y0|T1] = E[Y0]$$

- よって、割付(被験者の個人特性)に関係なく $E[Y1] - E[Y0]$ が正しく推測される。

潜在反応モデルRubin Causal Model、potential outcomes

- Rubin, Donald. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies, Journal of Educational Psychology, Vol. 66, No.5, (1974), pp. 689
- ある個体 u
 - もし 治療T1を割り振れば $y_1(u)$
 - もし、対照T0を割り振れば $y_0(u)$
 - 同じ被験者で、 $y_1(u)$ 、 $y_0(u)$ を観察できれば平均因果効果は推定できる
 - $y_1(u) - y_0(u) \doteq E[Y_1] - E[Y_0]$

でも、同じ人に2つの治療は割り付けられない

潜在反応モデルRubin Causal Model、potential outcomes

- 同一個体 u に 2つの治療 T_0 、 T_1 が割り付けられ、 $y_1(u)$ 、 $y_0(u)$ を観察できれば平均因果効果は推定できる
- 実際には、 T_0 、 T_1 どちらかの治療しか行われないうえ、 $y_1(u)$ 、 $y_0(u)$ のどちらかは欠測値となる。
- このような欠測値が存在しても平均因果効果がPropensity Scoreを使用すれば推定できることを以下に示す。

- 個体 u を特徴付けるものの被験者背景 Z
 Z (Z_1 : 重症度、 Z_2 : 年齢、 Z_3 : 性別、 \dots Z_n : xxx)
- n 個の被験者背景が同じなら、同一人物と同等と見なす
- $y_1(T=1|\mathbf{Z}=z) - y_0(T=0|\mathbf{Z}=z) \doteq E[Y_1] - E[Y_0]$
- 被験者の情報は観察された n 個の情報ですべて説明されるという仮定をおく = strongly ignorable
- Z 以外の情報は無視できるとする

- strongly ignorable
- Z を与えた下で Y と X は条件付独立
- 以後の理論展開のkey

$$\begin{pmatrix} Y0 \\ Y1 \end{pmatrix} \perp T \mid Z = z$$

strongly ignorableのもとでの 平均因果効果 $E[Y_1] - E[Y_0]$ の推定

- 傾向スコア $e(z) = E[T=1|z]$
- 被験者背景 z の被験者が $T=1$ 治療群に割り付けられる確率

傾向スコアを使えば平均因果効果が推定できる

$$E\left[\frac{Y1T}{e(z)} \mid z\right] = E[Y1 \mid z] \times E\left[\frac{T}{e(z)} \mid z\right] = E[Y1 \mid z]$$

∵ Y1とTは、z givenのもとで条件付独立

$$\therefore E[T \mid z] = e(z)$$

$$E[Y1 \mid z] = E\left[\frac{Y1T}{e(z)} \mid z\right] \quad (1)$$

任意の確率変数X、Yにおいて $E[Y] = E_x[E[Y \mid X]]$ だから

(1)式をzに関して期待値を求めると

$$E[Y1] = E_z\{E[Y1 \mid z]\} = E_z\left\{E\left[\frac{Y1T}{e(z)} \mid z\right]\right\} = E\left[\frac{Y1T}{e(z)}\right] \text{となる}$$

同様に

$$E[Y0] = E\left[\frac{Y0(1-T)}{1-e(z)}\right]$$

m人の観察者による治療群 T=1の 平均因果効果はE[Y1]は

$$E[Y1] = E\left[\frac{Y1T}{e(\mathbf{z})}\right]$$

$$\frac{1}{m} \sum_{i=1}^m Y1i \approx \frac{1}{m} \sum_{i=1}^m \frac{Y1i t_i}{e(\mathbf{z}_i)}$$

$$= \frac{1}{m} \sum_{i=(t_i=1)} \frac{Y1i}{e(\mathbf{z}_i)} \approx \frac{1}{\sum_{i=(t_i=1)} \frac{1}{e(\mathbf{z}_i)}} \sum_{i=(t_i=1)} \frac{Y1i}{e(\mathbf{z}_i)}$$

$$\therefore E\left[\sum_{i=(t_i=1)} \frac{1}{e(\mathbf{z}_i)}\right] = E\left[\sum_{i=(t_i=1)} \frac{T_i}{e(\mathbf{z}_i)}\right] = mE\left[\frac{T}{e(\mathbf{z})}\right] = m$$

- つまり、T=1の平均因果効果は、実際に治療群の効果の個々人の傾向スコアe(z)の重み付け平均と定義できる。



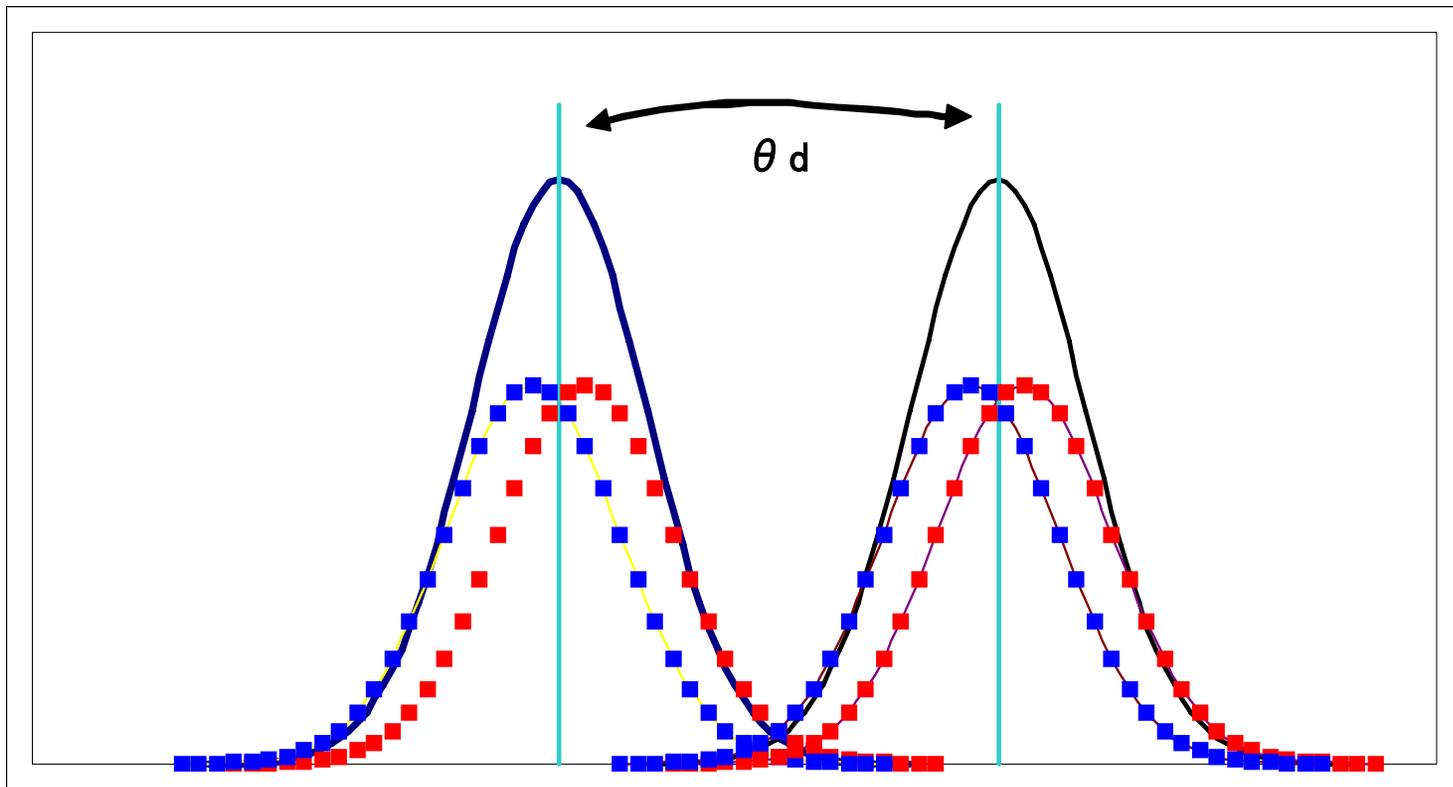
平均因果効果 $E[Y_1]$ は傾向スコアの
重み付け平均であることは説明しま
した

これをイメージで説明すると



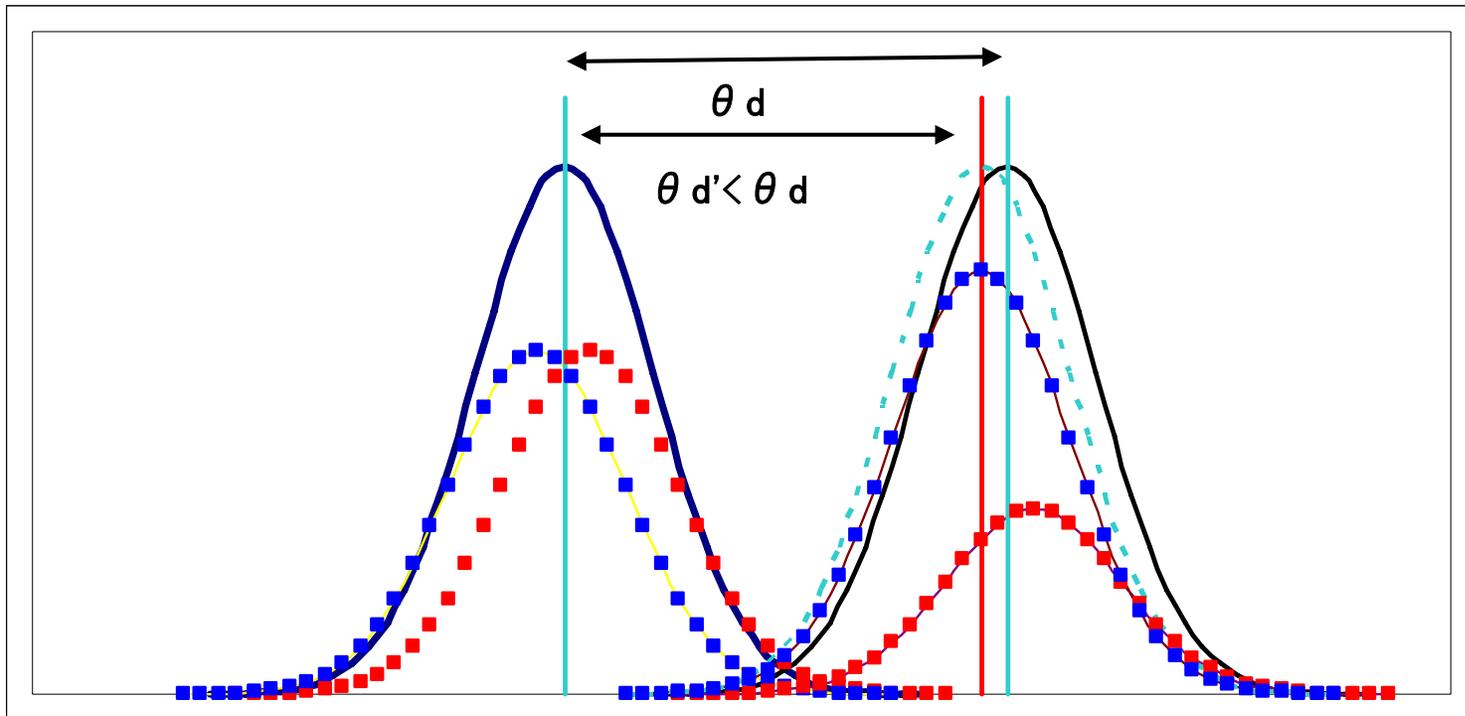
もし、背景因子(例:男=青、女=赤)の分布のバランスが治療群A、B間で保たれていれば

- 治療群効果 = $\theta a - \theta b = \theta d$
は正しく推定される。



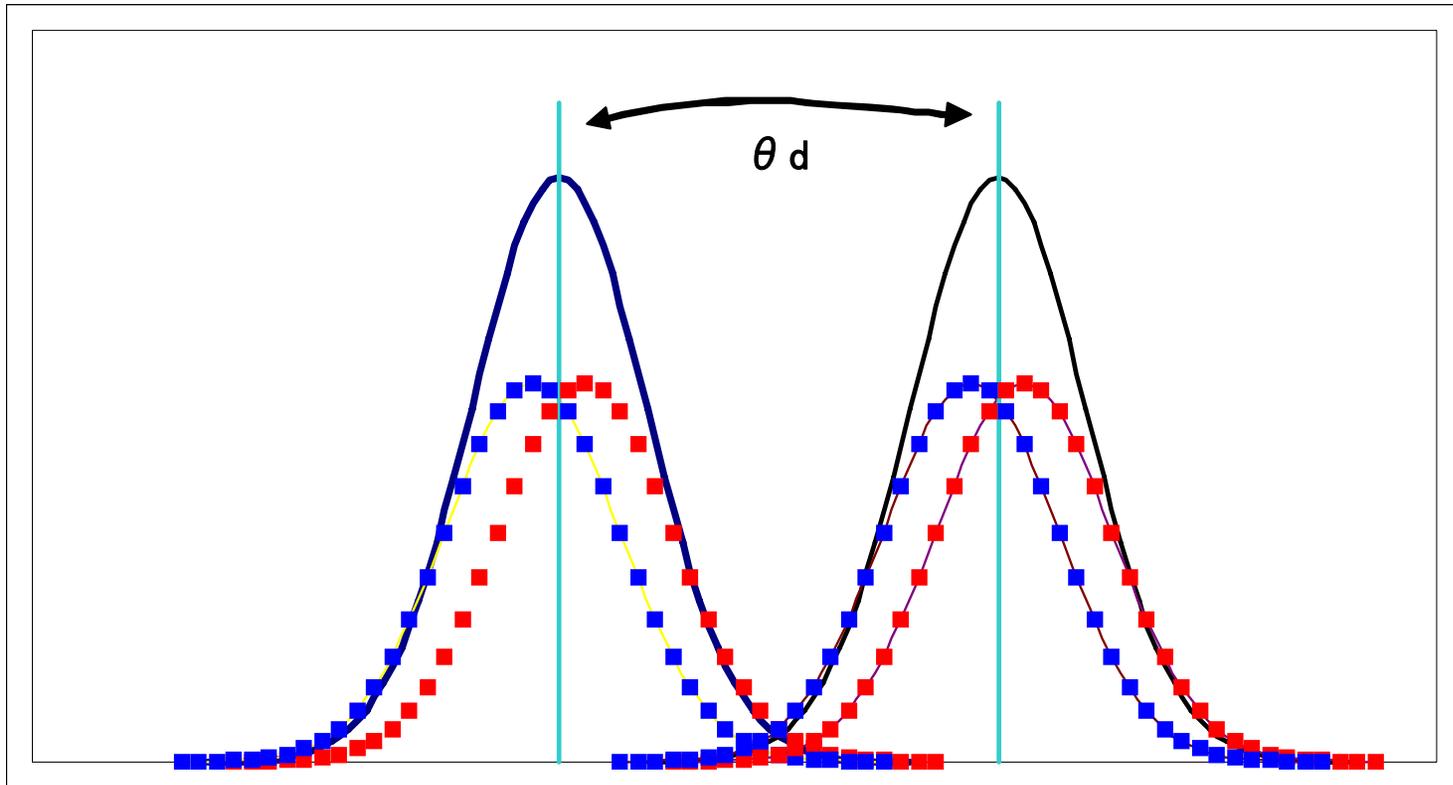
もし、治療群Bの男＝青の割合が治療群Aの割合よりも相対的に高くなると

- 治療群効果 = $\theta a' - \theta b' = \theta d' < \theta d$
- バイアスが発生する。



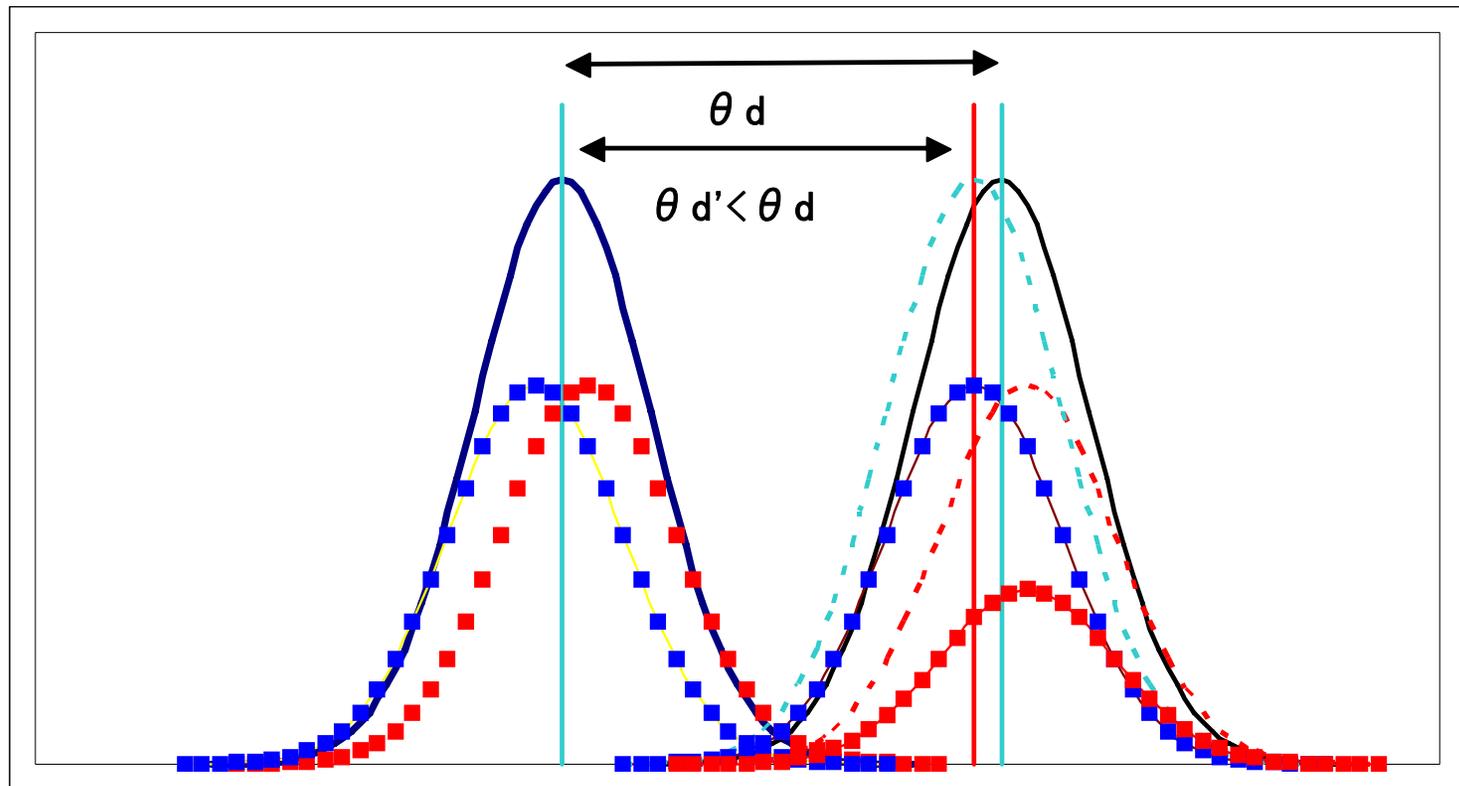
Propensity Score法

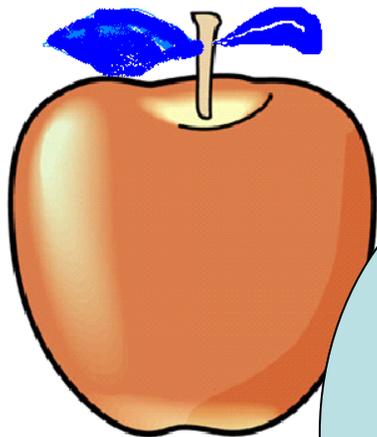
- Propensity Scoreと呼ばれるバランス変数を用いて治療群A、B群間の背景因子のバランスをとり
- θ_d を正しく推定する方法



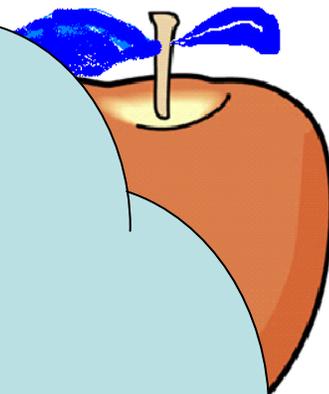
IPCW: Inverse Probability of Censoring Weighted

- 例えば、informativeな打ち切りにより、治療法Bの赤の分布に集中的にCensorが起こる場合
- Censor発生確率の逆数を重みとして、治療群A、B群間の背景因子のバランスをとり θ_d を正しく推定する方法





難しい話はここまで
ここから、傾向スコアの具体的な話を
進めます



共変量の調整：変数が多すぎれば変数を減らせばよい

- 下記は共変量が多すぎてだめ
- $Y=T+Z1+Z2+\dots+Z7$
- ならば、 $Z1\sim Z7$ を1つの変数 Ps にまとめてしまえばよい。
- $Y=T+Ps$
- これで、 T の効果は多項目性の問題を回避し適切に推定できる。

Psはどうやって作るのか

- T: 治療 T=1: 治療群、T=0: 対照群

- Proc logistic;

```
model T=z1 z2 z3...z7;
```

```
output out=データセット名 p=ps;
```

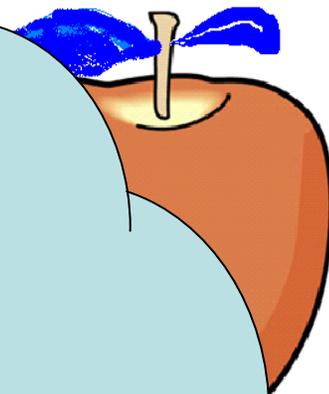
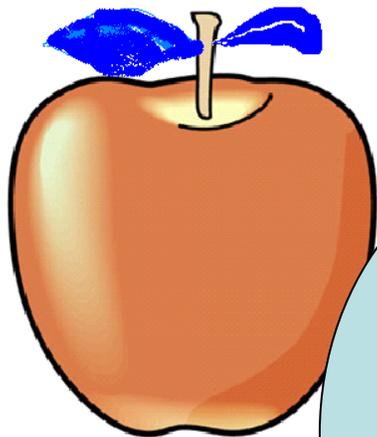
```
run;
```

被験者背景 z_1, z_2, \dots, z_7 によって予測される治療群の割付確率 p が ps という変数名に格納される

Ps: 傾向スコア

Psによる調整

- $Y = T + Ps$ というモデルでTを調整すればよい
- これは、 $y = z_1 + z_2 + z_3 \cdots + z_7$ という、T以外の共変量でyの予測値 $\text{Pred}(y)$ を求め
- $Y = T + \text{Pred}(y)$ というモデルでTを調整することと理論的には同じ
- モデルの線形性が成立すればどちらでも良い、でも線形性が成り立たないから傾向スコアを使う



具体的な応用例

あるデータの物語。

観察研究データ

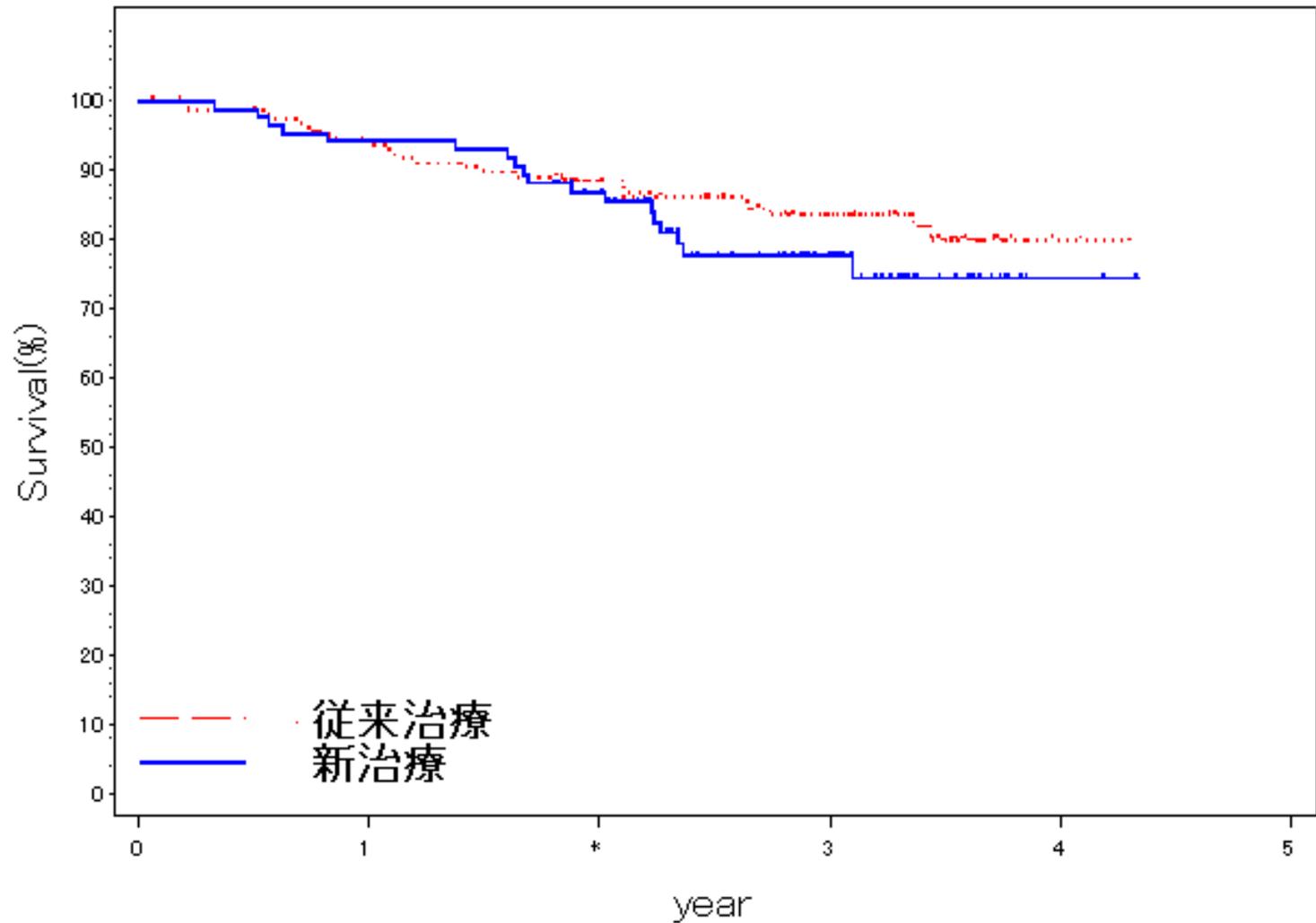
画期的な新治療が開発されましたが、新治療群の方が死亡率が高かった

新治療に予後の悪い患者が割り付けられた可能性が高い

割付のアンバランス

| | 例数 | 死亡 | 死亡率 | 生存 |
|------|-----|----|-------|-----|
| 従来治療 | 160 | 26 | 16.3% | 134 |
| 新治療 | 85 | 18 | 21.2% | 67 |
| 合計 | 245 | 44 | 18.0% | 201 |

Kaplan-Meier 新治療負けている



無調整な解析

ハザード比 新治療の死亡率が1.366
と高い

| | ハザード比 | 95%信頼区間 | | p値 |
|--------------|-------|---------|-------|--------|
| | 推定値 | 下限 | 上限 | |
| 従来治療=0、新治療=1 | 1.366 | 0.748 | 2.496 | 0.3096 |

とりあえず、多変量Cox解析 逐次変数増減法

- ハザードは1.0以下に
- でも選択される変数が7以上も・・・
- ハザードは正しく推定されているのか疑問

| | ハザード比 | | 95%信頼区間 | | p値 |
|--------------|-------|-------|---------|--------|----|
| | 推定値 | 下限 | 上限 | | |
| 従来治療=0、新治療=1 | 0.879 | 0.473 | 1.634 | 0.6841 | |
| 検査xxxx | 3.177 | 1.582 | 6.377 | 0.0012 | |
| 検査xxxx | 2.329 | 1.193 | 4.547 | 0.0132 | |
| 検査xxxx | 0.894 | 0.815 | 0.980 | 0.0169 | |
| 検査xxxx | 0.976 | 0.958 | 0.994 | 0.0105 | |
| 検査xxxx | 2.915 | 1.450 | 5.857 | 0.0027 | |
| 検査xxxx | 0.409 | 0.204 | 0.822 | 0.0120 | |
| 検査xxxx | 0.435 | 0.179 | 1.061 | 0.0672 | |
| 検査xxxx | 0.370 | 0.122 | 1.127 | 0.0802 | |

| 項目 | ハザード比 95%信頼区間 | | | p値 |
|--------|---------------|-----------|---------|----------|
| | 推定値 | 下限 | 上限 | |
| 年齢 | 1.015 | (0.991 - | 1.039) | 0.234 |
| 性別 | 1.225 | (0.569 - | 2.637) | 0.603 |
| 検査xxxx | 1.423 | (0.787 - | 2.572) | 0.243 |
| 検査xxxx | 0.940 | (0.899 - | 0.982) | 0.006 * |
| 検査xxxx | 1.334 | (0.687 - | 2.591) | 0.394 |
| 検査xxxx | 0.577 | (0.297 - | 1.120) | 0.104 |
| 検査xxxx | 0.873 | (0.312 - | 2.440) | 0.795 |
| 検査xxxx | 1.817 | (1.001 - | 3.300) | 0.050 * |
| 検査xxxx | 1.583 | (0.876 - | 2.861) | 0.128 |
| 検査xxxx | 0.442 | (0.061 - | 3.207) | 0.419 |
| 検査xxxx | 0.750 | (0.402 - | 1.399) | 0.365 |
| 検査xxxx | 1.217 | (0.645 - | 2.296) | 0.544 |
| 検査xxxx | 1.487 | (0.628 - | 3.523) | 0.367 |
| 検査xxxx | 0.926 | (0.848 - | 1.010) | 0.082 |
| 検査xxxx | 0.973 | (0.956 - | 0.991) | 0.003 * |
| 検査xxxx | 3.772 | (2.067 - | 6.884) | <0.001 * |
| 検査xxxx | 1.004 | (0.997 - | 1.012) | 0.251 |
| 検査xxxx | 1.408 | (0.771 - | 2.570) | 0.265 |
| 検査xxxx | 0.504 | (0.199 - | 1.279) | 0.149 |
| 検査xxxx | 1.002 | (0.996 - | 1.009) | 0.418 |
| 検査xxxx | 1.000 | (0.999 - | 1.002) | 0.596 |
| 検査xxxx | 2.747 | (1.372 - | 5.501) | 0.004 * |
| 検査xxxx | 1.503 | (0.806 - | 2.804) | 0.200 |
| 検査xxxx | 0.595 | (0.328 - | 1.081) | 0.088 |

被験者背景を確認したら
47項目もあり、
ほとんどの項目が死亡と
関連している
傾向スコア解析しかない

さあ、がんばって傾向スコアを作ろう

```
Proc logistic;
```

```
  model 治療法(従来治療=0、新治療=1)=  
    年齢 性別 検査1 ...検査45;
```

```
output out=データセット名 p=ps;
```

```
run;
```

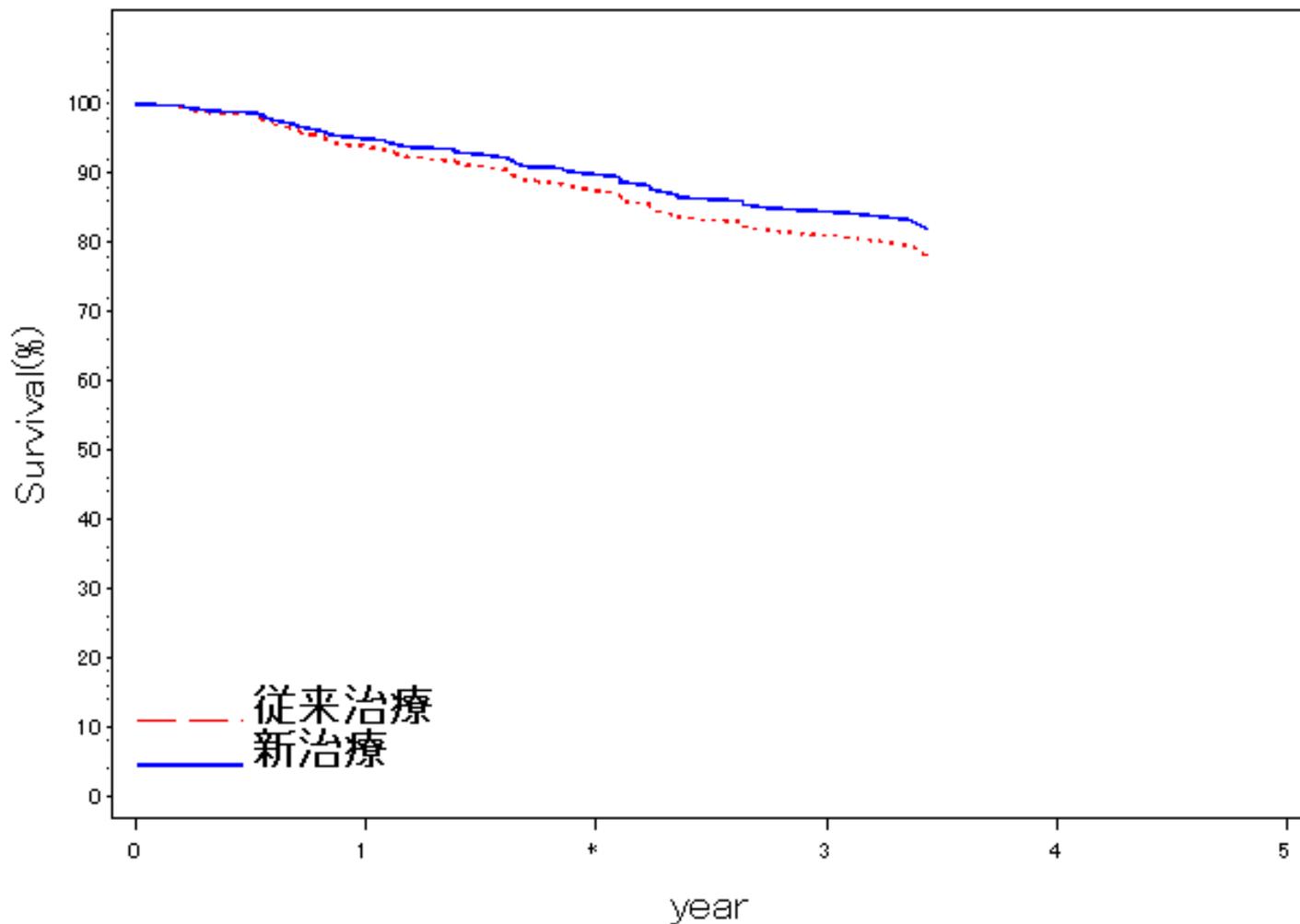
注意：現実には、変数選択や交互作用の確認等
を行い、上記ほど簡単ではありませんが

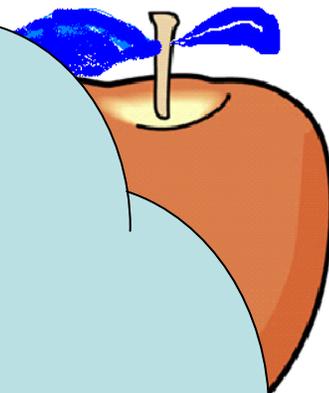
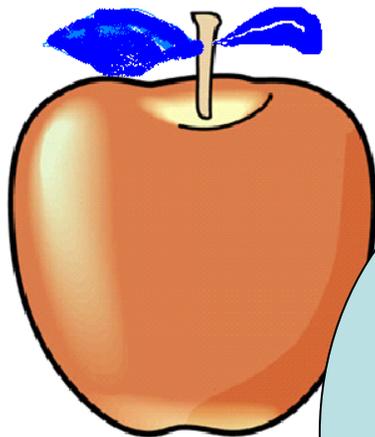
傾向スコアで調整すると

- やった、ハザードは0.80に
- 多変量の結果も同じ方向
- これで、良い雑誌に投稿可能？

| | ハザード比 | 95%信頼区間 | | p値 |
|----------------------|-------|---------|-------|-------|
| | 推定値 | 下限 | 上限 | |
| 従来治療=0、新治療=1 | 0.801 | 0.395 | 1.625 | 0.539 |
| 傾向スコア区分変数 (47項目から作成) | 1.189 | 1.046 | 1.352 | 0.008 |

調整済み生存曲線も 新治療が上回った





Propensity Score 作成上の注意

(1) 傾向スコア作成時の背景項目の選択
logisticモデル or 線形判別モデル

(2) 背景項目の選択 注意事項

(3) 選択したPropensity Scoreモデルの評価
Propensity Scoreによるバイアスの減少の確認

(4) Propensity Score 5区分変数を用いる理由

(5) 機能しないPS区分変数

(1)傾向スコア作成時の背景項目の選択 logisticモデル or 線形判別モデル

- ① 背景間で相関の高い、多重共線性のある項目は、代表的なもの1つを採用する。
- ② 少数例はずれ値の影響の高い背景項目はモデルから除外する。
- ③ それでもp値が0.70や0.80を超える項目がモデルにあったら、モデルから除外する。

最も重要なこと、

- ① 変数の選択によって最終的な結果が変わらない事を、感度分析で確認
＝何度も傾向スコア解析を行うこと。
- ① 複数の条件で結果が変わるようであれば、原因を追究し、最も妥当な条件を選択する。

Propensity Score法利用の実務的な問題

(2)背景項目の選択 注意事項

比較する治療法(曝露因子)とメカニズム的に関連する背景、予後因子は
Propensity Score構成項目に決して含めない。

例:手術関連治療の比較

ある特定の背景因子が存在するときにA(あるいはB)の治療を選択

理由:このような項目が Propensity Score構成項目に含まれると

治療群A、B間の治療効果差はほぼ0(ハザード比出れば1.0)となる。

実践では、多くのPropensity Score構成項目の中にこのような項目が紛れ込む場合が散見される

重要事項

- ・統計家と医学専門家のコミュニケーション
- ・数多くの感度分析

Propensity Score法利用の実務的な問題

(3)背景項目の選択 注意事項

Propensity Scoreを構成する項目間の交互作用

- Propensity Score構成項目のパラメータ空間の線形性からの乖離→バイアスの原因
- 現実的には、Propensity Score構成項目間のすべての交互作用を確認することができない
- Propensity Score解析の制約 (limitaion) を常に意識する必要がある。
- 感度分析

Propensity Scoreによるバイアスの減少

- Propensity Score作成に用いた背景因子のバイアスは、必ず減少する
- 採用したモデルの統計量に対し、背景因子の項目が多変量正規分布していると見なせるなら、バイアスはどの背景因子も等しく減少する。equal percent bias reducing(EPBR)

Propensity Score法利用の実務的な問題

(2) 選択したPropensity Scoreモデルの評価

- 作成したPropensity Score による、治療法A、Bの背景因子の偏りの評価
- バイアス減少率の評価
- 個々の背景項目の治療法A、B群の平均値の距離を評価
- Effect Size (Δ) = 平均値差 \div SE
- バイアス減少率 = PS調整 Δ' / PS非調整 Δ

年齢の新治療と従来治療群間の差と 傾向スコア調整後のt値

年齢の新治療と従来治療群間の差と傾向スコア調整後のt値

| | n | 平均 | 標準偏差 | 自由度 | t | p値 |
|-------------------------|-----|------|------|-----|------|--------|
| 従来治療 | 160 | 67.2 | 13.6 | | | |
| 新治療 | 85 | 65.0 | 12.2 | | | |
| 差 | | 2.2 | 13.2 | 243 | 2.25 | 0.0253 |
| 傾向スコア調整後 バイアス減少率 (%) | | 0.11 | | 239 | 0.11 | 95.1 |

Propensity Score法利用の実務的な問題

(2) 選択したPropensity Scoreモデルの評価

- SASプログラム例
治療法 Treat(A:1、B:0)
Propensity Score:PROPENSITY(区分1:1、区分2:2、・・・区分5:5)

- 治療法A、B間の距離
PROC MIXED;
MODEL Z1=TREAT /S;
RUN;

- Propensity Scoreで調整した治療法A、B間の距離

```
PROC MIXED;  
CLASS PROPENSITY;  
MODEL Z1=TREAT PROPENSITY /S;  
RUN;
```

Propensity Scoreによるバイアスの減少の確認

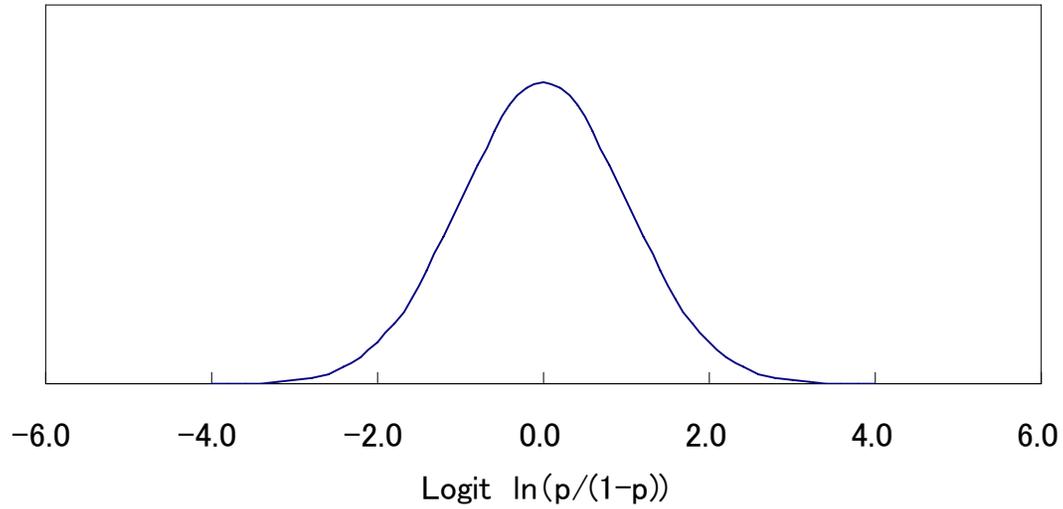
- Propensity ScoreによるBias Reductionを個々の背景因子(PSモデルに含む、含めない合わせて)ごとに吟味する
- 平均的なバイアス減少率を評価し、例えば70%以上のバイアス減少が得られなければ、Propensity Score法の効果は低いことを認識し、できれば90%以上のバイアスの減少が得られるようなPropensity Scoreを検討する。

- ① Propensity Scoreを調整に用いる場合、そのままの値は調整に使用せず、Propensity Scoreをその値の順序に対し5区分したものを
用いること(以下5区分変数と呼ぶ)

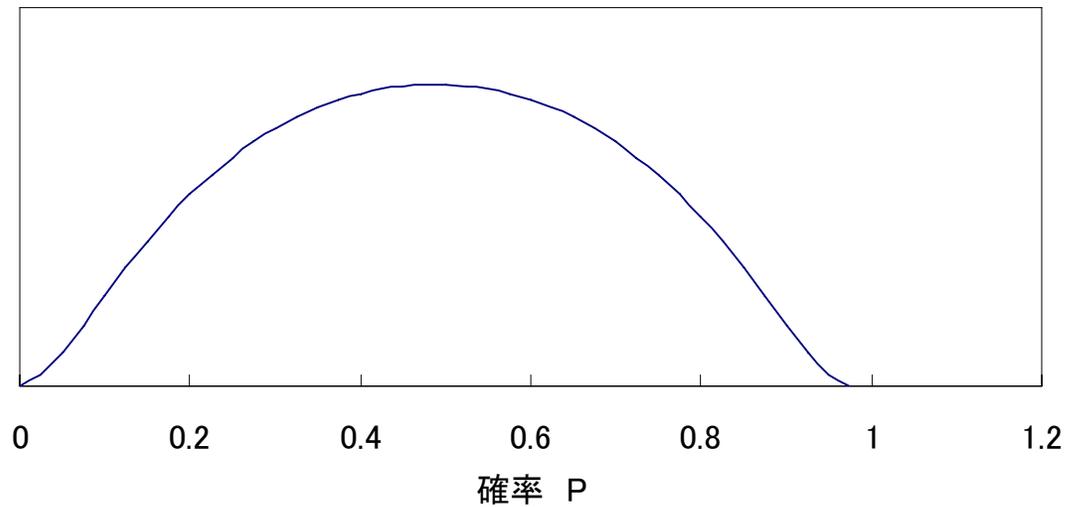
Propensity Score 5区分変数を用いる理由

- Propensity Score自体はResponse Outcomeと線形性が仮定できるわけではない
- Propensity Scoreの裾、すなわち、p値が1.0あるいは0に近い場合の数值は、背景項目の分布と特に乖離が大きい
- 実際的にイベント数から、5区分以上の区分は難しい
- Cochranや Rosenbaum and Rubinによれば
- PSによる背景因子の偏りによるバイアスの除去
 - ① 5区分変数 90%減少
 - ② 4区分変数 85%減少
 - ③ 3区分変数 80%減少

傾向スコア $\text{Logit } \ln(p/(1-p))$ 上での分布



傾向スコア確率P上での分布



Propensity Score 5区分変数 作成法

- 治療群と対照群を合わせたイベントが発生した症例のPropensity Scoreの順位に基づき適切に5区分する
- 治療群か対照群、解析対象のイベントが少ない方の群のイベントが発生した症例のPropensity Scoreの順位に基づき適切に5区分する
- Propensity Scoreの任意な値に対し、適切に5区分する

PS5区分変数 作成のポイント

- 治療群と対照群のイベントがPropensity Scoreの5区分で必ず重なっていること

機能しないPS区分変数

| PS区分 | 治療群 A | | 治療群 B | |
|------|-------|------|-------|------|
| | 例数 | イベント | 例数 | イベント |
| 1 | 100 | 0 | 100 | 60 |
| 2 | 100 | 2 | 100 | 28 |
| 3 | 100 | 10 | 100 | 10 |
| 4 | 100 | 28 | 100 | 2 |
| 5 | 100 | 60 | 100 | 0 |

適切な(理想的)PS区分変数

| PS区分 | 治療群 A | | 治療群 B | |
|------|-------|------|-------|------|
| | 例数 | イベント | 例数 | イベント |
| 1 | 320 | 20 | 30 | 20 |
| 2 | 85 | 20 | 35 | 20 |
| 3 | 50 | 20 | 50 | 20 |
| 4 | 35 | 20 | 85 | 20 |
| 5 | 30 | 20 | 320 | 20 |



PSマッチングは時間を見て



PS Matchingの適用

- 少数例の治療群と多数の対照群
- 必要な被験者背景は全例調査されている
- PS Matching後、治療群と選択された対照群のOutcomeを調査する

- コスト、時間の節約
- OutcomeをMatching後調査することにより解析中の選択バイアスを排除することが可能

- すでにOutcomeの調査が完了したデータに対してMatchingにより対照群の症例を選択することを散見するが、これは、上記のメリットが何も無く、また、症例の選択によるバイアスを否定できないので、筆者は推奨しない。

Propensity ScoreのMatchingへの応用 推奨方法

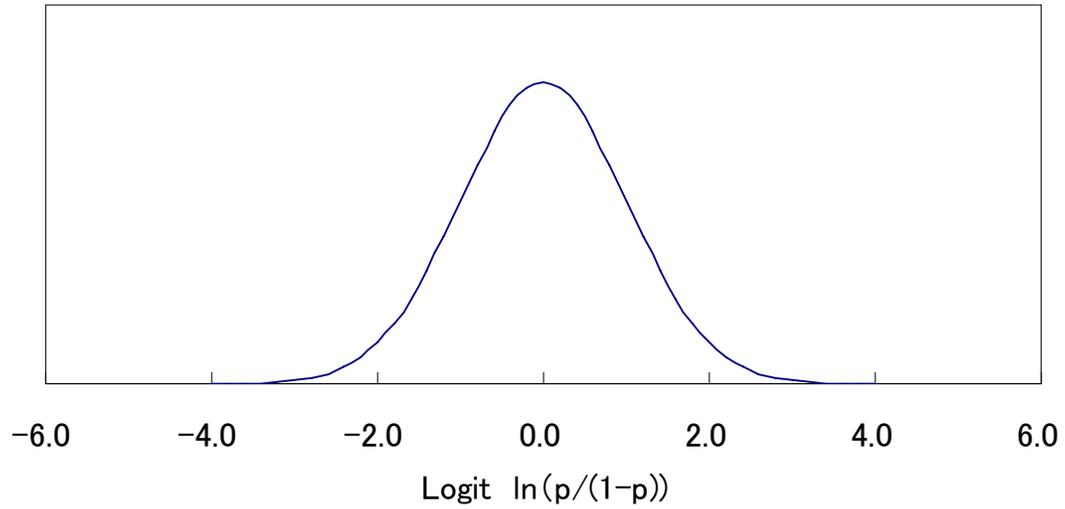
- ① すべての治療群、対照群の症例を用いて、適切な背景因子を選択したロジスティックモデルにより、Propensity Scoreのロジット値を計算する。
- ② 治療群から任意に1例 j を選択し、そのロジット値 $L_j \pm \text{Calipers}$ に入る症例を治療群、対照群から抽出する。
- ③ 抽出された症例を用い、より重要で少数の変数を用いた線形判別関数による治療群、対照群を判別するMahalanobisの汎距離を計算する。
 - 治療群 j 症例と最も近い Mahalanobisの汎距離を持つ症例を対照群より選択する。
- ②' 治療群から任意に1例 j を選択し、そのロジット値 $L_j \pm \text{Calipers}$ に入る症例を対照群から抽出する。
- ③' $L_j \pm \text{Calipers}$ に入る症例からランダムに1例抽出する。

Propensity ScoreのMatchingへの応用

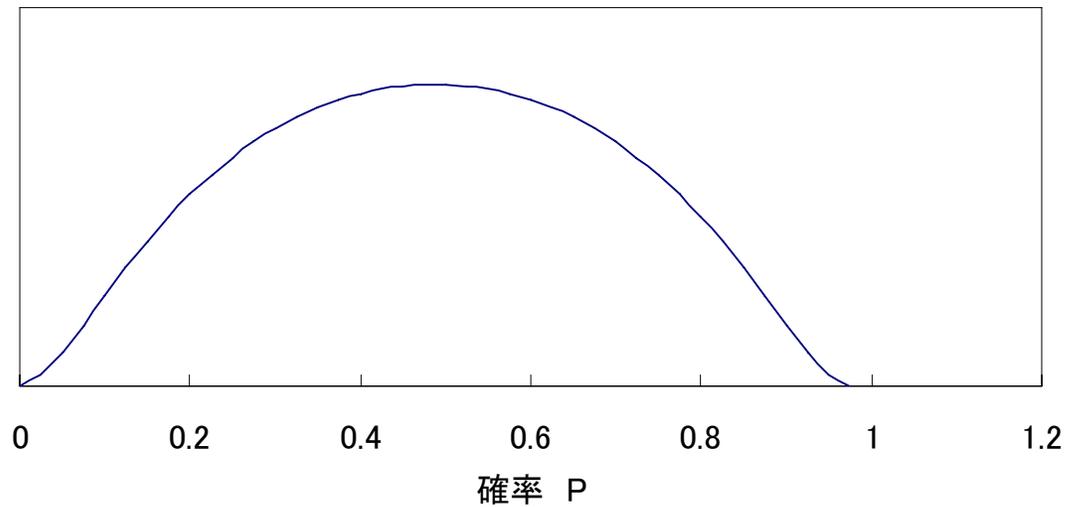
Matchingには確率値PではなくロジットLを使う

- ロジット値を用いる理由は、
- 条件が揃えば、ロジット値上の治療群、対照群の分布は、
- 分散の等しい正規分布となることが期待されるからである。

傾向スコア $\text{Logit } \ln(p/(1-p))$ 上での分布

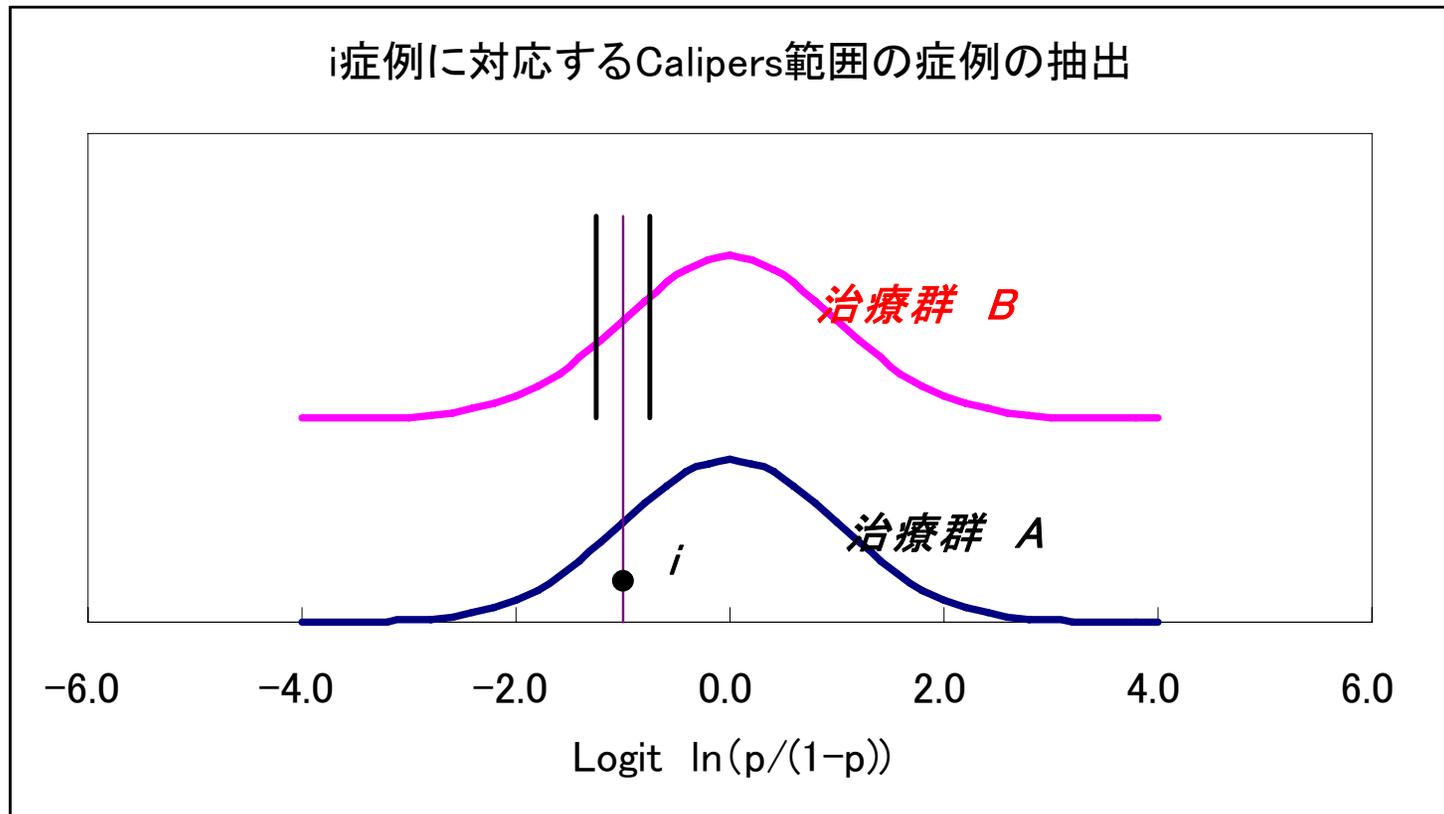


傾向スコア確率P上での分布



calipers

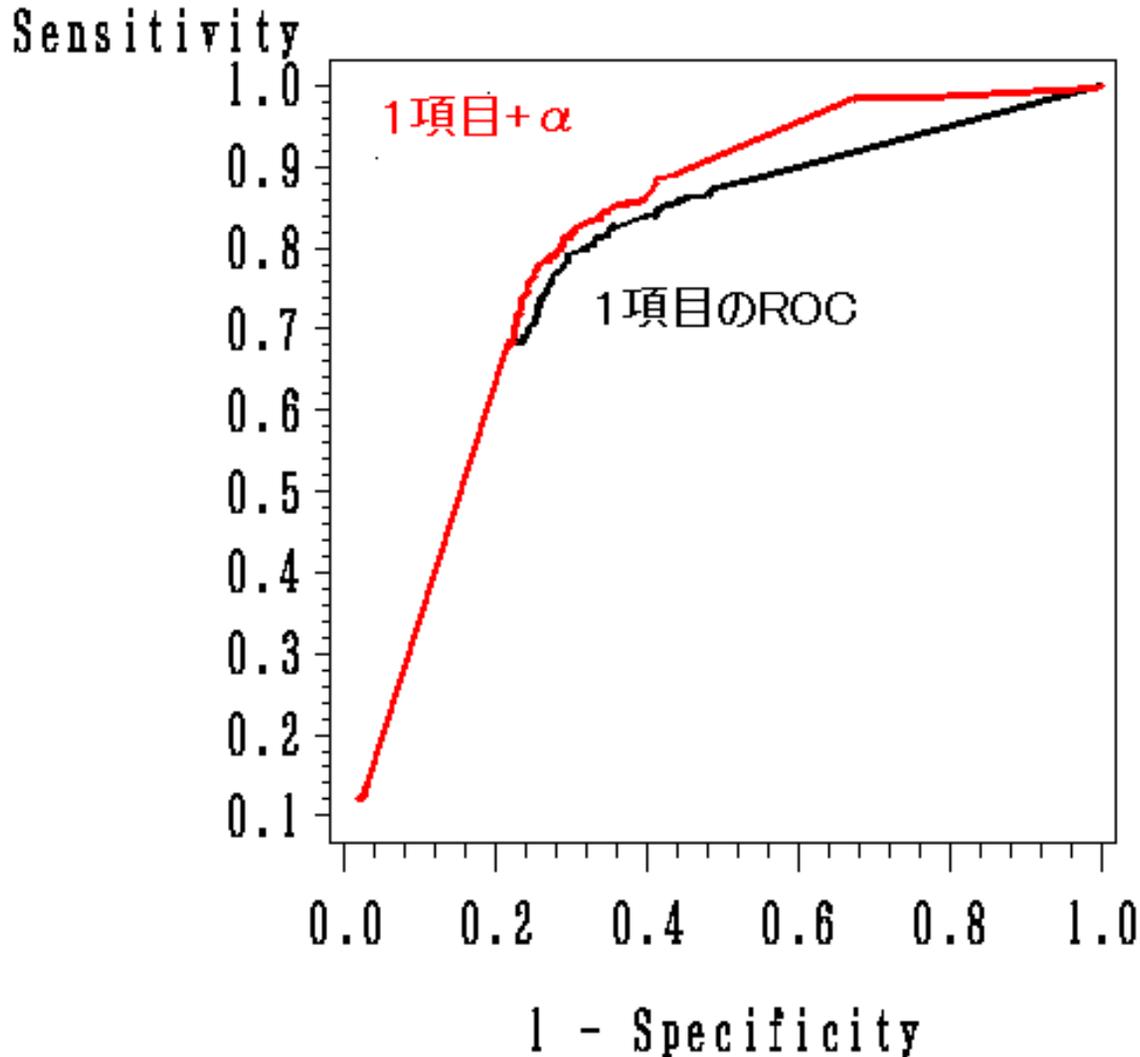
- Calipers: Rosenbaum、Rubinらは、propensity scoreロジット値の標準偏差の1/4を、caliperに推奨



Calipersならびに2段階抽出の理由

- PS Logisticモデルのパラメータ空間での非線形性
- Logisticモデル→辺縁部の重みが大きい
- 単純な1PSスコアによる治療群A症例*i*と最も近いPSを持つB症例の抽出 モデル依存バイアスの恐れ

Logisticモデル パラメーター空間の 非線形性



Propensity Score法利用のまとめ

- Propensity Score法の統計学的なバックグラウンドの理解が重要
- 対象となる臨床領域の基礎知識の理解し、Propensity Scoreを構成する背景項目の選択や、結果の妥当性の吟味が重要
- 統計家・臨床家のチームワークが重要
- Propensity Scoreを作成したら、そのbias reductionの大きさは常に確認すること

Propensity Score法利用のまとめ

- Propensity Score法はそれ自体が感度分析、
- 迷ったら感度分析、
- 結果の最終確認も感度分析