

バイアスを減少させるための 傾向スコアを応用した解析事例

長谷川 要

キリンビール株式会社 医薬カンパニー

開発部 臨床データ統括担当 統計解析G

E-mail: hasegawak@kirin.co.jp

今日話す内容

- 傾向スコア(Propensity Score)というものを紹介します
 - 新規的な話ではありません
- 国内の論文でこの手法を使ったものをあまり見うけません
- でも、結構便利な手法であり、応用範囲は広いと思います
 - ということで、傾向スコアを国内に広める布教活動的な発表となります

無作為化試験の話

- 無作為化試験の利点としては、
 - 特性の均一化
 - 選択による偏りを減少
 - 統計的推論が可能
- 無作為化比較試験なら、
 - 患者背景因子など比較群間でバランスが取れているから、例えば薬剤の効果などを群間で比較するのは容易である。

一方で、

- 無作為化試験でない場合、たとえば、疫学的な研究目的で事後的に対象集団を選択した場合、
 - しばしば、患者背景因子など比較群間でバランスが取れていないため、例えば薬剤の効果などを群間で比較するためには、何らかの調整が必要になってきます。

このような場合どう対処する？

- 教科書的には、
 - 多変量回帰分析
 - 交絡要因と考えられる多くの要因をモデルに入れ、自分が知りたい要因の影響を相対危険度の値として見ていく。
 - 層別化
 - 比較群間で交絡因子の影響を少なくするために、各交絡因子で層別を行い、その層の条件を満足する対象者のみを用いて比較する。
 - マッチング
 - 比較群間で交絡因子を強制的に同じにすることにより、群間の比較可能性を高める。

今日は、マッチングに着目！

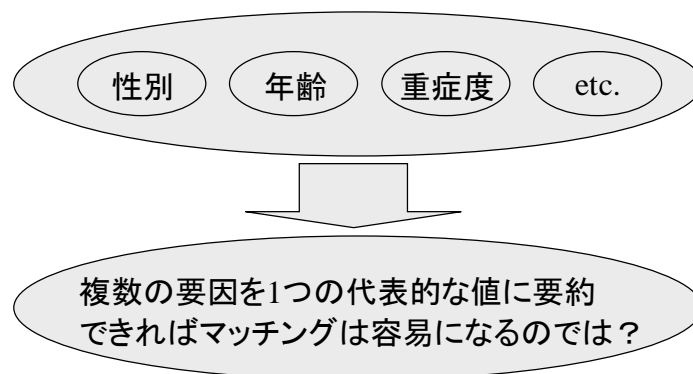
- マッチングとは、
 - 症例対照研究等において、交絡因子の条件ができるだけ似かうように症例群と対照群の2群を選び、交絡因子の影響をあらかじめ小さくする手法である。
 - 一般的にはマッチド・ペア法と呼ばれる方法で、症例群の性別、年齢等と同一な患者を対照群から選択し、マッチングを行います。

マッチングの問題点

- マッチングの問題点として、
 - 多数の因子についてマッチングを行ったり、きわめて厳密な条件を定めてマッチングを行うと、対となる対象を探せなかったりすることがある。
 - 例)
性別(2分類)、年齢(6分類)、疾患重症度(4分類)の3つの要因でマッチングを行う場合には、 $2 \times 6 \times 4 = 48$ 通りの組み合わせがある。さらに要因を追加することにより、組み合わせの数は指数的に増加する。

解決策はないのか？

- マッチングの問題点の解決策



傾向スコアの登場

- 複数の背景因子を1つの値に要約するための方法として、傾向スコア (Propensity Score) の利用が考えられる。
- 傾向スコアをマッチングの因子として利用すれば、容易に比較対照群の選択が可能となる。

傾向スコアとは？

- 傾向スコアとは共変量によって与えられる治療を受ける連続確率値として定義される。
- 傾向スコアはロジスティックモデルによって算出することができる。

$$p = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)}$$

傾向スコアの応用方法

- 傾向スコアが等しい対象者間では共変量が似通う性質がある。
- この性質を利用することにより、共変量を調整した解析を行うことができる。具体的には、傾向スコアの値を
 - マッチングの因子
 - 層別因子
 - 共分散分析の共変量

SASによる算出方法

- 傾向スコアはロジスティックモデルにより計算されることからLOGISTICプロシジャを用いて容易に求められる。

```
PROC LOGISTIC DATA=SAMPLE;  
    MODEL TREAT=AGE SEX HEIGHT WEIGHT;  
    OUTPUT OUT=OUTDATA PROB=PR;  
RUN;
```

SASの出力結果

- LOGISTICプロシジャにより計算され、出力されたデータセット中の傾向スコアの一部を紹介します。

SAS output.

OBS	TREAT	SEXCD	AGE	WEIGHT	HEIGHT	PR
1	なし	女	71	34.0	151.5	0.04277
2	なし	女	48	50.7	165.0	0.14289
3	なし	男	62	64.0	160.0	0.18827
4	なし	男	37	65.0	172.5	0.28297
5	あり	女	78	43.6	158.0	0.04280
6	あり	女	50	46.9	164.5	0.14297
7	あり	男	74	50.4	160.5	0.18827
8	あり	男	43	69.9	169.0	0.28303

解析事例

- エスポー®注射液
 - エリスロポエチン(Epoetin alpha)
 - 165個のアミノ酸残基からなる遺伝子組換え糖タンパク質
- 薬理作用
 - 前期赤芽球細胞および後期赤芽球細胞に作用して赤血球の造血を亢進させる
- 主な適応症
 - 腎性貧血、自己血貯血、未熟児貧血など

腎性貧血

- 発現機序
 - 腎機能低下に伴い腎臓の細胞から造血ホルモンであるエリスロポエチンの分泌が低下し、赤血球の産生が低下し貧血症状が来す状態
- どんな患者に発症するのか
 - 腎性貧血患者のほとんどは透析施行中であり、また、透析前の腎不全患者も発症することがある

腎性貧血患者における問題点

- 腎機能悪化に伴い種々の合併症を併発しているために様々な薬剤が併用されている
 - ビタミン剤
 - 炭酸カルシウム
 - ACE阻害薬、etc.←降圧剤
- 併用薬剤がEpo製剤の貧血効果に影響を与えるとする報告多数あり
 - ACE阻害薬との関係の論文に着目

ACE阻害薬のEpo製剤の薬効 に及ぼす影響のメカニズム

- Epo産生阻害
 - ACE阻害薬が腎臓に作用し、体内でのEpoの産生能力を低下させるため
- 造血肝細胞の分化・増殖阻害
 - ACE阻害薬が赤芽球系細胞に作用し、分化増殖を阻害するため
- 本当にそうなのか？
 - だとしたら、臨床症状に現れるはず

そこで...

- 目的
 - ACE阻害薬のEpo製剤に対する薬効の影響を検討する
- 対象
 - エスポー注射液使用成績調査のデータを用いる
- 解析方法
 - ACE阻害薬の併用の有無別に患者を2群に分け、貧血の改善効果を比較する

患者背景

変数	分類	ACE 阻害薬		群間比較
		併用なし n=1,884	併用あり n=329	
性別	男	1,025(54.4)	215(65.3)	P<0.001
	女	859(45.6)	114(34.7)	
年齢(歳)	平均	56.9	53.6	P<0.001
	SD	13.3	13.4	
体重(kg)	平均	53.1	55.0	P<0.001
	SD	9.5	10.1	
透析歴	1年未満	1,006(53.4)	156(47.4)	P=0.045
	1年以上	878(46.6)	173(52.6)	

データの問題点

- 交絡因子がひょっとしたらあるかも？
 - 患者背景を見るとACE阻害薬併用の有無群間で性別、年齢、体重、透析歴、etc.で有意な差が認められている
 - そもそも使用成績調査のデータであり、所詮無作為割付を行った試験データではない
- 交絡因子の影響を取り除いた上で比較する必要があるのでは？

再び傾向スコアの登場

- 交絡因子の影響を排除するために、患者背景因子を群間でそろえた上で2群間の比較を行えば、いいのでは？
- そのためには、どうする？
 - 傾向スコアを用いて症例のマッチングを行い、患者背景因子をそろえる方法がよいのでは？

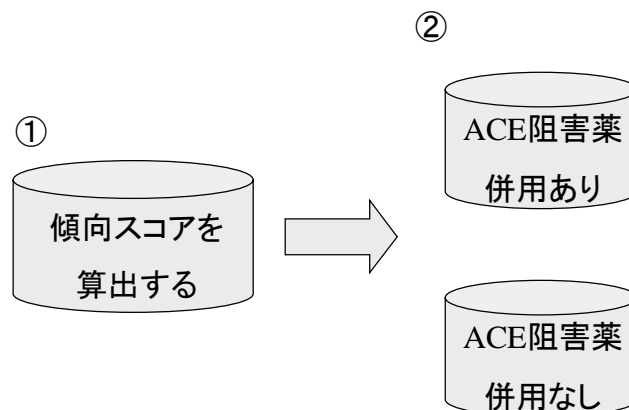
マッチング後の患者背景

変数	分類	ACE 阻害薬		群間比較
		併用なし n=329	併用あり n=329	
性別	男	218(66.3)	215(65.3)	P=0.805
	女	111(33.7)	114(34.7)	
年齢(歳)	平均	54.1	53.6	P=0.629
	SD	12.9	13.4	
体重(kg)	平均	55.0	55.0	P=0.963
	SD	9.4	10.1	
透析歴	1年未満	157(47.7)	156(47.4)	P=0.938
	1年以上	172(52.3)	173(52.6)	

マッチングを行うプログラム

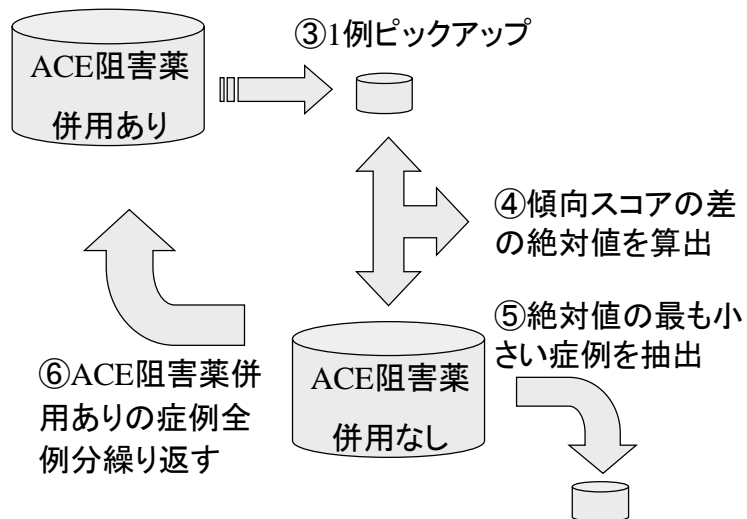
- 傾向スコアの算出
 - SAS/STATのLOGISTICプロシジャの利用
- マッチングの処理
 - SASにはマッチングを行うための機能は提供されていない
 - そのため、マッチングを行うためのプログラムを作成する必要がある

プログラムのロジック①



プログラムのロジック②

Logic.



いくつかの問題点

Problem.

- 共変量に欠損データが存在すると、傾向スコアの算出できなくなる。
 - 欠損データを何らかの手法で補ってから計算すれば回避できる
 - SAS Ver.8が提供するPROC MIを利用すればいいのかな？
- オーバーマッチング
 - マッチングによってかえって、偏る場合があることが知られている。
 - この辺はもう少し勉強が必要そう...

今後の課題

- 多群比較に対する拡張
 - 傾向スコアはロジスティック回帰モデルを前提として算出するために2群間の比較に対してのみ利用可能
→多群に拡張するには？
- 臨床試験データへの応用
 - ヒストリカルコントロールとの比較の際に？
 - 海外データと国内の試験データとの比較に使って見たら？

参考文献

- Rosenbaum PR、Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;70:41-55.
- D'Agostino RB. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine* 1998;17:2265-2281.
- Hayashi K、Hasegawa K、Kobayashi S. Effect of ACE inhibitors on Treatment of Anemia with Erythropoietin. *Kidney International*.(Accepted)