

観察研究において選択biasを制御するために用  
いられるPropensity Score IPTWと層化調整法  
の、頑健性の観点からの使い分けについて

2011年7月29日

古川敏仁

バイオスタティスティカル リサーチ

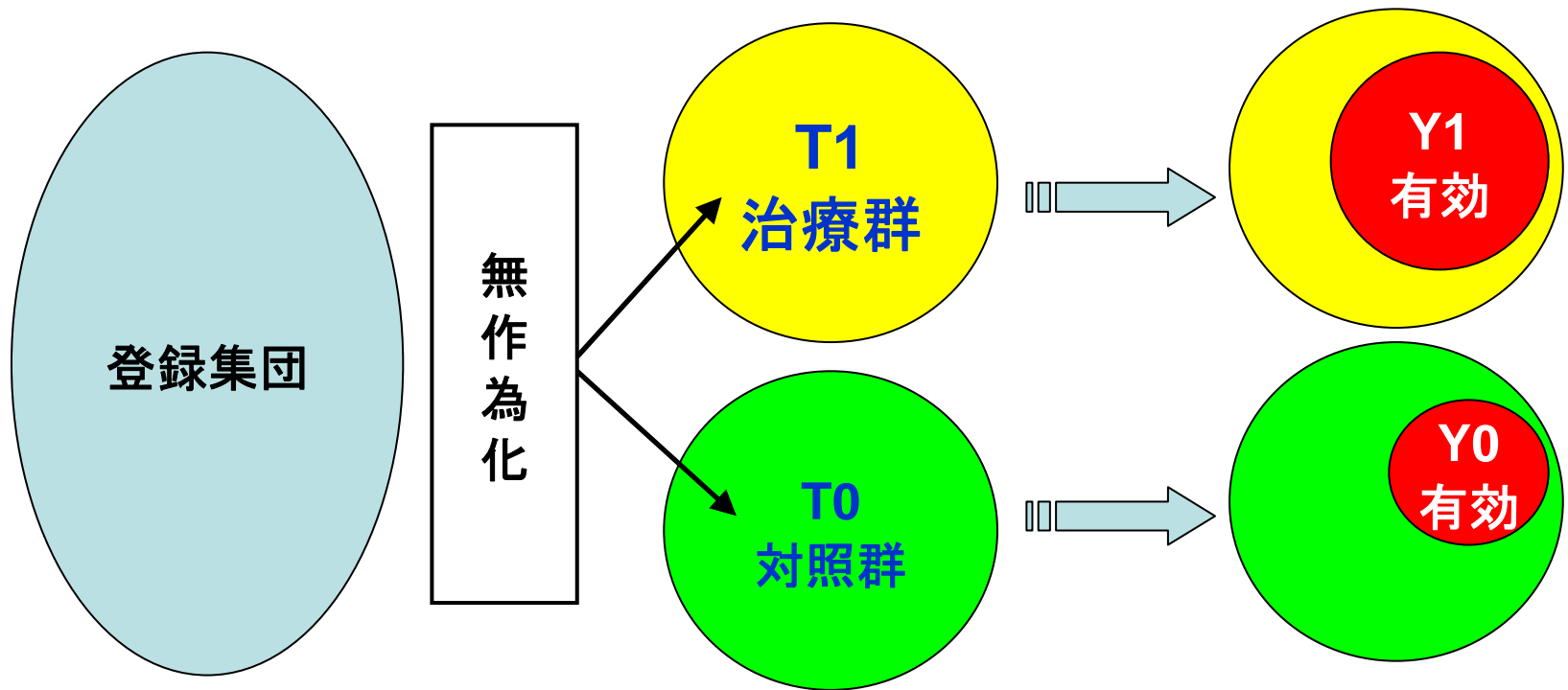
# 基本的な問題

- 2つの治療群T1、T0の応答Y1、Y0の差  $E[Y1 - Y0]$  を当該解析対象集団に対する治療効果として考える
- T1: 治療群    T0: 対照群
- Y1: 治療群における応答
- Y0: 対照群における応答

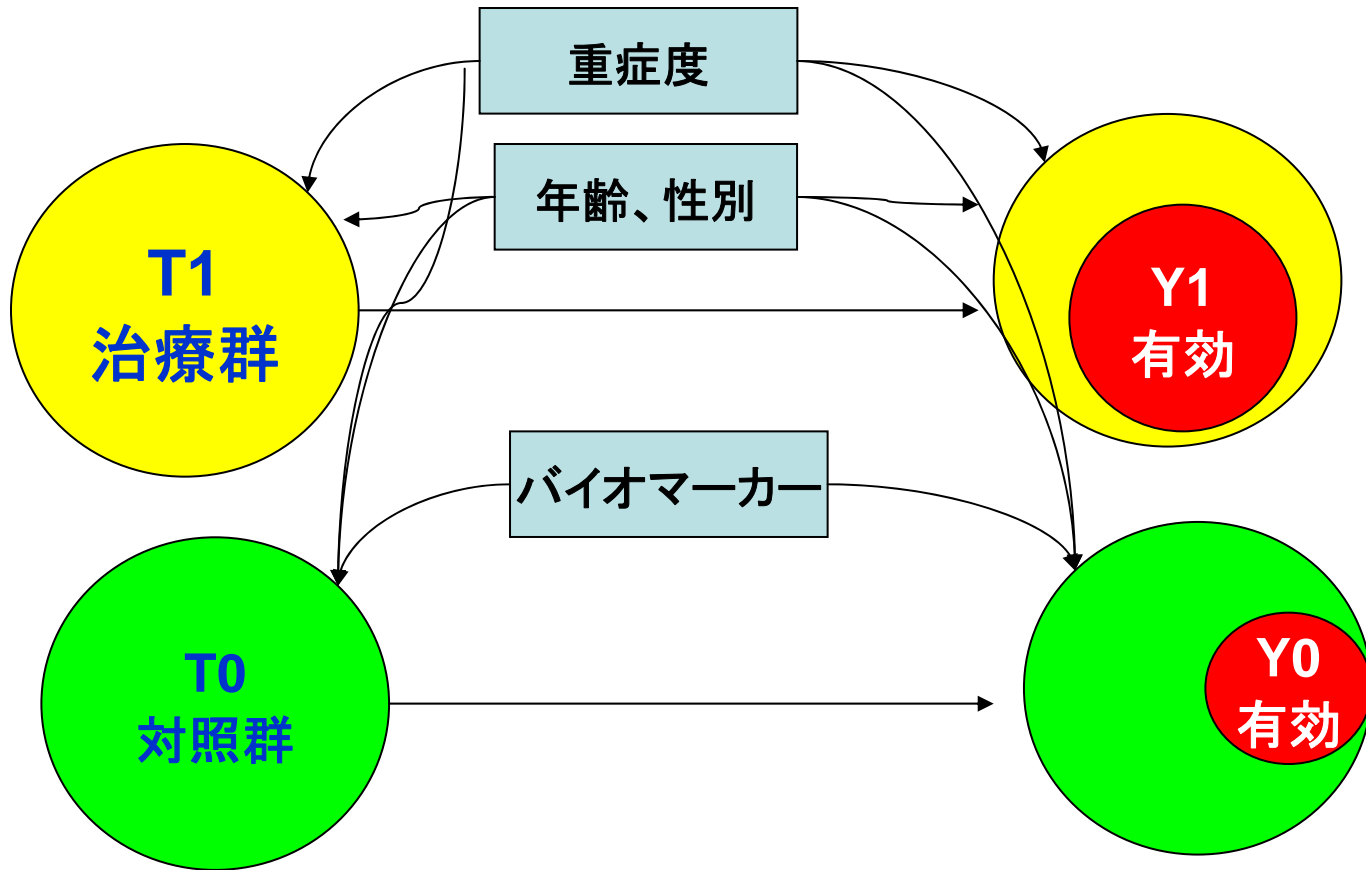
# 基本的コンセンサス

- 治療の応答 $Y$ には被験者背景 $Z$ が影響を与えるものがある。
  - がん 進行度(Stage)が高い症例は生存時間が短い
- 治療応答 $Y_1$ 、 $Y_0$ の比較を行うためには、比較する治療群以外の背景は等しい必要がある。
- 無作為化比較試験は $E[Y_1 - Y_0]$ の推定を可能にする
- では、観察研究は？

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



# 観察研究



# 基本的コンセンサス

- 治療T1とT0の治療効果の差が知りたければ
- 同じ被験者に同時にT1とT0の治療を行い、応答Y1とY0を観察し、その差の集団平均 $E[Y1 - Y0]$ (平均因果効果 average causal effects)を求めればよい。
- そうすれば、被験者背景の偏りは、比較する2群間で起こらない
- これもそのとおり、しかし、実際には残念ながら1人の被験者には1つの治療しか行うことができない
- つまり、Y1とY0のどちらかしか観察されない

# Rubin Causal Model

## 観察研究データから平均因果効果の推定

- 3つの確率変数を考え

Y: 応答

Z: 被験者背景

T: 割付 1=治療群、0=対照群

- $\text{Prob}(T=1|z_i) = e(z_i)$  propensity score

被験者背景  $z_i$ のもとで、治療群  $T=1$ に割り付けられる確率

# Rubin Causal Model

3つの基本的な確率の公式を利用するだけで、  
同じ被験者に同時にT1とT0の治療を行った(仮定)場合に計算  
できる平均因果効果が  
Y1、あるいはY0のどちらかが観察されない観察データから計  
算できる。

- 平均因果効果  $E[Y1 - Y0] = E[Y1] - E[Y0]$

$$= \{ E(Y1|T=0) - E(Y0|T=0) \} \text{Prob}(T=0) + \\ \{ E(Y1|T=1) - E(Y0|T=1) \} \text{Prob}(T=1)$$

$$E[Y1 - Y0] \approx \frac{1}{\sum_{i=(m+1)}^n \frac{1}{e(\mathbf{z}_i)}} \sum_{i=(m+1)}^n \frac{Y1_i}{e(\mathbf{z}_i)} - \frac{1}{\sum_{i=1}^m \frac{1}{1 - e(\mathbf{z}_i)}} \sum_{i=1}^m \frac{Y0_i}{1 - e(\mathbf{z}_i)} \quad (10)$$



# IPTW(Inverse Probability of Treatment Weighted)

$$E\{Y_1 - Y_0\} \approx \frac{1}{\sum_{i=(m+1)}^n \frac{1}{e(\mathbf{z}_i)}} \sum_{i=(m+1)}^n \frac{Y_{1i}}{e(\mathbf{z}_i)} - \frac{1}{\sum_{i=1}^m \frac{1}{1-e(\mathbf{z}_i)}} \sum_{i=1}^m \frac{Y_{0i}}{1-e(\mathbf{z}_i)} \quad (10)$$

10式をそのまま応用、応答 $Y_{1i}$ 、 $Y_{0j}$ に以下の重みを用いた重み付け平均の差が平均因果効果

$$\text{治療群の重み } w_{1i} = \frac{p_1}{e(\mathbf{z}_i)}$$

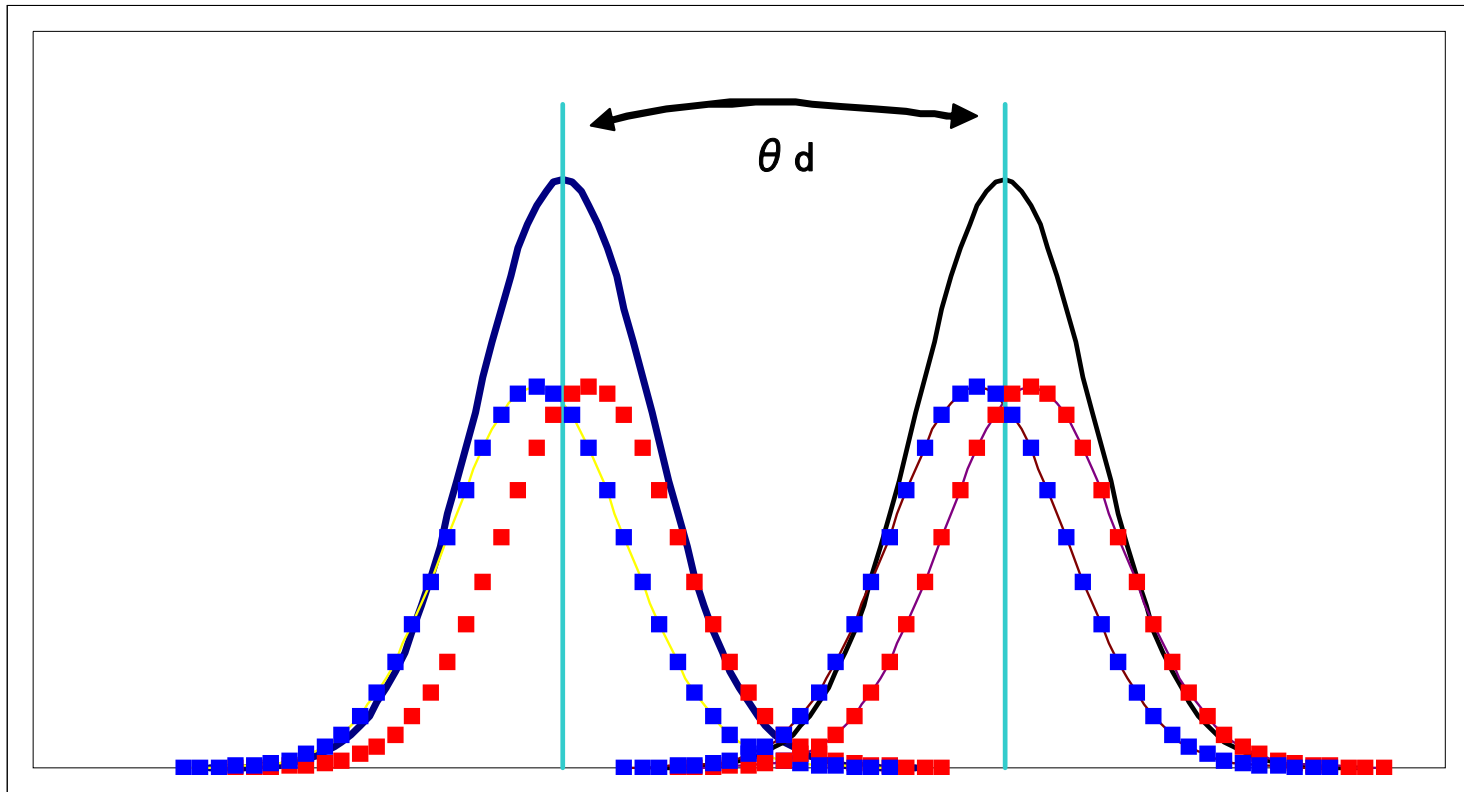
$$\text{対照群の重み } w_{0j} = \frac{1-p_1}{1-e(\mathbf{z}_j)}$$

# Propensity Scoreによる調整

- $e(z_i)$ 、あるいは、その導出変数 $f(e(z_i))$ を共変量、あるいは層化因子として用いる。
- 複数の背景因子を1つの変数 $f(e(z_i))$ に縮約したものと考えることができる。
- 複数の背景変数で治療効果パラメーターを調整する場合の多項目特性から来る問題を1つの共変量により置き換えることで回避する手段とも言える。

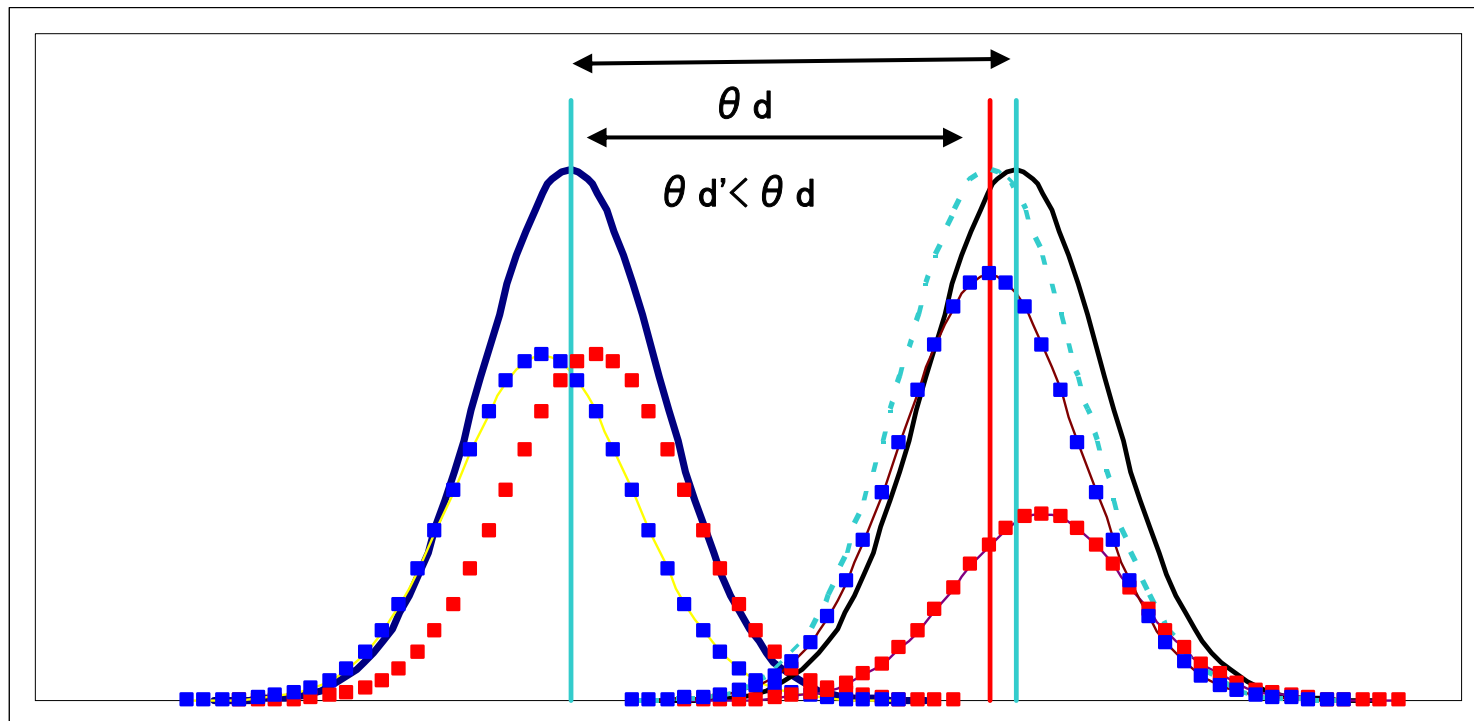
もし、背景因子(例:重症=青、軽症=赤)の分布のバランスが治療群A、B間で保たれていれば

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



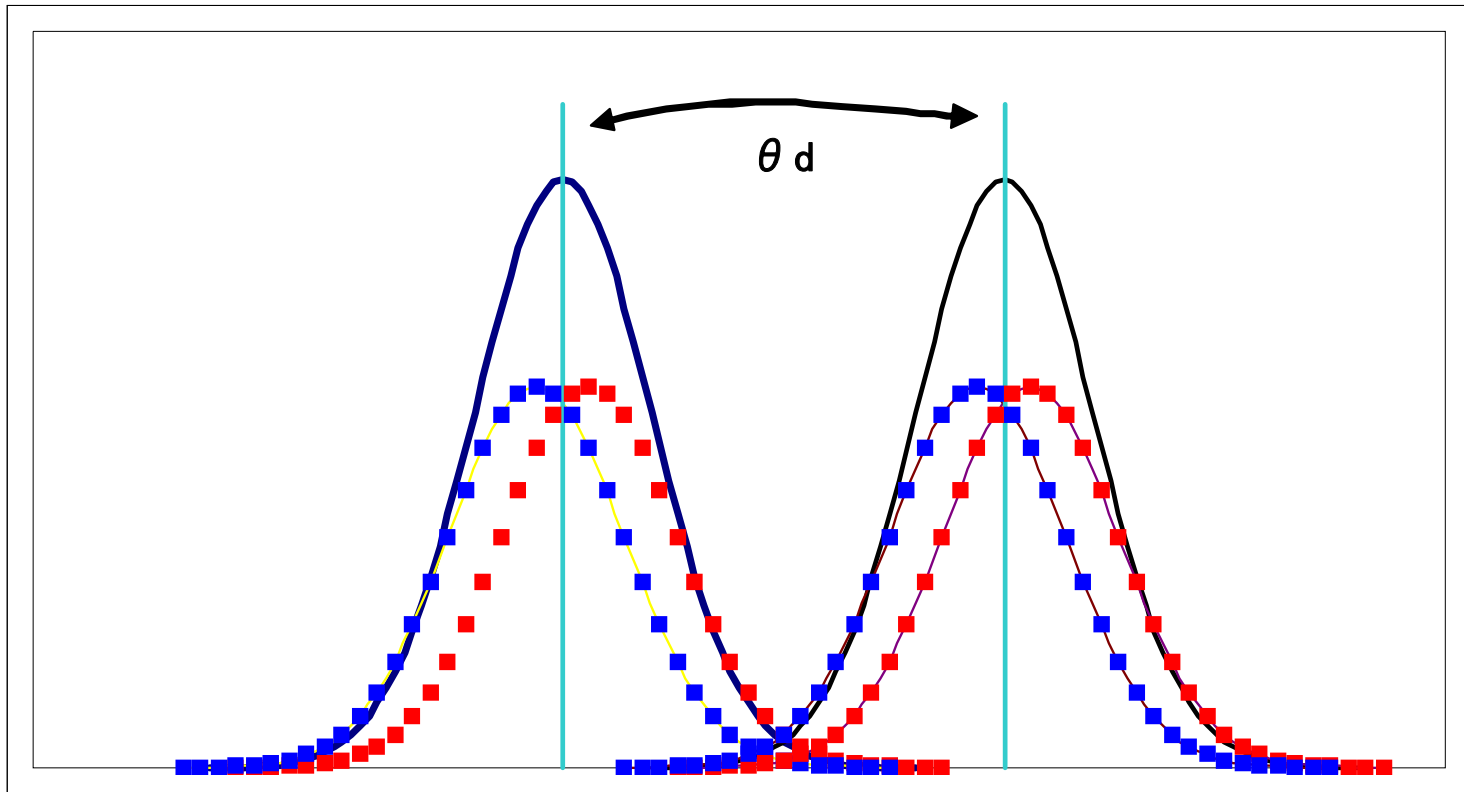
# もし、治療群Bの重症＝青の割合が治療群Aの割合よりも相対的に高くなると

- 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

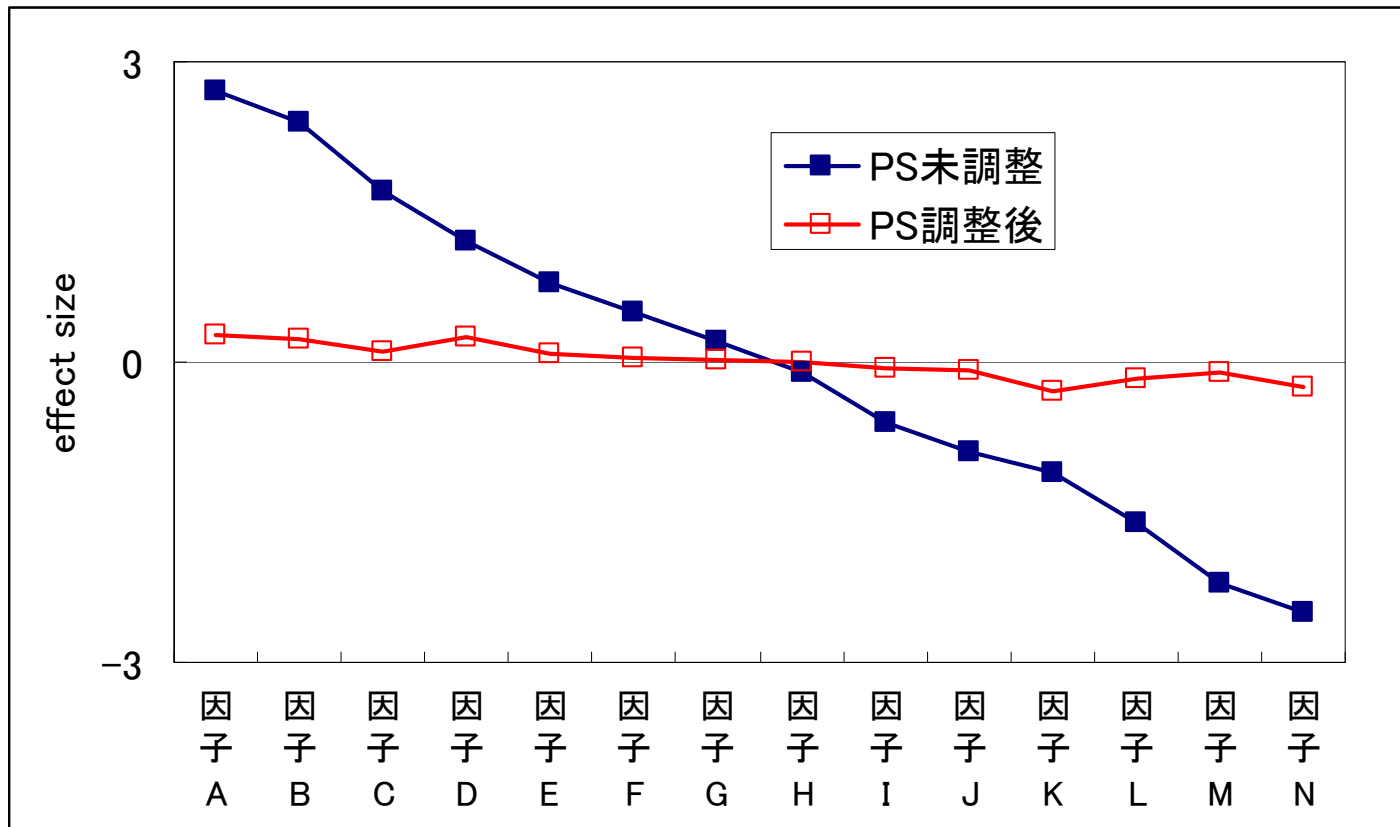
- Propensity Scoreと呼ばれるバランス変数を用いて相対的に当該因子に少ない群の重みを重くすることにより、治療群A、B群間の背景因子のバランスをとり
- $\theta_d$ を正しく推定する方法



# IPTW法とPSによる調整解析法に共通する重要事項

- 作成したpropensity scoreは、目的とする被験者背景のbiasをきちんと減少させるか。
- propensity score作成に用いた変数(被験者背景)は、治療効果にあたえるbiasをすべて説明できるものなのか=strongly ignorableな仮定を満たすといえるのか。

- *equal percent bias reducing*
- 基本的なPS確認方法
- $P > 0.80$ の背景項目はPSから除外

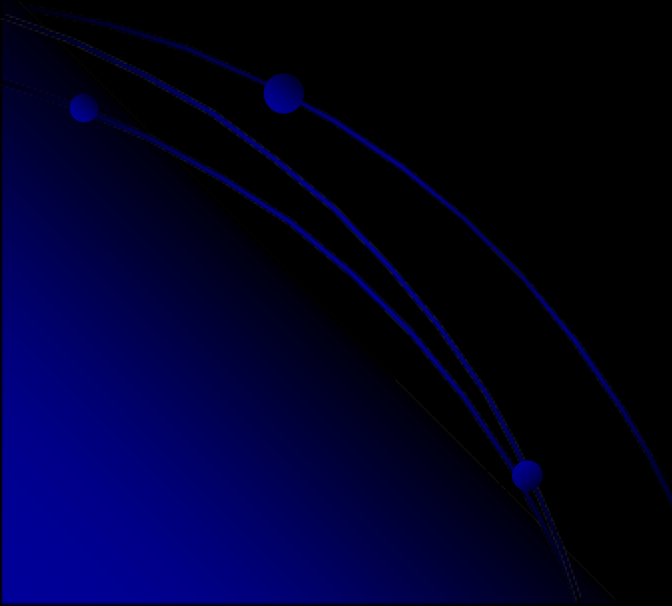


# IPTW法とPSによる調整解析法の使い分け

- 教科書によれば
- IPTWは調整解析法と比較して以下のような利点があることが記載されている8)
- Directに統計計量が計算できる。
- 調整解析法のようにPSの治療群、対照群の分布の重なりがあまりなくても計算できる。
- 真実だが問題も多い



# 実例 冠動脈疾患ステントデータ

- 2325例のデータ
  - イベント TLR発生までの時間 153例
  - 血管因子 あり:118例 なし:2207
  - 48項目の被験者背景によるPSスコア作成
- 

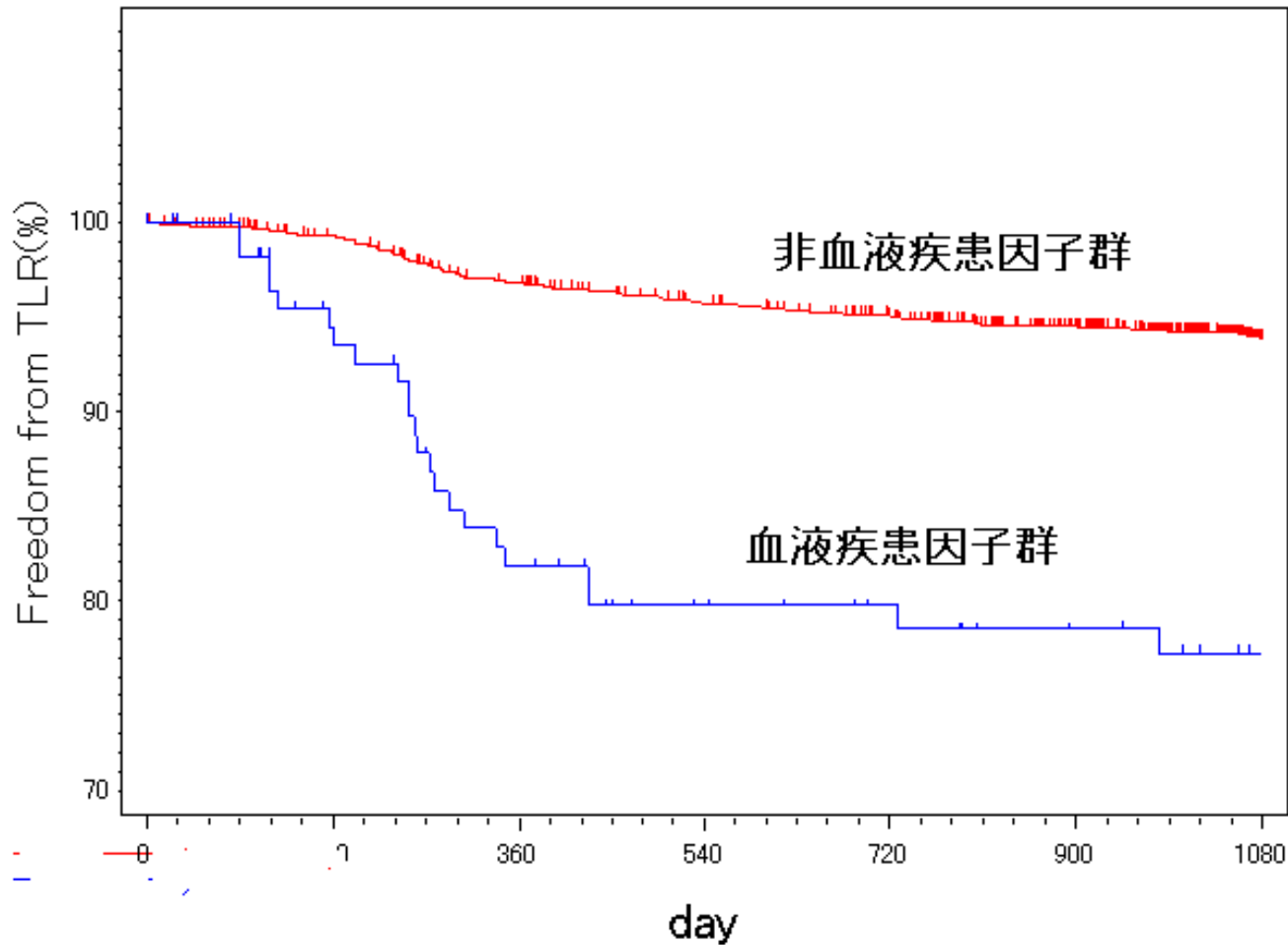
# 観察研究 ハザード比の探索

## 血管因子が一番大きなハザード比

被験者背景の偏りを調整しても、血管因子は意味がある項目か

項目	ハザード比 95%信頼区間			p値
	推定値	下限	上限	
血管疾患因子	4.423	( 2.836 - 6.897 )	<0.001 *	
手技A	2.496	( 1.413 - 4.406 )	0.002 *	
ステント長	1.987	( 1.376 - 2.869 )	<0.001 *	
糖尿病	1.846	( 1.336 - 2.552 )	<0.001 *	
再ステント術	1.813	( 1.243 - 2.645 )	0.002 *	
病変長	1.702	( 1.105 - 2.623 )	0.016 *	
Osital	1.674	( 1.159 - 2.416 )	0.006 *	
因子A	1.642	( 1.178 - 2.289 )	0.003 *	
因子B	1.483	( 0.907 - 2.426 )	0.116	
因子C	1.397	( 0.994 - 1.964 )	0.054	
因子D	1.393	( 0.948 - 2.049 )	0.092	
因子E	.....			

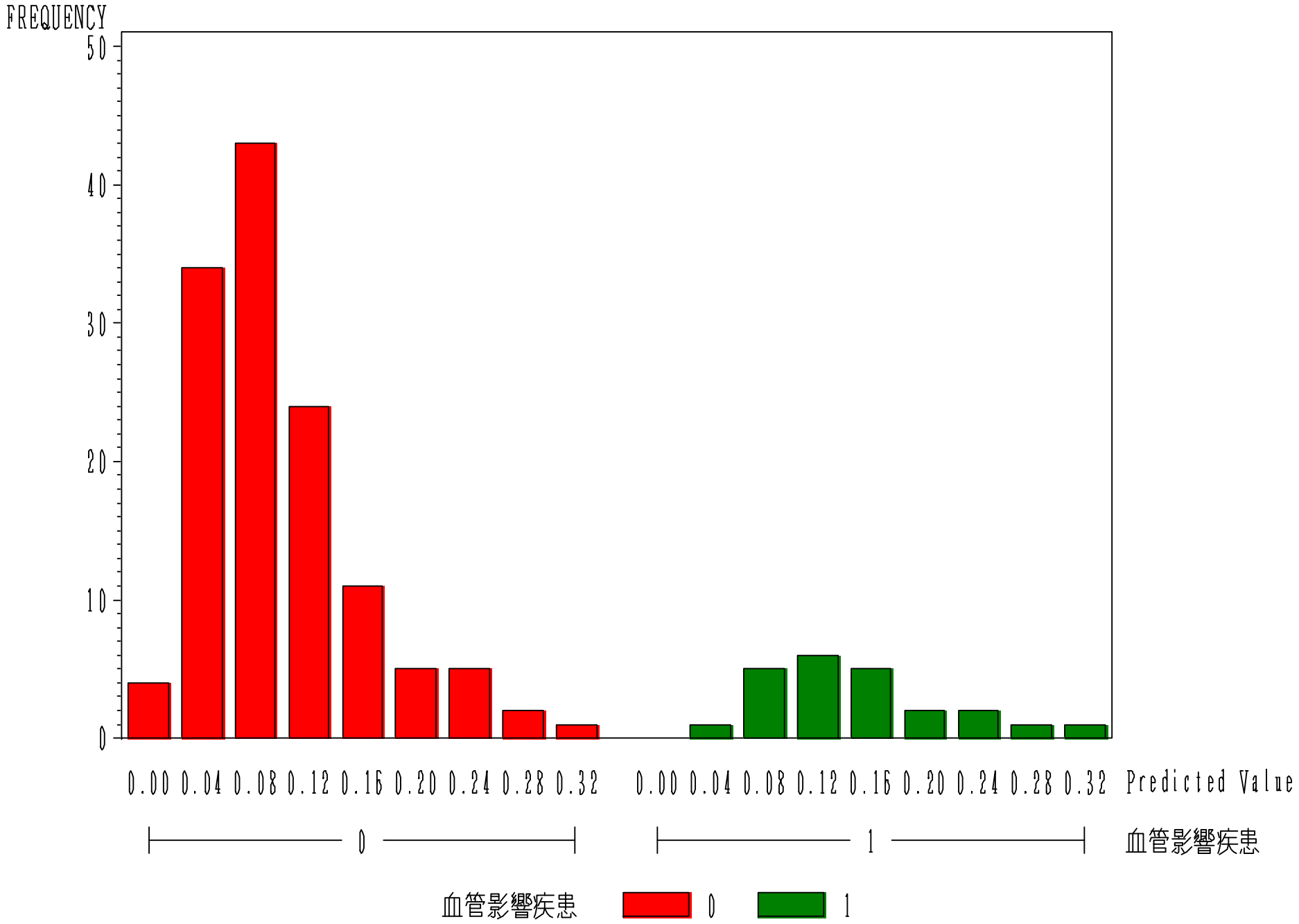
# 非調整Kaplan-Meier曲線



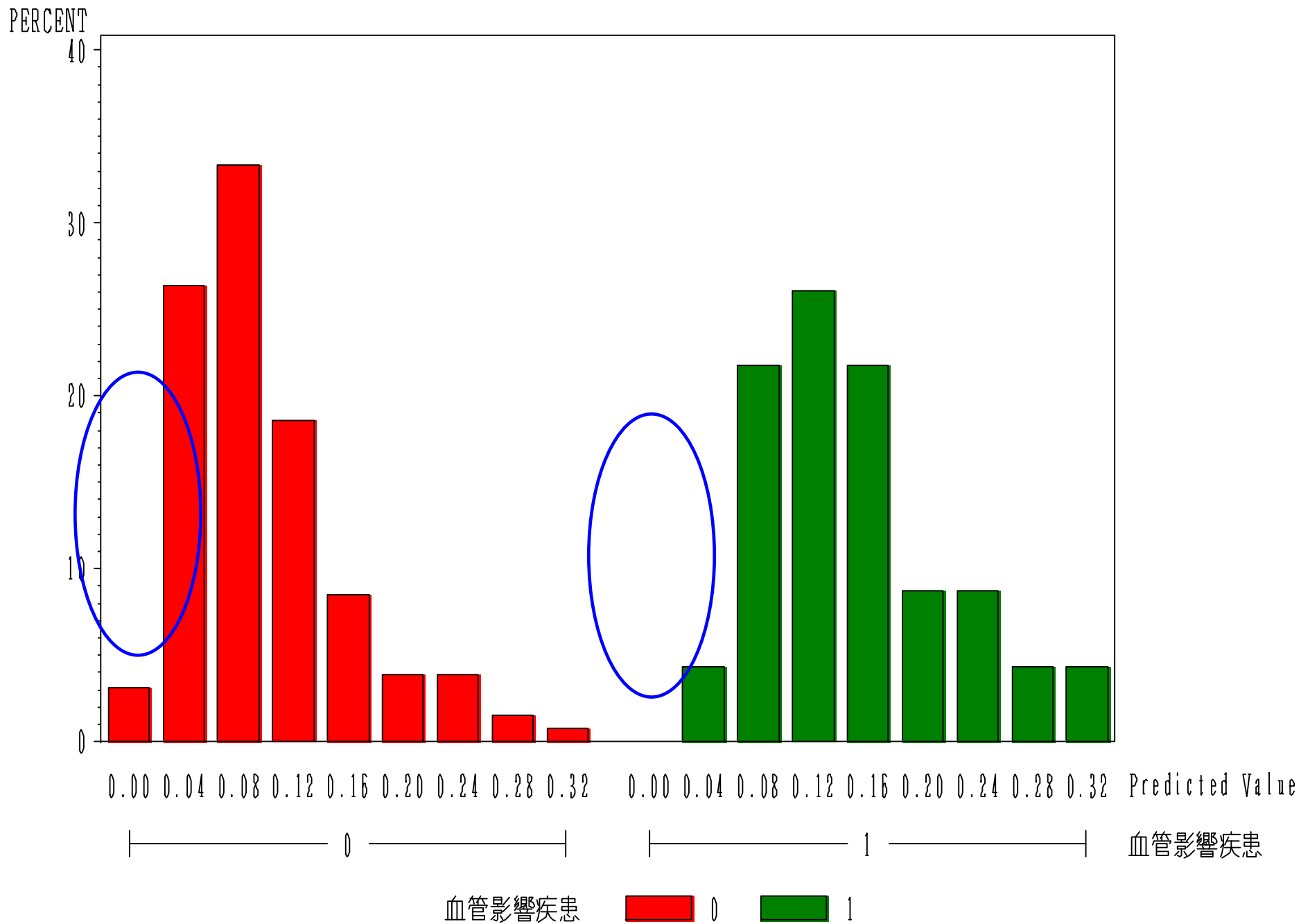
- ハザード比の感度分析
- 9項目の共変量調整多変量モデルはあまりハザードは低下しない 4.423→3.659
- しかし、PS調整するとハザード比は2.860、しかし、p値はまだ有意
- IPWではハザード比は2.201にp値も0.021に
- 一見IPTWが最も調整されているように見えるがこれでよいのか？

	ハザード比		95%信頼区間		p値
	推定値	下限	上限		
因子単独	4.423	2.836	6.897	<.0001	
9項目の共変量調整多変量解析	3.659	2.305	5.807	<.0001	
傾向スコア3区分調整 (共変量)	3.342	2.125	5.257	<.0001	
傾向スコア3区分調整 (層化)	3.260	2.069	5.135	<.0001	
傾向スコア4区分調整 (共変量)	2.935	1.866	4.616	<.0001	
傾向スコア4区分調整 (層化)	2.860	1.816	4.503	<.0001	
IPTW	2.201	1.126	4.300	0.021	

# PS上のイベントの頻度分布 実数



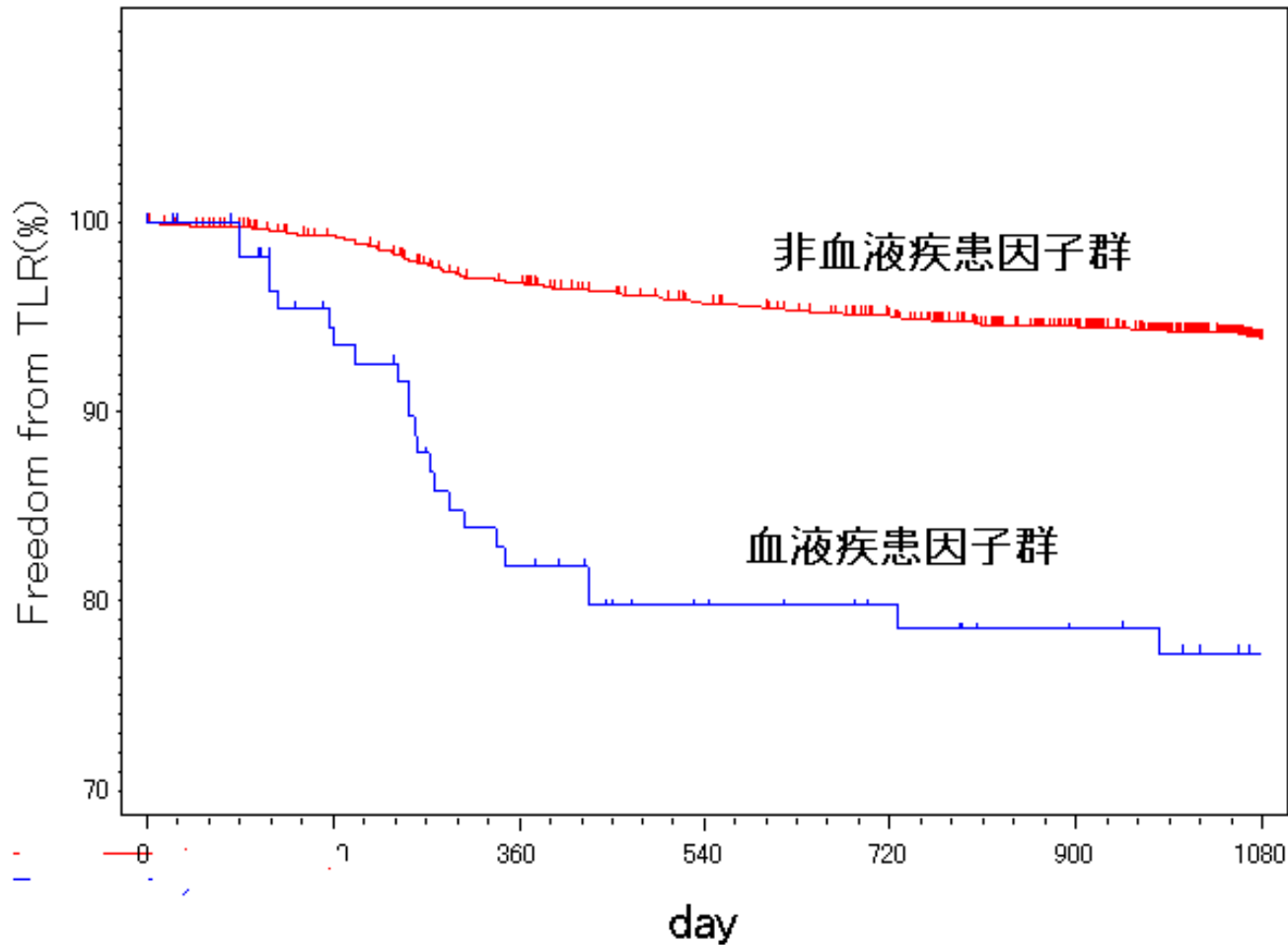
# PS上のイベントの頻度分布 相対度数



# IPTW法とPS層化調整解析

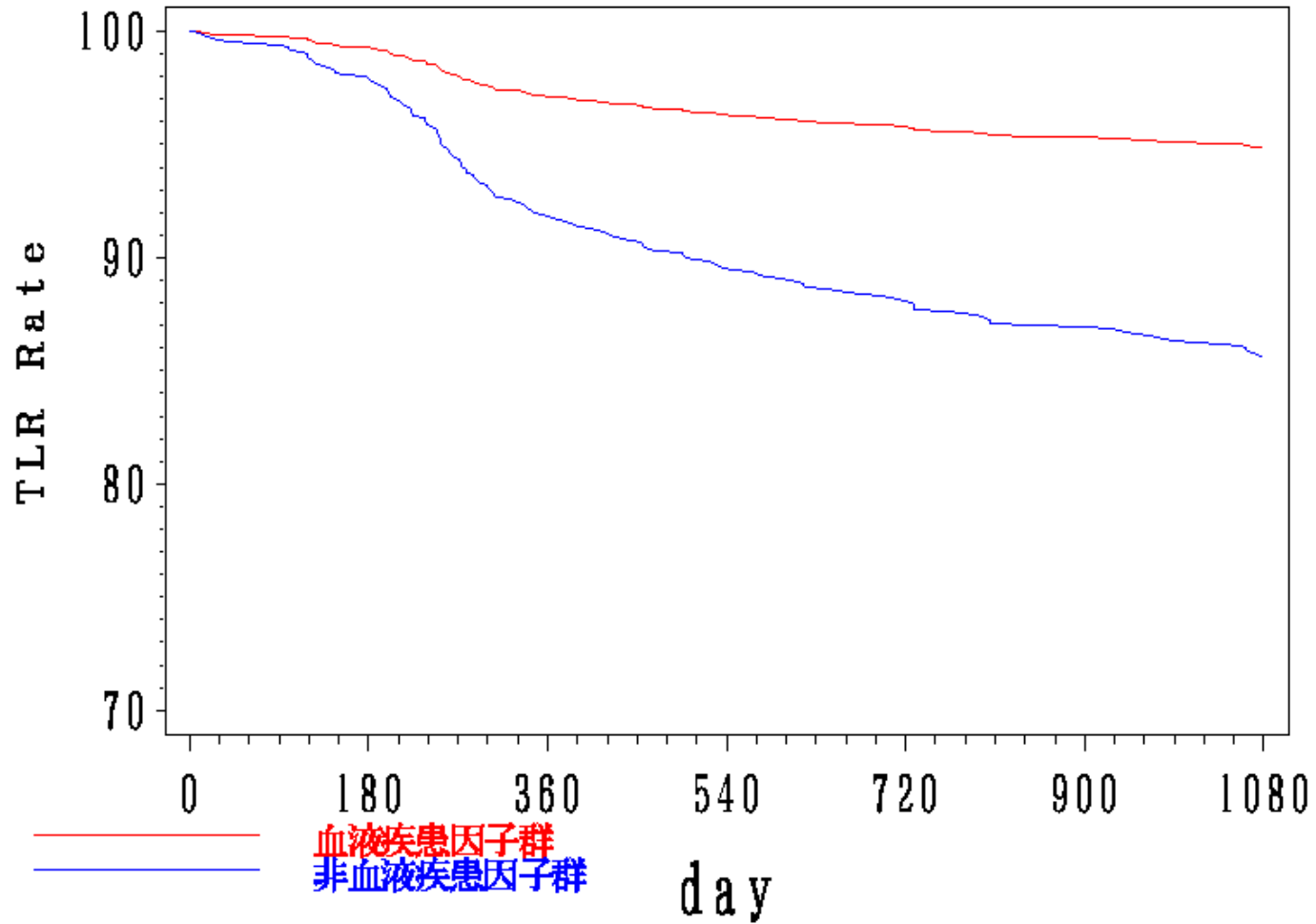
- 一見、前グラフの低PS値において、血管因子あり群は度数がほとんどなく、なし群と分布がほとんど変わらない
- ゆえに、PS層化調整では、1/3のTLRイベントが比較対象からは除外されているのと同じ
- IPTWの方が一見性能が良い？

# 非調整Kaplan-Meier曲線

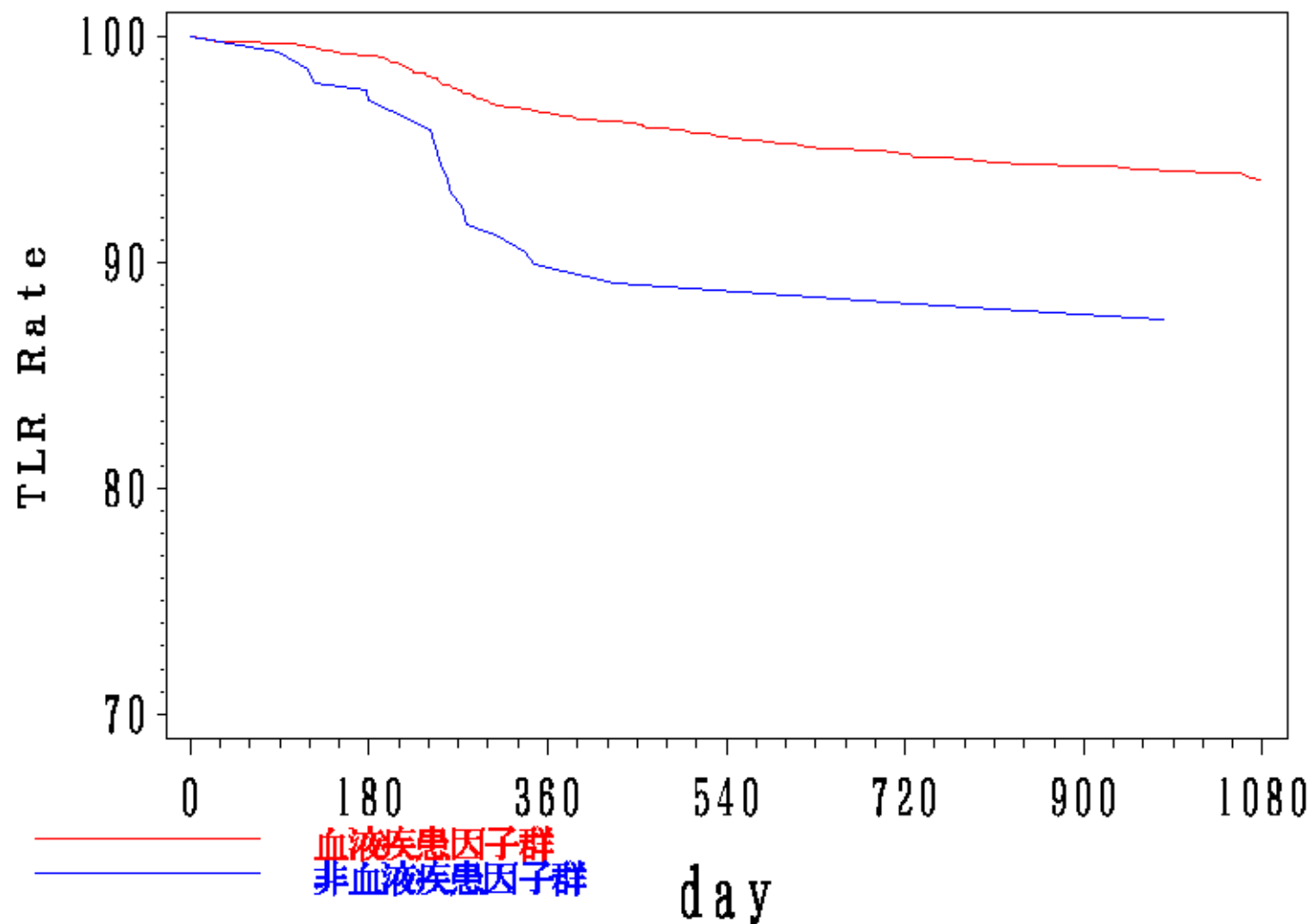




# PS層化調整Kaplan-Meier曲線



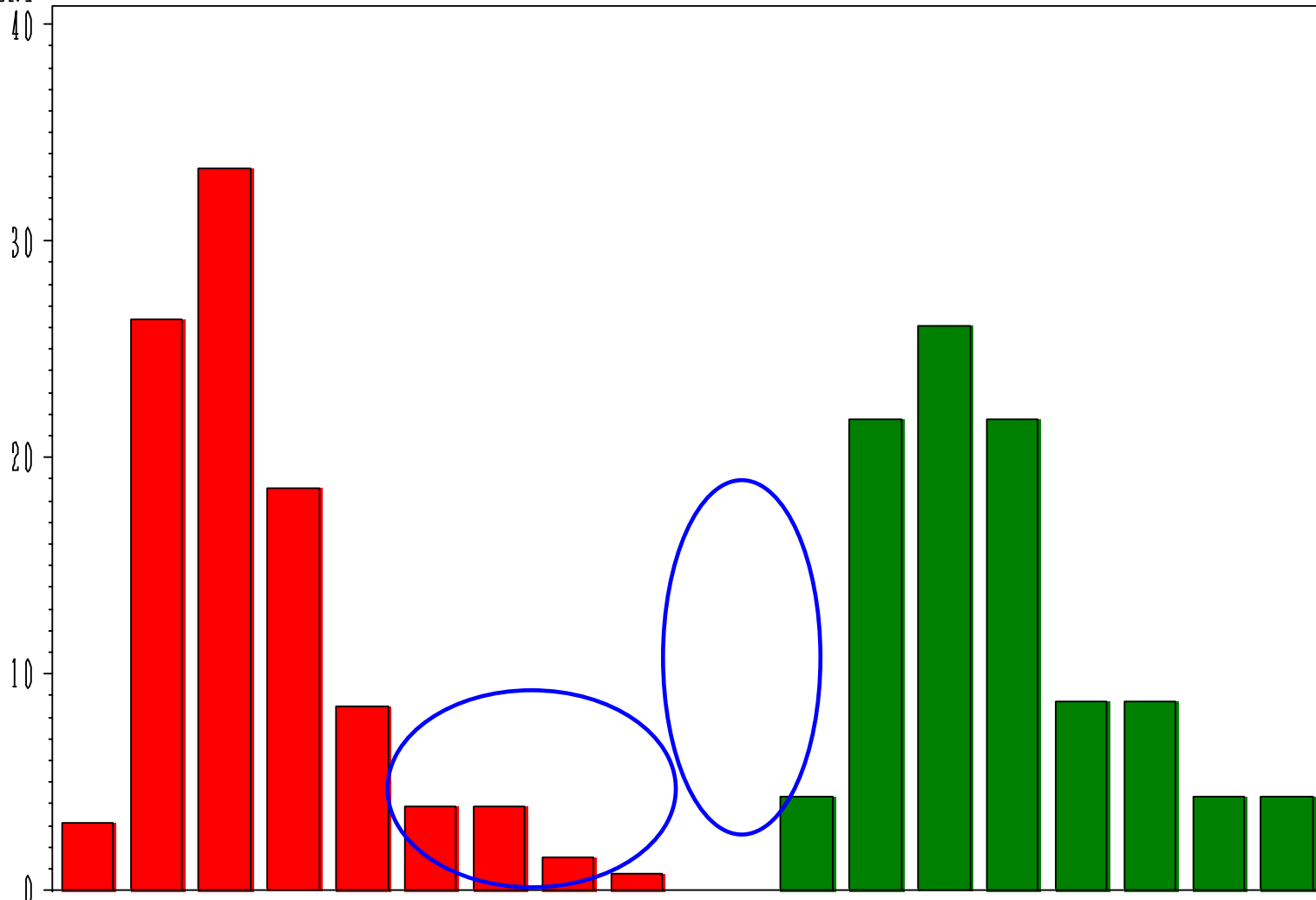
# IPTW 重み付けKaplan-Meier曲線



- Rosenbaum-Rubin 流のstrongly ignorableな仮定 の感度分析
- PSスコアに含まれない、有意な未観察背景項目が存在すると仮定 HR=1.39  
このパラメータ推定値0.329の20%、40%、60%、80%、100%が仮にIPTWのパラメータ推定値に含まれると仮定
- 有意な未観察項目の40%の影響でHRの有意性は消滅
- 感度分析(調整)のHRが減少する傾向、および本感度分析の結果から、この血管因子は本当は有意な項目ではない？
- でも何かが引っかかる

		ハザード比 推定値	95%信頼区間 下限      上限		パラメータ 推定値	標準誤差	
IPTW		2.201	1.126	4.300	0.789	0.342	
未観察 仮定		1.39			0.329		
Bias	20%	2.060	1.055	4.026	0.723	0.066	0.342
Bias	40%	1.929	0.987	3.769	0.657	0.132	0.342
Bias	60%	1.806	0.924	3.529	0.591	0.198	0.342
Bias	80%	1.691	0.865	3.304	0.525	0.263	0.342
Bias	100%	1.583	0.810	3.093	0.459	0.329	0.342

PERCENT



0.00 0.04 0.08 0.12 0.16 0.20 0.24 0.28 0.32 0.00 0.04 0.08 0.12 0.16 0.20 0.24 0.28 0.32 Predicted Value

0 1

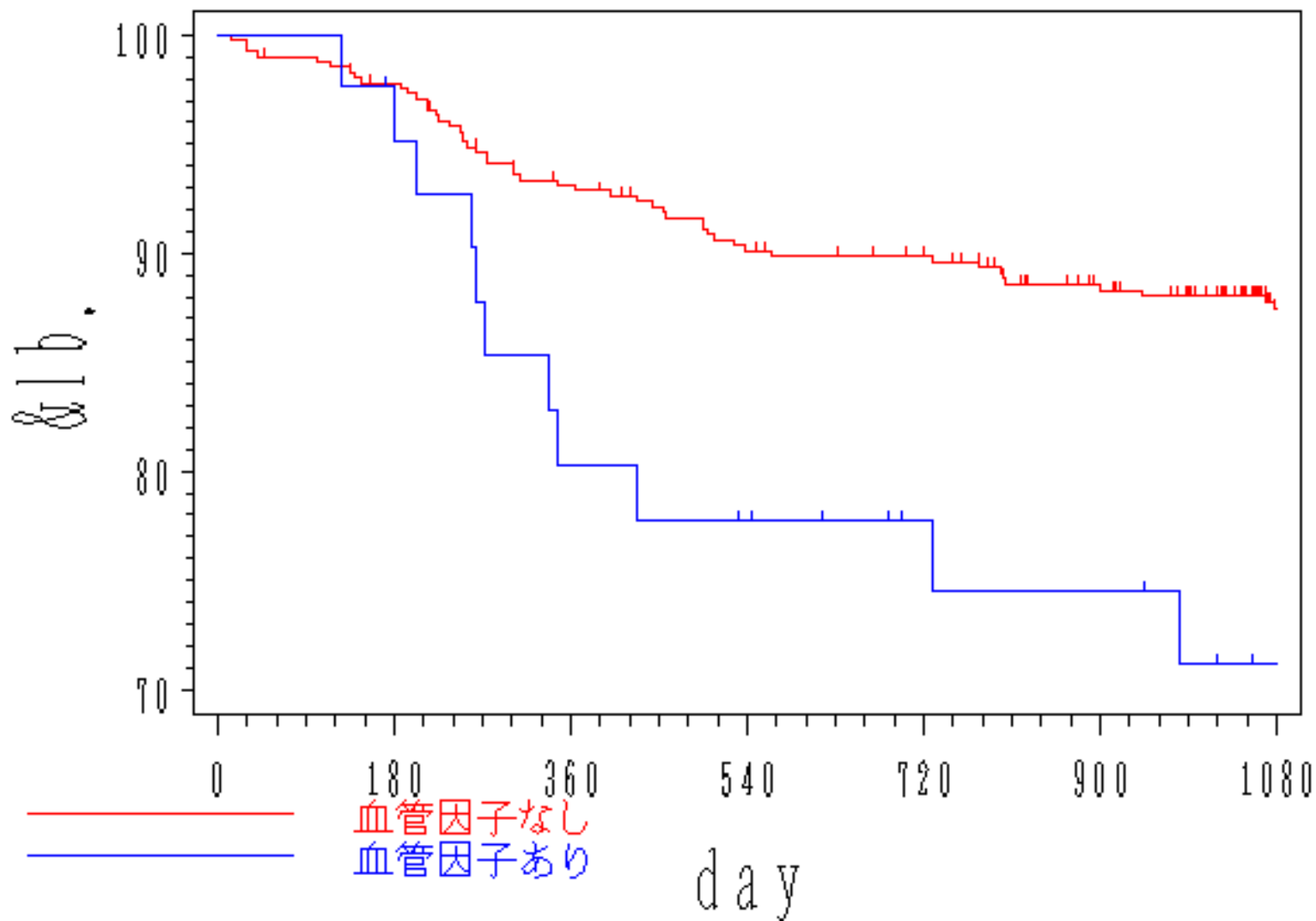
血管影響疾患

血管影響疾患 0 1

# IPTWの問題点

- イベントのPSスコア分布の重心(背景因子の重心)から離れたデータの重みが大きい
- 重心から離れたデータ→集団を代表しない特異なSubgroupの影響が大きい可能性あり

- 平均的(重心付近)のデータによるsubgroup解析
- 



# Conclusion

- IPTWの利点

- Directに統計計量が計算できる。
- 調整解析法のようにPSの治療群、対照群の分布の重なりがあまりなくても計算できる。

- IPTWの問題点

- 被験者背景の重心から離れた地点の重みが大きい
- つまり、集団を代表しない特異的subgroupの影響を受けやすい。

# Conclusion

- IPTWを積極的に採用する場合
  - 臨床試験データのように治療群間の背景に極端な差がなく、集団自体の均一性がある程度保証される、かつ
  - データ数が十分大きい(イベント数で数百を超える)。
  - 理由:スマートだから
- PS層化調整を積極的に採用する場合
  - 上記以外の場合
  - 理由:頑健性が期待できるから
  - 理由2:スマートでない分、制約もはっきりしているから misleadingが少ない



# Conclusion

- IPTWとPS層化調整

- いつも、どちらを採用するかは迷いの種

- 結果を見てからでないと判断できない

- 恣意性

- 解析担当者の能力

- 単変量、多変量、IPTW、PS層化調整すべての結果が一致すれば...

- Subgroup解析を含めた感度分析が重要