

Propensity Score法によるバイアスの調整法に関する実務的な問題点

古川敏仁、杉本典子
株式会社バイオスタティスティカル リサーチ

SASユーザー会総会2010

2010/7/27 11:00

1

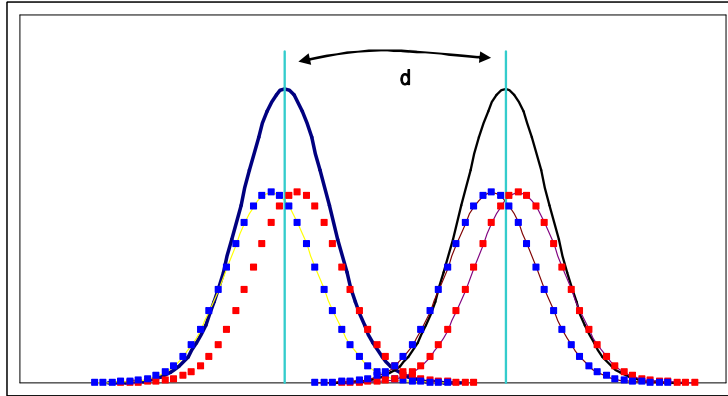
Propensity Score法とは

- 無作為化比較試験
 - 2つの治療法A、Bの比較において
 - 例数大ならば、
 - 観察された i 個の背景因子 z_1, z_2, \dots, z_i ,
 - 観察していない背景因子 z_{i+1}, z_{i+2}, \dots
 - が無作為化割付によってバランスが保たれる。
- 例えば 観察研究
 - 2つの治療法A、Bの比較において
 - 観察された背景因子、観察していない背景因子のバランスは保たれない可能性がある。

2

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

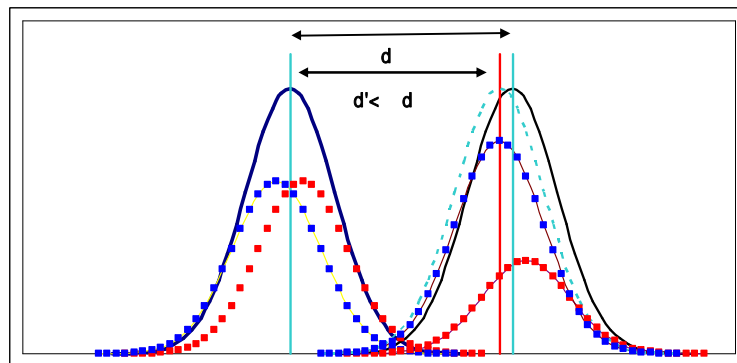
- 治療群効果 = $\theta_a - \theta_b = \theta_d$ は正しく推定される。



3

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

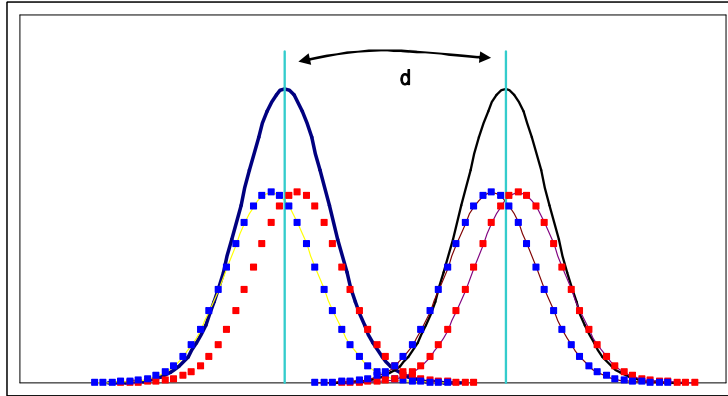
- 治療群効果 = $\theta_{a'} - \theta_{b'} = \theta_{d'} < \theta_d$
- バイアスが発生する。



4

Propensity Score法

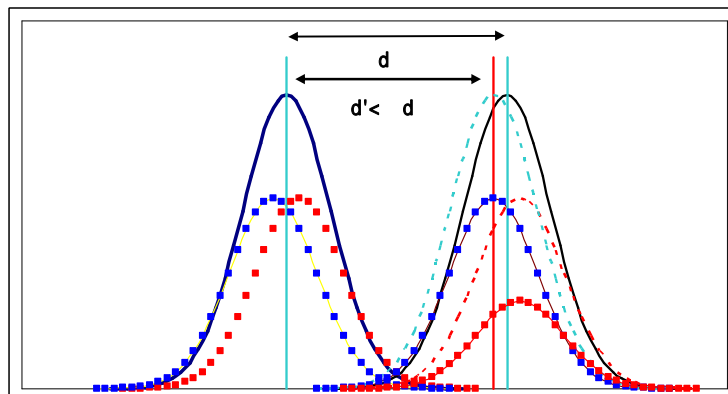
- Propensity Scoreと呼ばれるバランス変数を用いて治療群A、B群間の背景因子のバランスをとり
- θd を正しく推定する方法



5

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 θd を正しく推定する方法



6

観察研究の問題点

- 2つの治療法A、Bの比較において
 - 観察された*i*個の背景因子 z_1, z_2, \dots, z_i 、
 - 観察していない背景因子 z_{i+1}, z_{i+2}, \dots
 - バランスが保たれない可能性がある。
- であるならば
 - 観察された*i*個の背景因子 z_1, z_2, \dots, z_i
 - のバランスを保つ工夫をすることを考える。
 - 観察していない背景因子 z_{i+1}, z_{i+2}, \dots ごめんなさい

7

Propensity Score法

- 観察された予後因子 z_1, z_2, \dots, z_i から
- 個々の症例*j*が治療群AあるいはBが曝露される確率 P_j を求め

$$p_j = \text{prob}(\text{Drug} = A \mid z_{1j}, z_{2j}, \dots, z_{ij})$$

- A群とB群で確率 P が等しい症例を選択し比較する = matching
- 統計モデル的にA群とB群が同等な P のもとでの効果差を推定する = 調整

8

Propensity Score

- 治療群 Treat 1:A 0:B とコーディング
- 下記のSASプログラムで 背景因子 Z_1, Z_2, \dots, Z_i から予測される治療群Aが観察される予測確率Pを求めることができる
- Pを Propensity Score と呼ぶ。

```
PROC LOGISTIC;  
  MODEL Treat(event=1)=Z1 Z2 ... Zi;  
  OUTPUT OUT=OUT P=P;  
RUN;
```

9

Propensity Scoreの別な側面 例: 生存時間解析 (1)

- 治療群 A、Bのハザード比の推定

```
proc PHREG data=;  
  model Time*Censor(0) =Treat ;  
run;
```

- 共変量 Z_1, Z_2, \dots, Z_i の影響を除く

```
proc PHREG data=;  
  model Time*Censor(0) =Treat z1 z2 ... zi;  
run;
```

10

新たなる処置が必要となるまでの時間に対する解析
(単変量解析)

- 治療群 Bに対するAのハザード比は0.488
- $P < 0.0001$

Analysis of Maximum Likelihood Estimates								
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	
Treat	1	-0.7178	0.1735	17.12	<.0001	0.488	0.347	0.685

11

新たなる処置が必要となるまでの時間に対する解析
(変数選択による多変量)

- 22項目から7つの背景因子が逐次変数増減法で選択
- 治療群 Bに対するAのハザード比は0.669
- $P = 0.0512$

Analysis of Maximum Likelihood Estimates								
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	
Treat	1	-0.4025	0.1850	4.74	0.0512	0.669	0.465	1.002
性別	1	-0.4661	0.1933	5.81	0.0159	0.627	0.43	0.917
Z2	1	1.3675	0.2479	30.43	<.0001	3.925	2.415	6.381
Z3	1	-0.2395	0.1582	2.29	0.1301	0.787	0.577	1.073
Z4	1	0.3488	0.1440	5.87	0.0154	1.417	1.069	1.879
Z5	1	0.3738	0.2235	2.80	0.0944	1.453	0.938	2.252
Z6	1	0.4075	0.0770	28.01	<.0001	1.503	1.292	1.748
Z7	1	-0.7720	0.1846	17.49	<.0001	0.462	0.322	0.664

12

多変量解析の問題点

- 変数の選択は妥当か
- 多重共線、項目間相関によりパラメータ推定値の解釈が困難
- パラメータ空間の線形性は保証されているか
- パラメータ数増大により、パラメータ推定値の分散が大きくなる

例:分散 単項目:0.1735

多変量:0.1850

13

Propensity Scoreによる多項目解析の回避

- Propensity Scoreは
- 共変量 z_1, z_2, \dots, z_i を代表する
- 1変数のバランス変数とも解釈できる

14

Propensity Scoreによる多項目解析 の回避 (Matching)

- 性別、年齢 (65歳未満、65歳以上)、病型 (I、II、III)、総コレステロール (正常、高)、HDL (正常、低)、……
- Matchingすべき要因が多いとき、これらをすべてマッチした治療群A、治療群B集団を構成することは難しい
- 複数の背景要因を1変数に縮約したPropensity Scoreであれば、複数の背景要因を考慮したMatchingが可能

15

Propensity Score法利用の実務的な問題 (1)背景項目の選択

例えば、生存時間解析であれば、2つの治療法 A、Bの比較時に影響を与える可能性のある予後因子に関して

潜在的な背景因子の偏りをできるだけ防ぐため、可能な限り多くの背景項目をモデルに含める。

モデルの安定性の確保のため、例えば $P > 0.500$ となる項目はモデルから除外する。

- この2つの原則が教科書に、散見される。

16

Propensity Score法利用の実務的な問題 (1)背景項目の選択 推奨

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。

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。

それでもp値が0.70や0.80を超える項目がモデルに残ったら、そのとき考える。

最も重要なこと、

考えら得る変数の選択ルールをすべて試してみて、すべての結果が変わらないことを確認する。

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。

17

Propensity Score法利用の実務的な問題 (1)背景項目の選択 注意事項1

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

18

Propensity Score法利用の実務的な問題 (1)背景項目の選択 注意事項2

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

19

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

20

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;

21

表1 Propensityスコアで調整した場合のっ治療群間の偏り

	治療法		自由度	Logisticパラメーター				Effect Size	Bias Reduction (%)
	A	B		推定値	標準誤差	Wald ²	p値		
	例数	例数							
単変量ロジスティック回帰分析									
Sex	500	200	1	-0.115	0.338	0.12	0.7334	0.34	
Age	500	200	1	-0.020	0.012	2.59	0.1078	1.61	
NYHA	500	200	1	-0.281	0.476	0.35	0.5556	0.59	
.....							
心機能検査1	500	200	1	0.045	0.022	4.07	0.0436	2.02	
心機能検査2	500	200	1	0.023	0.015	2.36	0.1242	1.54	
QOL	499	180	1	-0.009	0.006	2.17	0.1404	1.47	
Propensity Score調整ロジスティック回帰分析									
Sex Women vs Men	500	200	1	-0.047	0.369	0.02	0.8987	0.13	37.4
Age	500	200	1	-0.004	0.014	0.07	0.7926	0.26	16.3
NYHA	500	200	1	-0.224	0.530	0.18	0.6733	0.42	71.5
.....							
心機能検査1	500	200	1	0.015	0.025	0.38	0.5387	0.62	30.4
心機能検査2	500	200	1	0.001	0.017	0.01	0.9332	0.08	5.4
QOL	499	180	1	-0.010	0.006	2.47	0.1161	1.57	106.8
Bias Reduction平均									30.6

22

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))

23

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。

24

表1 Propensityスコアで調整した場合のっ治療群間の偏り

	治療法		自由度	Logisticパラメーター				Effect Size	Bias Reduction (%)
	A	B		推定値	標準誤差	Wald ²	p値		
単変量ロジスティック回帰分析									
Sex	500	200	1	-0.115	0.338	0.12	0.7334	0.34	
Age	500	200	1	-0.020	0.012	2.59	0.1078	1.61	
NYHA	500	200	1	-0.281	0.476	0.35	0.5556	0.59	
.....
心機能検査1	500	200	1	0.045	0.022	4.07	0.0436	2.02	
心機能検査2	500	200	1	0.023	0.015	2.36	0.1242	1.54	
QOL	499	180	1	-0.009	0.006	2.17	0.1404	1.47	
Propensity Score調整ロジスティック回帰分析									
Sex Women vs Men	500	200	1	-0.047	0.369	0.02	0.8987	0.13	37.4
Age	500	200	1	-0.004	0.014	0.07	0.7926	0.26	16.3
NYHA	500	200	1	-0.224	0.530	0.18	0.6733	0.42	71.5
.....
心機能検査1	500	200	1	0.015	0.025	0.38	0.5387	0.62	30.4
心機能検査2	500	200	1	0.001	0.017	0.01	0.9332	0.08	5.4
QOL	499	180	1	-0.010	0.006	2.47	0.1161	1.57	106.8
Bias Reduction平均									30.6

25

Propensity ScoreのMatchingへの応用

- Propensity Score でMatchingした対照群に対して治療群の効果を比較すれば、
- Quasi-randomized 比較治療群間で被験者背景のバランス化
- 治療効果の比較からバイアスを除外することができる。

26

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。

27

Propensity ScoreのMatchingへの応用 推奨方法

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。

治療群から任意に1例を選択し、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。

抽出された症例を用い、より重要で小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。

治療群から任意に1例を選択し、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。

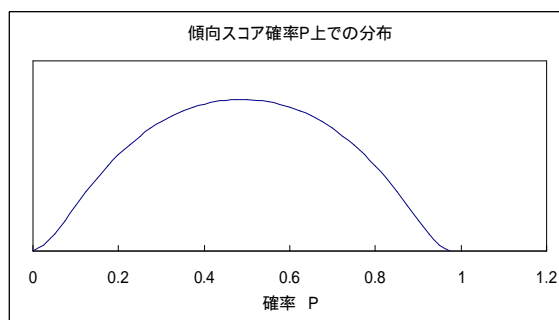
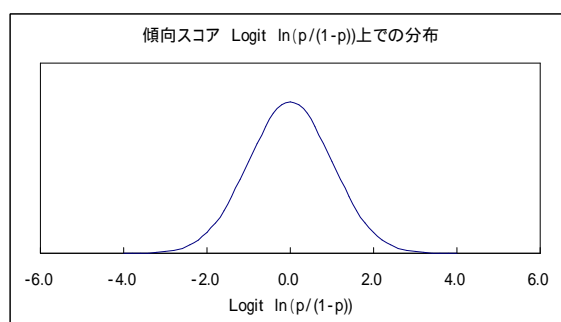
$L_j \pm \text{Calipers}$ に入る症例からランダムに1例抽出する。

28

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。

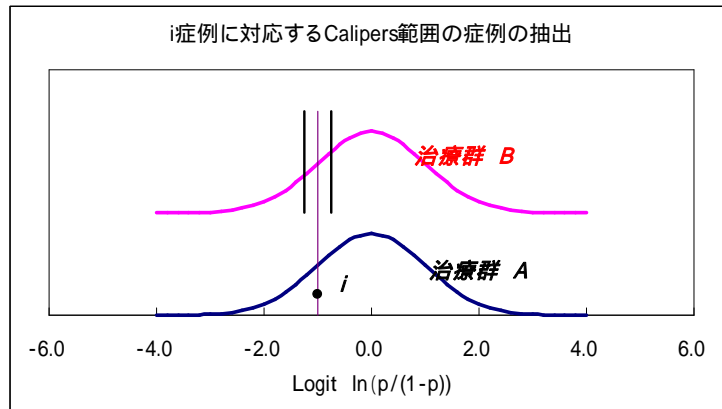
29



30

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→辺縁部の重みが大い
- 単純な1PSスコアによる治療群A症例と最も近いPSを持つB症例の抽出 モデル依存バイアスの恐れ

Propensity Scoreの調整への応用

治療群vs対照群を応答変数とし、適切な背景因子を説明変数としたロジスティックモデルで、Propensity Scoreを算出する。

- Propensity Scoreをその値の順序に対し5区分する(以下5区分変数と呼ぶ)。
- 5区分変数を層別因子、あるいは名義カテゴリ変数として共変量に用い、調整した治療効果の差を推定する。

33

PS調整 SASプログラム例

- 共変量として調整: ダミー区分変数
- PS1: 区分1=1, else 0, ... PS5: 区分5=1, else 0

```
proc PHREG data=;  
  model Time*Censor(0)=Treat ps1 ps2 ps3 ps4 ;  
run;
```

- 層化要因として調整

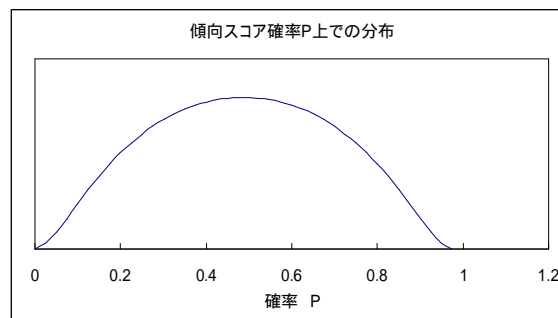
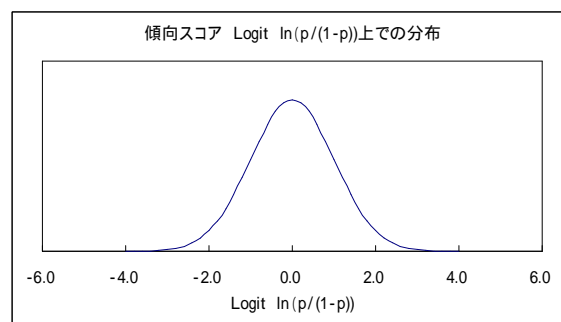
```
proc PHREG data=;  
  model Time*Censor(0)=Treat ;  
  strata PS;  
run;
```

34

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
- 傾向スコアによる背景因子の偏りによるバイアスの除去
 - 5区分変数 90%減少
 - 4区分変数 85%減少
 - 3区分変数 80%減少

35



36

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

37

PS 5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

38

機能しないIPS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

39

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

40

生存時間解析におけるPS区分変数の使い分け

- 5区分変数を層別因子として使う
- ハザード比の推定はより妥当(5区分間の生存関数に比例ハザード性を仮定しなくても良い)
- 基準ハザード関数を層ごとに仮定するため、治療効果間のハザード比の信頼区間は大きくなってしまいうデメリット

- 5区分変数を名義尺度の共変量として用いる。
- 背景因子の影響を推定できる
- 基準ハザード関数が1つなため、治療効果間のハザード比の推定精度が、5区分変数を層別因子として用いるときよりも狭い可能性
- しかし、多くの場合、5区分変数に比例ハザード性を期待するには無理な場合がある。

41

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

42

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