

傾向スコア

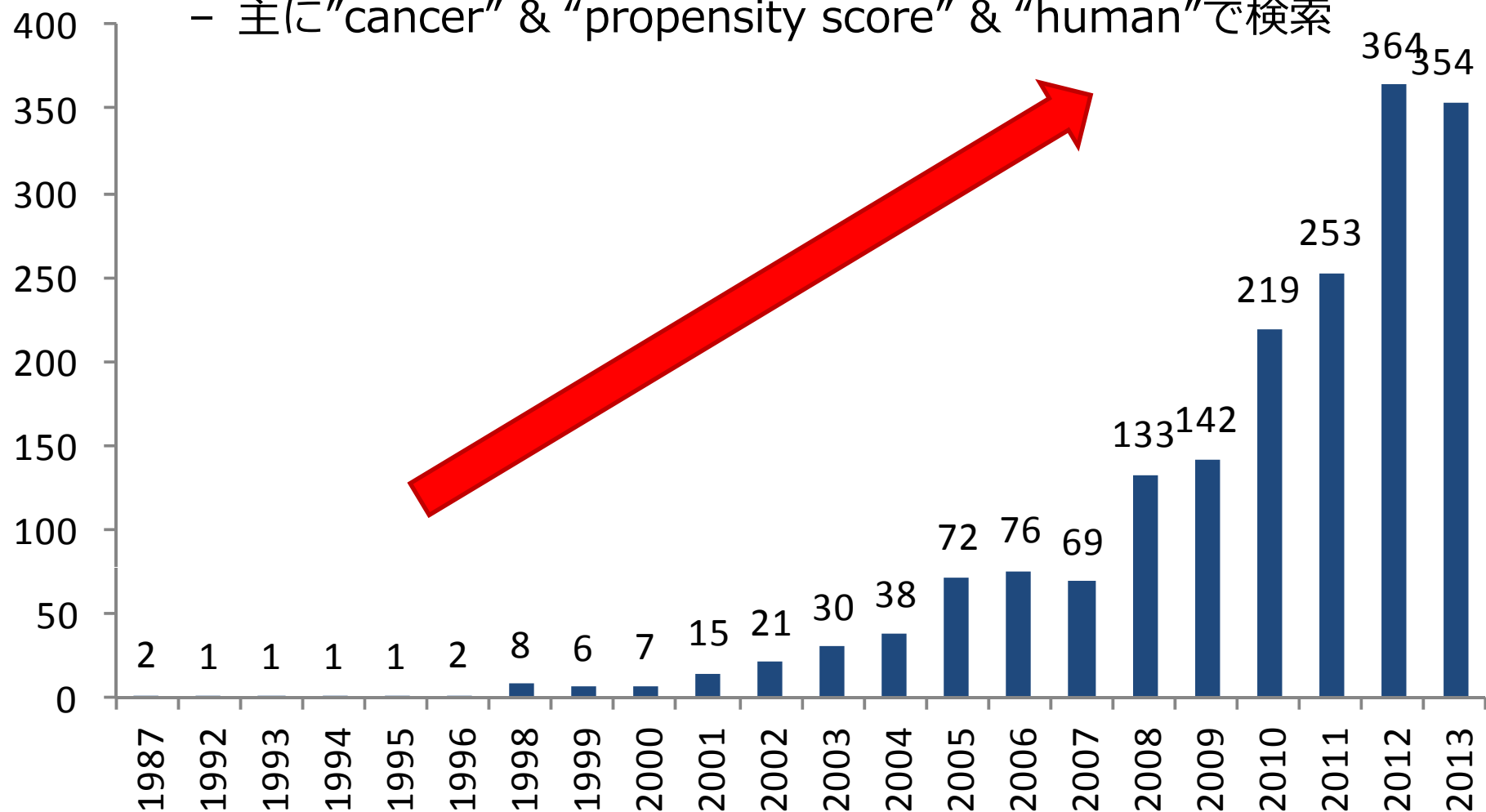
生物統計部門/JCOGデータセンター
水澤 純基

2014. 5.14 (水)

傾向スコアを用いたがん臨床研究の論文数

- PubMed[<http://www.ncbi.nlm.nih.gov/pubmed>]

– 主に“cancer” & “propensity score” & “human”で検索



傾向スコアを用いた臨床研究の実例

- 高齢者の進行(Stage IIIB or IV)非小細胞肺癌 (NSCLC) 患者に対し、標準治療のカルボプラチン+パクリタキセル(CP)にベバシズマブ(Bev)を上乗せすることは有効か？

ORIGINAL CONTRIBUTION

Carboplatin and Paclitaxel With vs Without Bevacizumab in Older Patients With Advanced Non-Small Cell Lung Cancer

Junya Zhu, MS, MA

Dhruv B. Sharma, PhD

Stacy W. Gray, MD, AM

Adleen B. Chen, MD, MPP

June C. Weeks, MD, MS

Deborah Schrag, MD, MPH

Context A previous randomized trial demonstrated that adding bevacizumab to carboplatin and paclitaxel improved survival in advanced non-small cell lung cancer (NSCLC). However, longer survival was not observed in the subgroup of patients aged 65 years or older.

Objective To examine whether adding bevacizumab to carboplatin and paclitaxel chemotherapy is associated with improved survival in older patients with NSCLC.

Design, Setting, and Participants Retrospective cohort study of 4168 Medicare beneficiaries aged 65 years or older with stage IIIB or stage IV non-squamous cell NSCLC diagnosed in 2002-2007 in a Surveillance, Epidemiology, and End Results (SEER) region. Patients were categorized into 3 cohorts based on diagnosis year and type of initial chemotherapy administered within 4 months of diagnosis: (1) diagnosis in 2006-2007 and bevacizumab-carboplatin-paclitaxel therapy; (2) diagnosis in 2006-2007 and carboplatin-paclitaxel therapy; or (3) diagnosis in 2002-2005 and carboplatin-paclitaxel therapy. The associations between carboplatin-paclitaxel with vs without bevacizumab and overall survival were compared using Cox proportional hazards models and propensity score analyses including information about patient characteristics recorded in SEER-Medicare.

Main Outcome Measure Overall survival measured from the first date of chemotherapy treatment until death or the censoring date of December 31, 2009.

Results The median survival estimates were 9.7 (interquartile range [IQR], 4.4-18.6) months for bevacizumab-carboplatin-paclitaxel, 8.9 (IQR, 3.5-19.3) months for carboplatin-paclitaxel in 2006-2007, and 8.0 (IQR, 3.7-17.2) months for carboplatin-paclitaxel in 2002-2005. One-year survival probabilities were 39.6% (95% CI, 34.6%-45.4%) for bevacizumab-carboplatin-paclitaxel vs 40.1% (95% CI, 37.4%-43.0%) for carboplatin-paclitaxel in 2006-2007 and 35.6% (95% CI, 33.8%-37.5%) for carboplatin-paclitaxel in 2002-2005. Neither multivariable nor propensity score-adjusted Cox models demonstrated a survival advantage for bevacizumab-carboplatin-paclitaxel compared with carboplatin-paclitaxel cohorts. In propensity score-stratified models, the hazard ratio for overall survival for bevacizumab-carboplatin-paclitaxel compared with carboplatin-paclitaxel in 2006-2007 was 1.01 (95% CI, 0.89-1.16; $P = .85$) and compared with carboplatin-paclitaxel in 2002-2005 was 0.93 (95% CI, 0.83-1.06; $P = .28$). The propensity score-weighted model and propensity score-matching model similarly failed to demonstrate a statistically significant superiority for bevacizumab-carboplatin-paclitaxel. Subgroup and sensitivity analyses for key variables did not change these findings.

Conclusion Adding bevacizumab to carboplatin and paclitaxel chemotherapy was not associated with better survival among Medicare patients with advanced NSCLC.

JAMA. 2012;307(15):1593-1601

www.jama.com

Author Affiliations: Center for Patient Safety (Dr Zhu), Departments of Biostatistics and Computational Biology (Dr Sharma, Medical Oncology (Dr Gray), and Radiation Oncology (Dr Chen), and Center for Outcomes and Policy Research (Dr Schrag) (Dr Chen, Weeks, and Schrag).

Dana-Farber Cancer Institute, Boston, Massachusetts; Corresponding Author: Deborah Schrag, MD, MPH, Center for Outcomes and Policy Research, Dana-Farber Cancer Institute, 450 Brookline Ave, Boston, MA 02215 (dr_schrag@dfci.harvard.edu).

©2012 American Medical Association. All rights reserved.

JAMA, April 18, 2012—Vol 307, No. 15 1593
Corrected on April 17, 2012

JAMA. 2012;307(15):1593-1601

研究背景（1）

- 進行したNSCLCは予後不良（3年生存割合：約25%）
- 標準治療はCP療法だったが、ECOGによって行われたランダム化比較試験(RCT)によってCPにBevを上乗せすることで全生存期間(OS)が延長することが示され、2006年にFDAに承認された
 - HR 0.79 (95% CI, 0.67-0.92) N Engl J Med. 2006;355(24):2542-2550
- しかし、4つのRCTのメタアナリシスではBevの上乗せによる有意な延長は示されなかった*Clin Drug Investig. 2010;30(4):229-241.*
- さらに、ECOGのRCTでは65歳以上の患者、または70歳以上の患者ではBevの効果がない可能性が示唆された
 - 65歳以上：HR 0.89 (95%CI, 0.70-1.14)
 - 70歳以上：生存期間中央値 11.3 か月vs 12.1か月

研究背景（２）

- FDAの承認に従いMedicare(高齢者のための医療保険) & Medicaidでは保険償還されている
- NSCLCの2/3以上が65歳以上
- Medicareの対象でBevの上乗せがOSの改善に寄与するか確認することは臨床家の意思決定に意味がある
- MedicareとリンクしているSEER programのデータベースを用いて、進行NSCLCの高齢者を対象に、CPにBevを上乗せすることがOSを改善するかどうかを調べる
 - SEER program : 米国NCI主導の地域がん登録

使用するデータの概要

- レトロスペクティブコホート
- 65歳以上のStage IIIB or IV期, NSCLC
- 診断時期2002年-2007年
- N=4168
 - CP:N=3846 (2002-05: N=2664, 2006-07:N=1182)
 - CP+Bev:N=318
- 背景因子（合計11因子、35カテゴリー）
 - 年齢、性別、人種、配偶者の有無、収入、学歴、SEER登録地域、都市部での居住、Modified Charlson comorbidity score、分化度、Stage

Bev使用有無による背景因子の比較（１）

- Bev有に多い
 - 年齢低い
 - 女性
 - 白人
 - 収入多い

Characteristics	Bevacizumab- Carboplatin- Paclitaxel, No. (%) (n = 318)	Carboplatin- Paclitaxel 2006-2007 (n = 1182) ^{b,d}	
		No. (%)	P Value ^e
Age at diagnosis, y			
65-69	127 (39.9)	396 (33.5)	.07
70-74	84 (26.4)	356 (30.1)	
75-79	78 (24.5)	278 (23.5)	
≥80	29 (9.1)	152 (12.9)	
Sex			
Female	156 (49.1)	529 (44.8)	.18
Male	162 (50.9)	653 (55.2)	
Race/ethnicity			
Non-Hispanic white	279 (87.7)	981 (83.0)	.07
Non-Hispanic black	15 (4.7)	98 (8.3)	
Other	24 (7.5)	103 (8.7)	
Marital status			
Not married	117 (36.8)	467 (39.5)	.40
Married	201 (63.2)	715 (60.5)	
Median income (Census tract quintile) ^f			
1 (Lowest)	60 (18.9)	240 (20.3)	.20
2	58 (18.2)	242 (20.5)	
3	63 (19.8)	237 (20.1)	
4	58 (18.2)	242 (20.5)	
5 (Highest)	79 (24.8)	221 (18.7)	

JAMA. 2012;307(15):1593-1601

Bev使用有無による背景因子の比較（2）

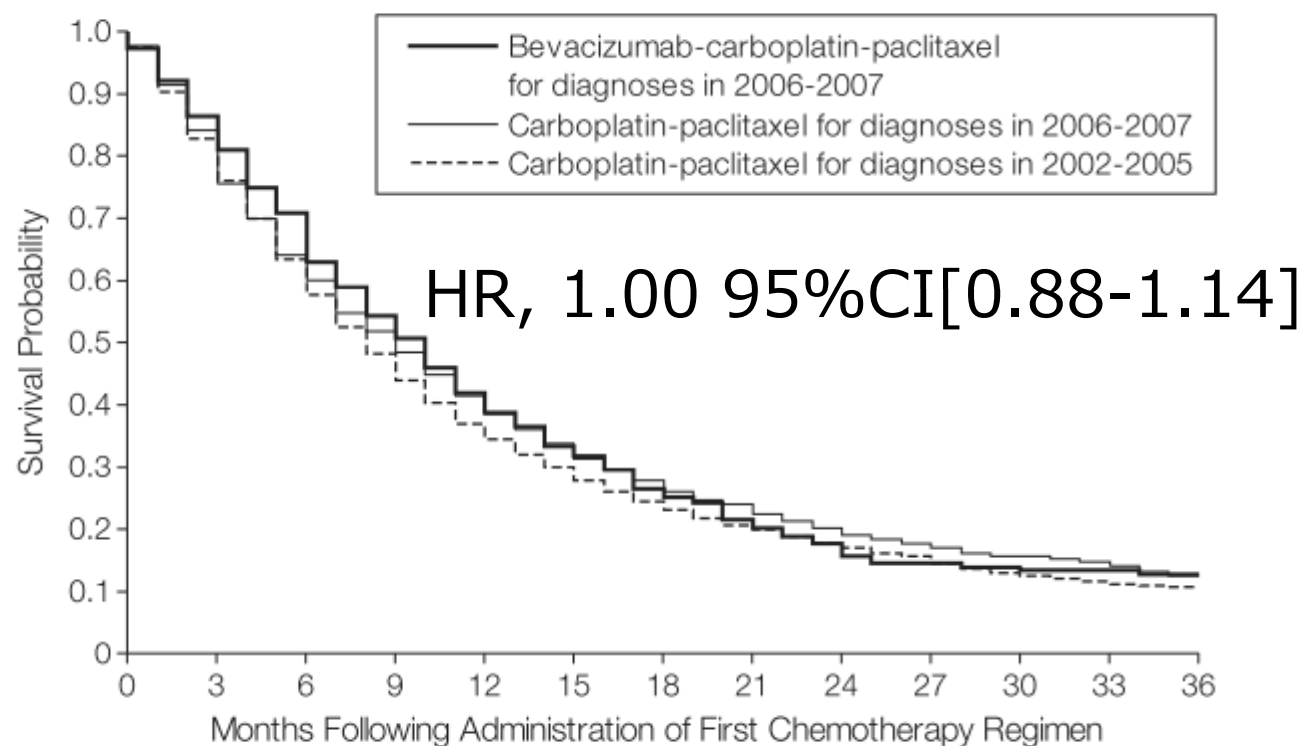
- Bev有に多い
 - 学歴高
 - 居住地西部
 - 都市部でない
 - 併存症・既往歴少
 - 分化度高
 - **Stage IV（進行例）**

Characteristics	Bevacizumab- Carboplatin- Paclitaxel, No. (%) (n = 318)	Carboplatin- Paclitaxel 2006-2007 (n = 1182) ^{b,d}	
		No. (%)	P Value ^e
College educated (Census tract quintile) ^g			
1 (Lowest)	66 (20.8)	234 (19.8)	.38
2	58 (18.2)	244 (20.6)	
3	62 (19.5)	238 (20.1)	
4	57 (17.9)	241 (20.4)	
5 (Highest)	75 (23.6)	225 (19.0)	
SEER region			
Northeast	65 (20.4)	295 (25.0)	.40
South	69 (21.7)	244 (20.6)	
Midwest	51 (16.0)	173 (14.6)	
West	133 (41.8)	470 (39.8)	
Urban residency ^h			
No	35 (11.0)	92 (7.8)	.07
Yes	283 (89.0)	1090 (92.2)	
Modified Charlson comorbidity score ⁱ			
0	210 (66.0)	665 (56.3)	<.001
1	88 (27.7)	324 (27.4)	
≥2	20 (6.3)	193 (16.3)	
Level of differentiation (tumor grade)			
Well/moderately	50 (15.7)	128 (10.8)	.06
Poorly/undifferentiated	85 (26.7)	345 (29.2)	
Unknown	183 (57.5)	709 (60.0)	
AJCC stage			
IIIB	56 (17.6)	344 (29.1)	<.001
IV	262 (82.4)	838 (70.9)	

JAMA. 2012;307(15):1593-1601

単純なBevあり vs なしの生存曲線

Figure 2. Kaplan-Meier Survival Curves for Medicare Beneficiaries Diagnosed as Having Advanced Non-Squamous Cell, Non-Small Cell Lung Cancer, by Year of Diagnosis and First-line Chemotherapy Administration With or Without Bevacizumab

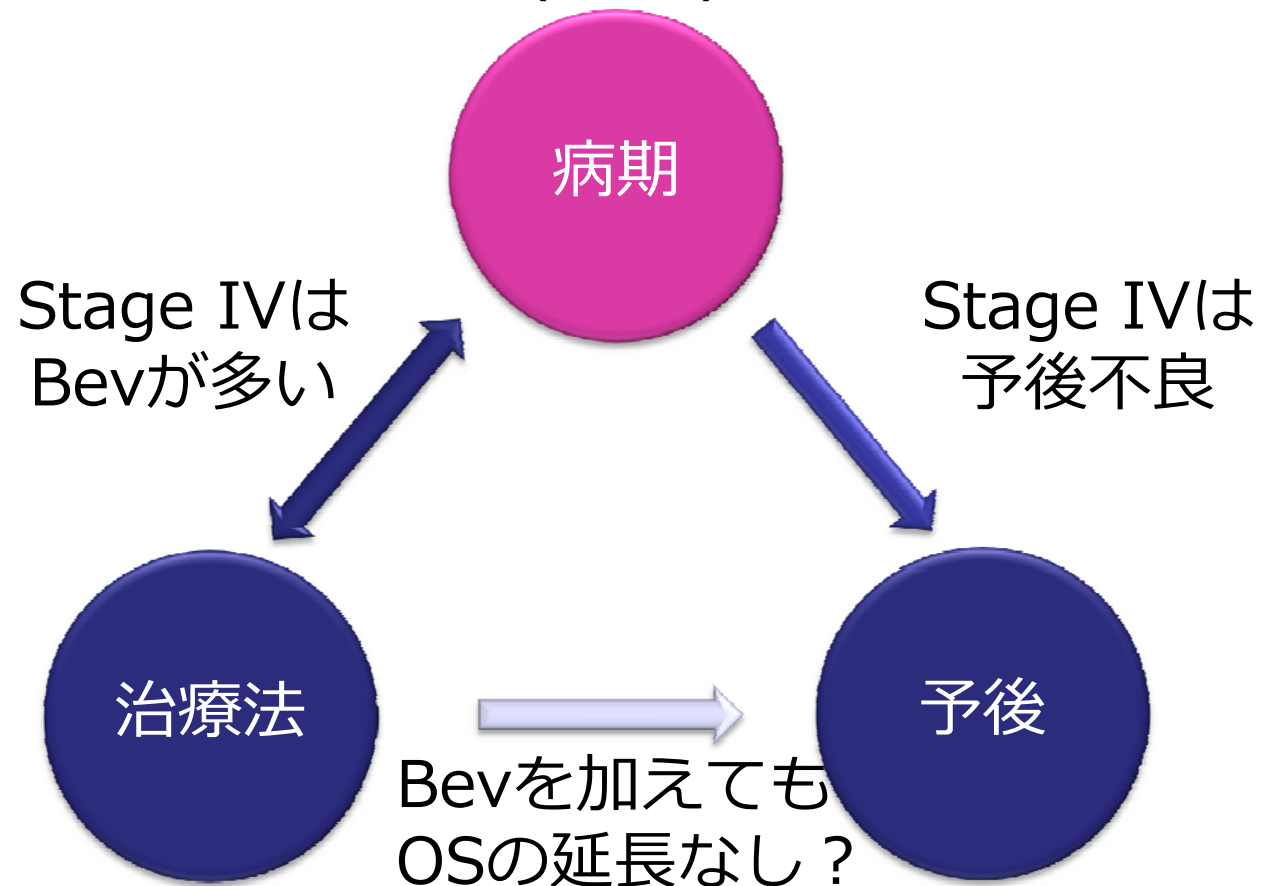


No. at risk							
Bevacizumab-carboplatin-paclitaxel	318	225	133	84	54	31	17
Carboplatin-paclitaxel 2006-2007	1182	756	489	328	228	133	68
Carboplatin-paclitaxel 2002-2005	2664	1688	982	652	472	348	283

どのように妥当な比較を行うか？

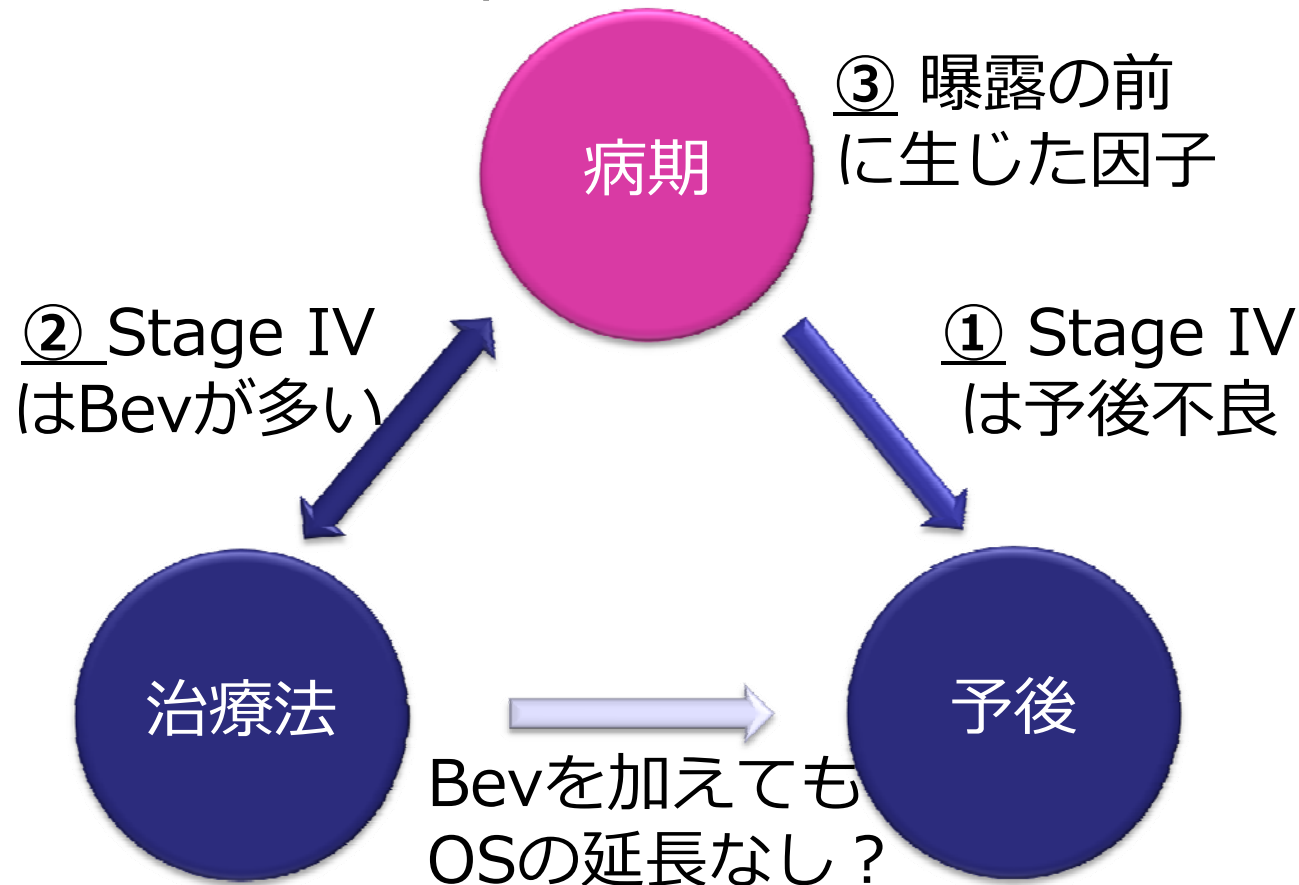
復習：交絡

- 治療法と予後に関連する第3の因子（病期）によって見かけ上の関連が生じてしまう or 消えてしまう現象のこと
 - 交絡を引き起こす因子(=病期)のことを交絡因子という



復習：交絡の必要条件

- ① アウトカム(予後)と関係がある
- ② 曝露(治療法)と関係がある
- ③ 中間変数ではない（治療法の前に判明している事象）



交絡がないことを保証するには

- 治療群間で予後に関係する背景因子を揃える
 - Stage
 - 年齢
 - 併存症・既往歴
 - その他（未知の因子も含めて）

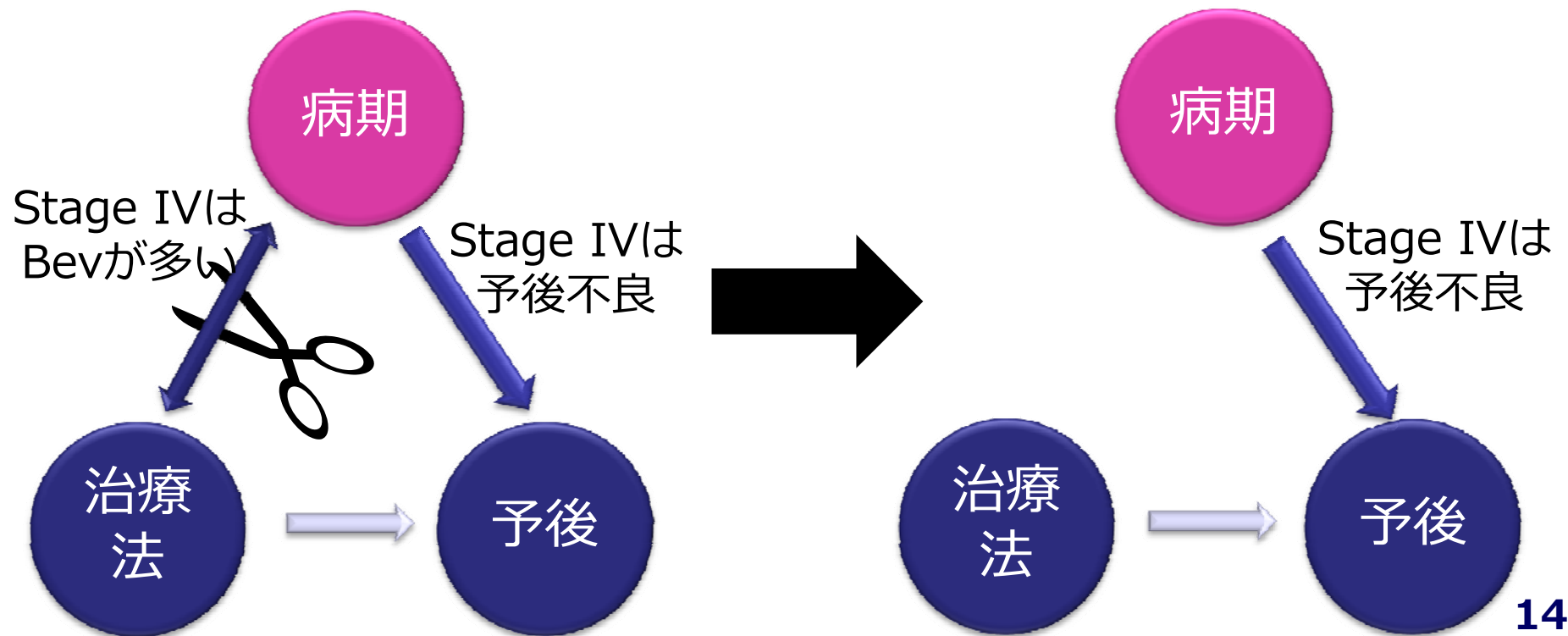
因子がたくさんある・未知の因子があるために
全てを考慮できない



ランダムに決める

ランダム化の意義

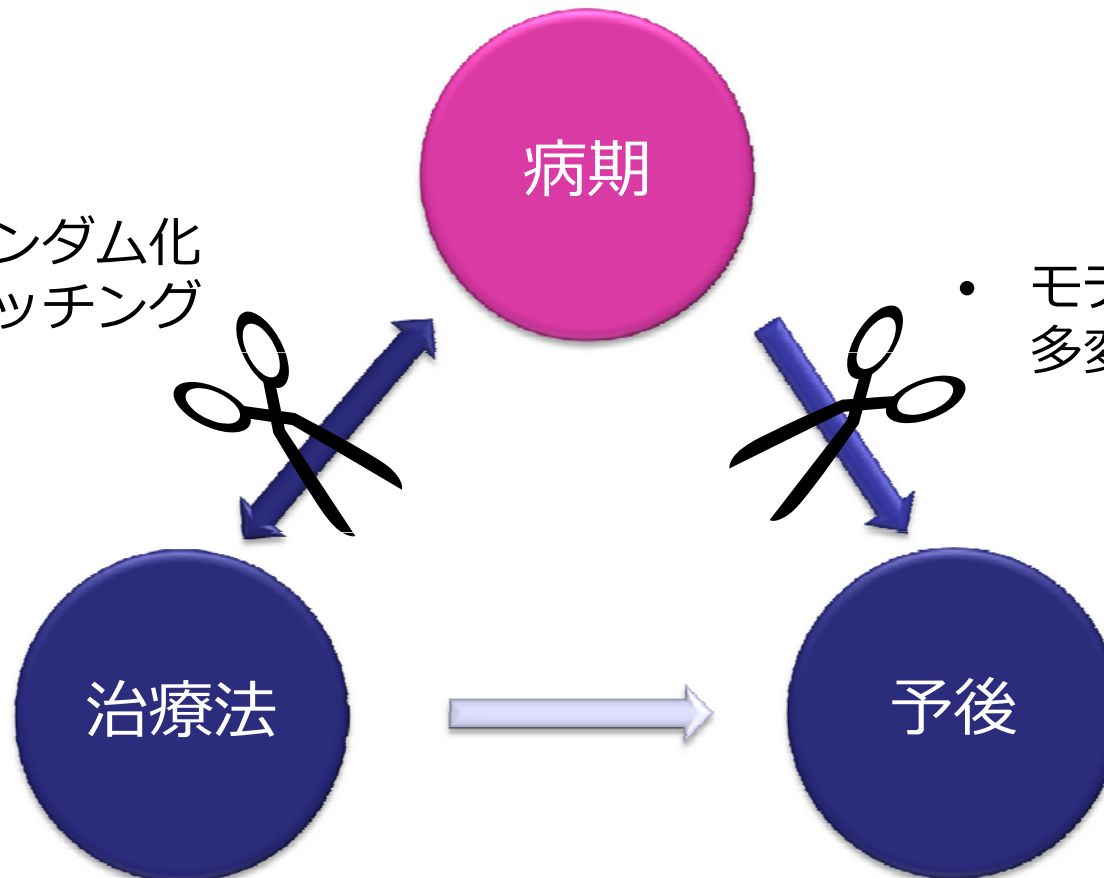
- 治療法と病期の関連を除去できる
 - どの患者も一方の治療法に割り付けられる確率は等しい
 - 1:1割り付けなら、一方の治療法に割り付けられる確率は50%
 - 治療群間で背景因子が等しくなるので、交絡が除去される



交絡を除去するには？

- 少なくともどちらか一方の関係を除去すれば良い
 - 治療法と病期の関係
 - 予後と病期の関係

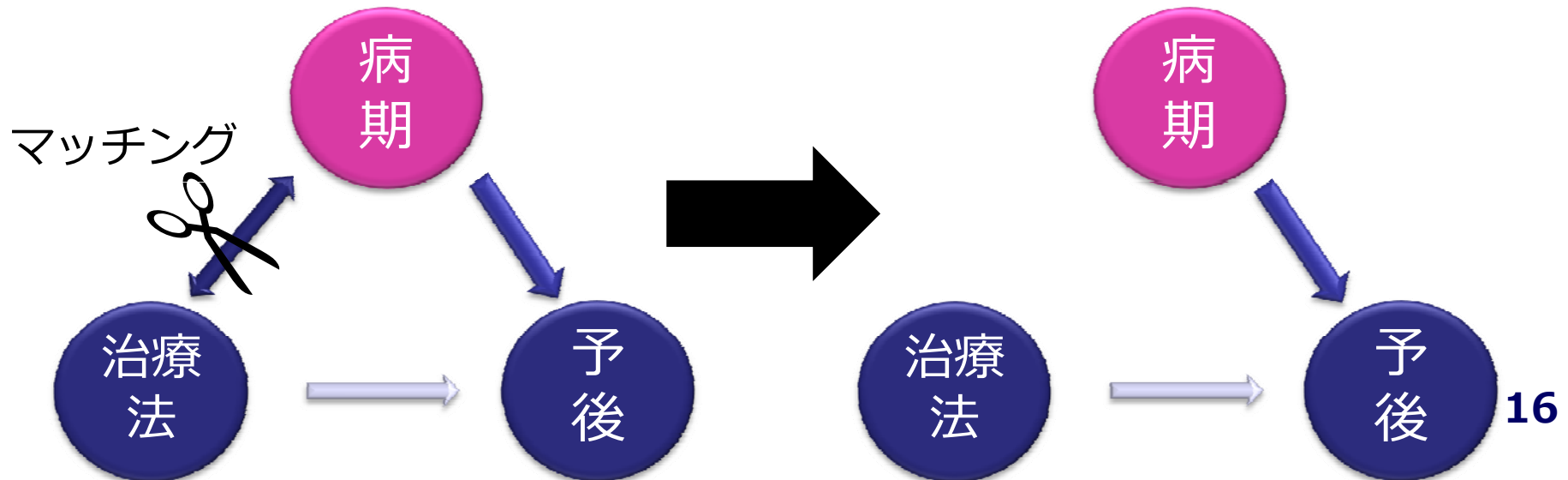
- ランダム化
- マッチング



- モデルを用いた多変量解析

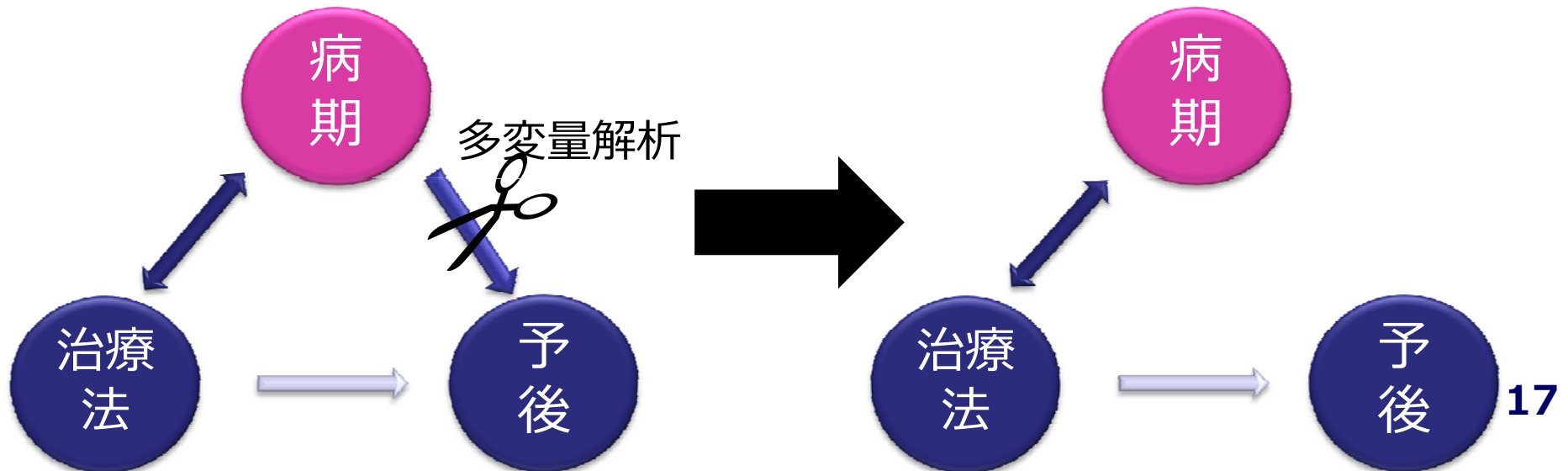
マッチング

- 偏っている変数でマッチングする
 - 年齢、性別、人種 . . .
- 全ての変数でマッチングすることは出来ない
 - 背景が一致する患者がいなくなってしまう
 - 70歳&男性&白人&収入高&学歴低& . . . は、いない
 - 交絡を除去するには、同時に解析での処理も必要



モデルを用いた多変量解析

- ロジスティック回帰やCox回帰の使用
 - 複数の交絡因子を同時に考慮できる
 - モデルが正しければ強力な方法
- オーバーフィッティングの問題
 - データの偶然誤差に過度にフィットしていないか？
 - モデルの妥当性、イベント数は？

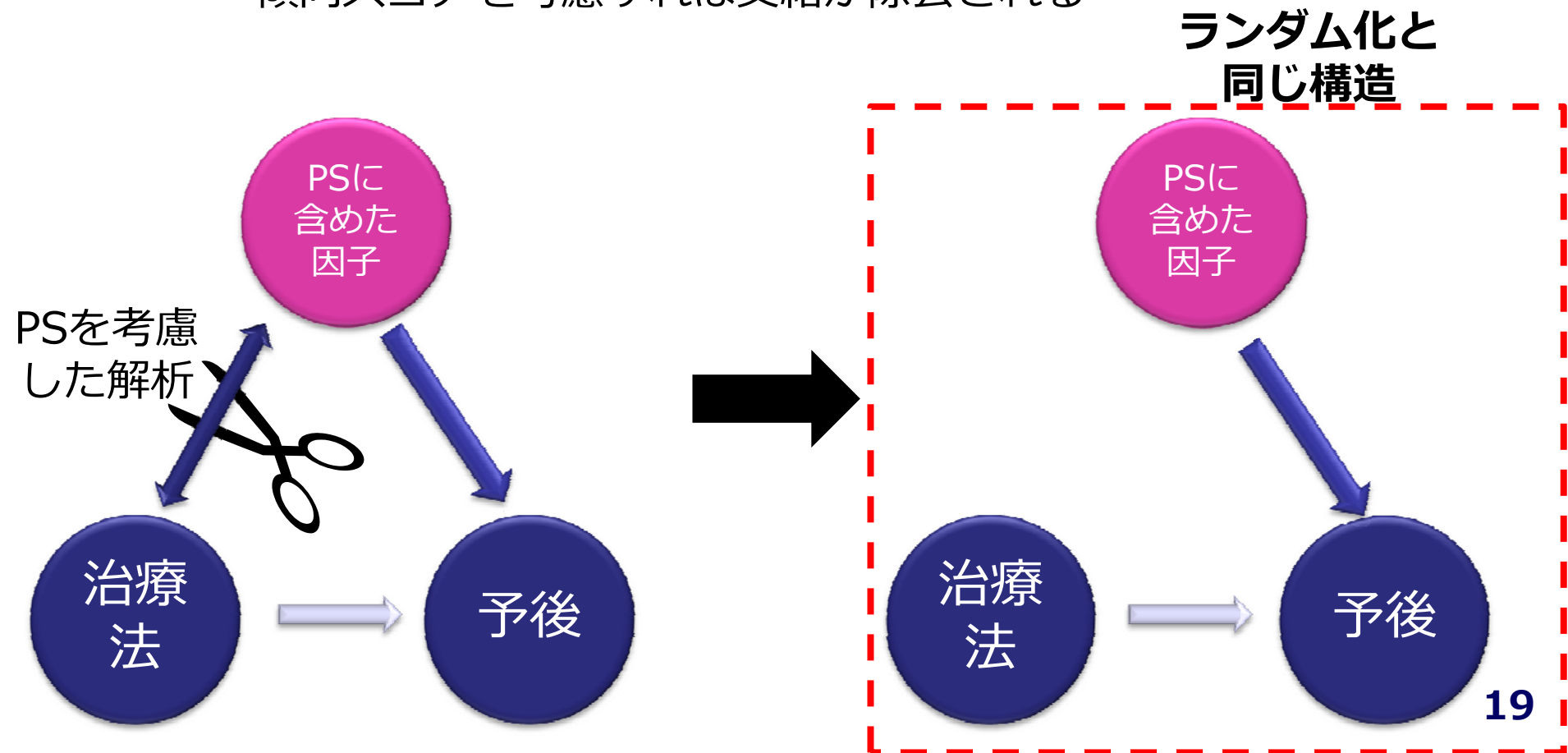


傾向スコア (Propensity Score)

- 傾向スコアとは
 - ある背景因子を持つ対象が、一方の治療法に割り付けられる確率
 - 曝露要約スコア
 - 取り得る値の範囲：0～1
- 治療法の選択に係る複数の因子を要約した1つのスコアにする
 - 1つ1つの背景因子は違っても、傾向スコアが同じであれば、一方の治療法への割り付けられやすさは同じとみなせる
 - 傾向スコアが同じ対象を集団で考えれば、治療群間で背景因子が平均的に同じであることが期待できる
 - ある意味ランダム化と同様の状況が作り出せる

傾向スコアで調整した比較の意義

- 治療法と傾向スコア[PS]（≡背景因子）の関連を除去できる
 - 交絡を除去する構造がランダム化と同じになる
 - 傾向スコアを考慮すれば交絡が除去される



傾向スコアを用いた解析手順

傾向スコアを用いた解析法

- 2段階の手順で行う

① 傾向スコアの推定

- － 患者ごとの傾向スコアを推定する

② 傾向スコアを用いて調整した解析

- － マッチング
- － 層別解析
- － 多変量解析
- － IPTW

①患者ごとの傾向スコア(PS)の推定

- 傾向スコア (PS:Propensity score)はBevを受ける確率で、 $0 < PS < 1$ の値を取る
 - $\doteq 0$: Bevを受ける確率はほとんどない
 - $\doteq 1$: ほぼ間違いなくBevを受ける

- 通常はロジスティック回帰モデルを用いて推定する

- Bevをうけていれば1、受けていなければ0として、

$$\log\left(\frac{PS}{1-PS}\right) = \beta_1 \cdot \text{年齢} + \beta_2 \cdot \text{stage} + \beta_3 \cdot \text{性別} + \dots + \beta_i \cdot \text{併存症}$$



$$PS = \frac{\exp(\beta_1 \cdot \text{年齢} + \beta_2 \cdot \text{stage} + \beta_3 \cdot \text{性別} + \dots + \beta_i \cdot \text{併存症})}{1 + \exp(\beta_1 \cdot \text{年齢} + \beta_2 \cdot \text{stage} + \beta_3 \cdot \text{性別} + \dots + \beta_i \cdot \text{併存症})}$$

- 傾向スコアは「予後(アウトカム)」の情報を用いずに推定可

論文における傾向スコア推定法の統計記載

ロジスティック回帰モデルを使って推定した



Multivariable logistic regression was used to predict treatment (bevacizumab-carboplatin paclitaxel compared with carboplatin paclitaxel) based on confounding covariates, including **age, sex, race/ ethnicity, marital status, geographic region, urban residency, tumor grading, Census tract education, median income, modified Charlson comorbidities, and AJCC stage.**



ロジスティック回帰モデルに含めた変数は11変数

傾向スコア推定時に考えるべきこと

1. 傾向スコア推定のための変数の選び方は？
2. 推定された傾向スコアが妥当かどうかを確認するには？

1.傾向スコア推定のための変数の選び方(1)

- 基本原則

- 含める変数は“治療(介入)前”の因子

- 傾向スコアは「一方の治療に割り付けられる確率」なので、治療選択の決定前の因子を用いる
 - 中間変数（治療開始後にわかる因子）は含めない

- 例1：Stage IVのNSCLCの患者にBevをするかを推定するのに、治療開始後の全身状態や臨床検査値は用いてはいけない

- 例2：手術可能早期胃癌の患者に腹腔鏡下術をするかを推定するのに、手術中の出血量や手術後に得られる病理などの情報は用いてはいけない

1.傾向スコア推定のための変数の選び方(2)

- どのように介入前の変数を選択するかは統計家によって意見が分かれる
 - A) 治療法を選択に関係する（治療群間で差がある）因子を含める
 - B) 予後との関連を確認して、予後に関係する因子を全て含める
 - C) AとBに共通する因子(実際に交絡を引き起こす因子)だけを全て含める
- "A"を用いるのが良い
 - BやCでは、傾向スコアのメリット「アウトカムのデータを用いずに推定できる」が失われる *Journal of biopharmaceutical statistics* 22.6 (2012): 1272-9.
 - Aであれば、プロトコールに規定した範囲で傾向スコア推定のやり直しが可能
 - 交互作用項を入れる、変数変換など
 - 統計的有意差よりも過去の報告や経験などを重視すべき
 - ただし、統計的に検出力を最も高くするためにはCのみ入れるのが良い

傾向スコア推定時に考えるべきこと

1. 傾向スコア推定のための変数の選び方は？

2. 推定された傾向スコアが妥当かどうかを確認するには？

2. 推定された傾向スコアが妥当か確認

- 個々の背景因子の群間比較

- 例えば、傾向スコアが同じ患者でマッチした後の集団で、背景因子を群間比較する
 - 標準化したものを用いて比較することもある
- 著しく異なる因子があれば、傾向スコアを算出し直したり、解析時に調整することを検討する

個々の背景因子別に確認した例

- 早期胃がんに対して開腹切除術 vs 腹腔鏡切除術

Characteristics	All Patients					Propensity-Matched Patients				
	Laparoscopic Gastrectomy (n = 3937)		Open Gastrectomy (n = 5451)		P	Laparoscopic Gastrectomy (n = 2473)		Open Gastrectomy (n = 2473)		P
	n	%	n	%		n	%	n	%	
Sex (male)	2507	63.7	3689	67.7	<0.001	1691	68.4	1666	67.4	0.447
Age, yr										
≤59	837	21.3	771	14.1	<0.001	364	14.7	386	15.6	0.981
60–69	1242	31.5	1485	27.2		794	32.1	729	29.5	
70–79	1320	33.5	1978	36.3		904	36.6	944	38.2	
≥80	538	13.7	1217	22.3		411	16.6	414	16.7	
Cancer stage										
I	3650	92.7	3691	67.7	<0.001	2188	88.5	2186	88.4	0.929
II	287	7.3	1760	32.3		285	11.5	287	11.6	
Charlson comorbidity index										
≤2	2273	57.7	2790	51.2	<0.001	1381	55.8	1437	58.1	0.237
3	1187	30.1	1573	28.9		754	30.5	704	28.5	
≥4	477	12.1	1088	20.0		338	13.7	332	13.4	
Smoking index, pack years										
0	2109	53.6	2850	52.3	<0.001	1303	52.7	1326	53.6	0.807
1–49	1095	27.8	1392	25.5		702	28.4	688	27.8	
≥50	733	18.6	1204	22.1		468	18.9	459	18.6	
Body mass index, kg/m ²										
< 18.5	364	9.2	670	12.3	<0.001	243	9.8	246	9.9	0.901
18.5–25	2706	68.7	3557	65.3		1675	67.7	1676	67.8	
25–30	738	18.7	963	17.7		507	20.5	810	32.8	
≥30	70	1.8	121	2.2		48	1.9	41	1.7	
Type of hospital										
Teaching	1313	33.4	851	15.6	<0.001	159	6.4	163	6.6	0.818
Nonteaching	2624	66.6	4600	84.4		2314	93.6	2310	93.4	
Hospital volume, per yr										
Low (≤25)	998	25.3	2166	39.7	<0.001	913	36.9	923	37.3	0.687
Medium (26–46)	1348	34.2	1700	31.2		827	33.4	830	33.6	
High (≥47)	1591	40.4	1585	29.1		733	29.6	720	29.1	

傾向スコアを用いた解析法

- 2段階の手順で行う

① 傾向スコアの推定

- － 患者ごとの傾向スコアを推定する

② 傾向スコアを用いて調整した解析

- － マッチング
- － 層別解析
- － 多変量解析
- － IPTW

論文における解析法の統計記載（1）

- Cox proportional hazards models were conducted separately within each stratum to compare overall survival of patients receiving vs not receiving bevacizumab, and then the 5 HRs estimated from each stratum were combined into an overall HR for the whole cohort.
 - 層別解析

論文における解析法の統計記載（2）

- Cox models were also performed by applying propensity scores to adjust for group differences in 3 alternative ways:
- (1) regression adjustment (ie, inclusion of the propensity score as a linear predictor in the model);
 - モデルを用いた解析（多変量解析）
- (2) propensity score matching, which paired bevacizumabcarboplatin- paclitaxel and carboplatinpaclitaxel patients who were similar in terms of their measurable characteristics;
 - マッチング
- (3) use of the propensity score to create stabilized weights, defined as the inverse probability of treatment weighting.
 - 治療の逆確率重み付け解析（IPTW）

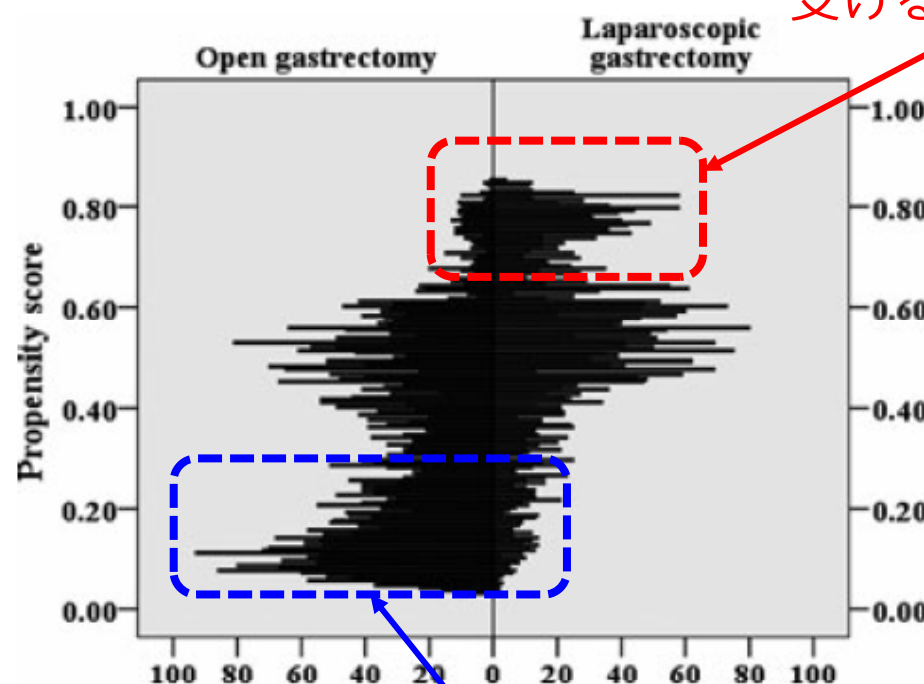
マッチングによる解析

- 各治療群から傾向スコアがほぼ同じ対象をペアにする
 - Untreated(Bevなし)の人数がTreated(Bevあり)よりもかなり多く必要
- 利点
 - わかりやすい
 - 他の傾向スコアを用いた解析法と比較してバイアスが入りにくい
- 欠点
 - マッチできない対象は解析から除外される
 - 一般化可能性の欠如
 - マッチングの恣意性
 - どのようなマッチングの方法を用いるかによって結果が変わる

マッチングの例

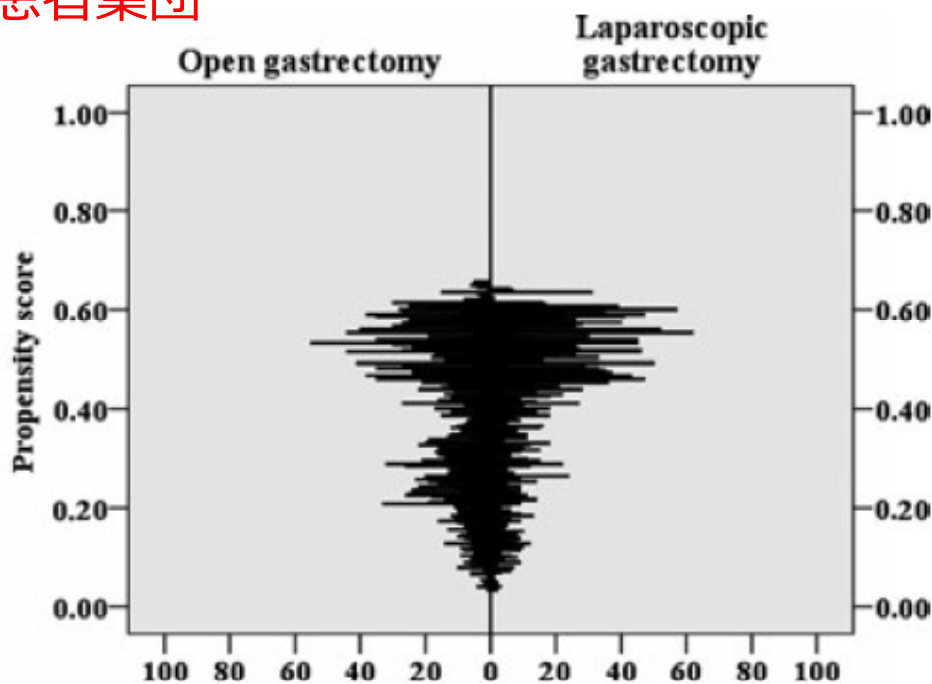
- 早期胃がんに対して開腹切除術(OP) vs 腹腔鏡切除術(LAP)

全患者を対象にした
傾向スコアの分布



ほぼ間違いなくLAPを
受ける患者集団

マッチング後の
傾向スコアの分布



ほぼ間違いなくOPを受ける患者集団

マッチングによる結果

Table 2. Effect of Bevacizumab Added to Carboplatin and Paclitaxel Chemotherapy on Hazard Ratios for Overall Survival

Models	Sample Sizes, No.		Hazard Ratio (95% CI)	
	Bevacizumab-Carboplatin-Paclitaxel vs Carboplatin-Paclitaxel 2006-2007	Bevacizumab-Carboplatin-Paclitaxel vs Carboplatin-Paclitaxel 2002-2005	Bevacizumab-Carboplatin-Paclitaxel vs Carboplatin-Paclitaxel 2006-2007	Bevacizumab-Carboplatin-Paclitaxel vs Carboplatin-Paclitaxel 2002-2005
Unadjusted model	318 and 1182	318 and 2664	1.00 (0.88-1.14)	0.95 (0.84-1.08)
Multivariable-adjusted model ^a	318 and 1182	318 and 2664	1.01 (0.88-1.15)	0.94 (0.83-1.06)
Propensity score-adjusted model ^b				
Stratification	318 and 1153	311 and 2512	1.01 (0.89-1.16)	0.93 (0.83-1.06)
Within-propensity score quintile				
1 (Lowest propensity)	33 and 261	51 and 716	0.75 (0.51-1.10)	0.97 (0.72-1.30)
2	69 and 337	84 and 795	1.06 (0.80-1.39)	0.92 (0.73-1.17)
3	110 and 361	107 and 640	1.06 (0.85-1.32)	0.97 (0.79-1.20)
4	78 and 151	52 and 299	1.16 (0.87-1.55)	0.81 (0.59-1.11)
5 (Highest propensity)	28 and 43	17 and 62	0.84 (0.48-1.47)	1.07 (0.61-1.91)
Regression adjustment	318 and 1153	311 and 2512	1.01 (0.89-1.16)	0.94 (0.83-1.06)
Weighting (stabilized IPTW)	318 and 1153	311 and 2512	0.99 (0.87-1.13)	0.93 (0.82-1.06)
Matching 1:1	318 and 318	318 and 318	0.99 (0.79-1.23)	0.90 (0.72-1.13)

- HR, 0.99 95%CI[0.79-1.23]

マッチングした集団に対する解析方法

- マッチしたことを考慮した解析が必須

- 平均値の比較 $\Rightarrow$ 対応のあるt検定
- 割合の比較 $\Rightarrow$ McNemar検定 & 条件付きlogistic回帰
- 生存時間の比較 $\Rightarrow$ 層別log-rank検定 & 層別Cox回帰 with $\square$ バスト分散

- マッチしたことを考慮した解析法を用いないと、過度に保守的になる（有意になりにくい）

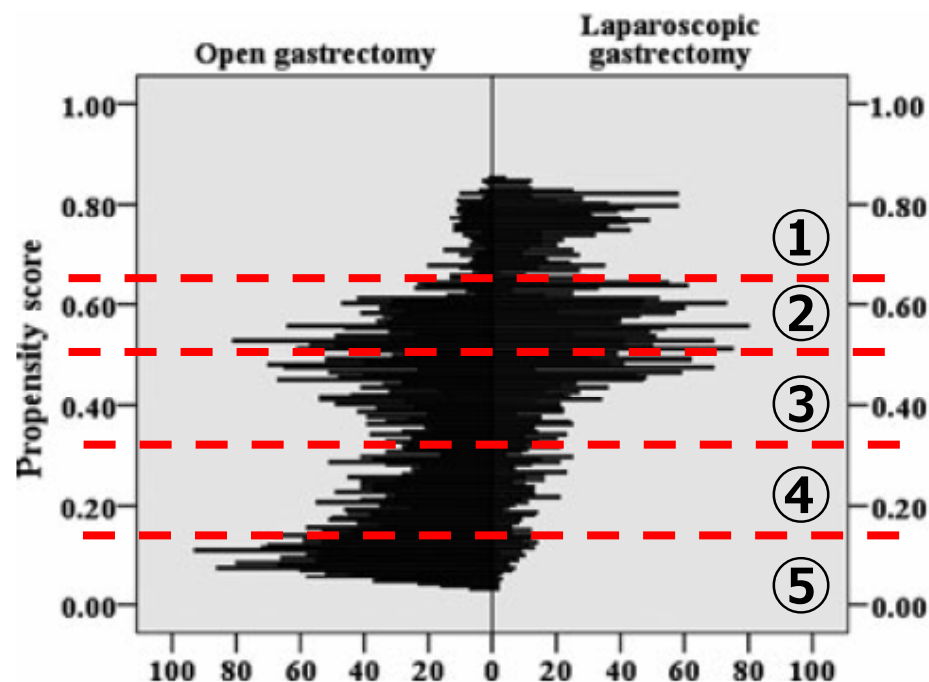
- 通常の解析法は各患者が独立であることが前提
- マッチングされたペアは独立とは言えず、似たものどうし
- 似たものどうしを考慮した方がバラツキが小さくなり、有意になりやすくなる

層別解析

- 傾向スコアを層に分けて、層ごとに治療効果を算出し、重み付き平均して統合する
 - 傾向スコアの分位点で、0-20%、20-40%、40-60%、60-80%、80-100%と分けることが多い
- 利点
 - 全対象を解析に含めることができる
- 欠点
 - 層の中で交絡していることがあり得る
 - どのように層別するかについての恣意性がある

層別解析の仮想例

- 早期胃がんに対して開腹切除術(OP) vs 腹腔鏡切除術(LAP)
- ①～⑤それぞれで治療効果を求めて、それらを統合(重み付き平均)して全体の結果を出す
 - ① $HR=0.7$
 - ② $HR=0.65$
 - ③ $HR=0.75$
 - ④ $HR=0.68$
 - ⑤ $HR=0.72$
 - $\Rightarrow$ 統合 $HR=0.7$



層別解析の結果

Table 2. Effect of Bevacizumab Added to Carboplatin and Paclitaxel Chemotherapy on Hazard Ratios for Overall Survival

Models	Sample Sizes, No.		Hazard Ratio (95% CI)	
	Bevacizumab-Carboplatin-Paclitaxel vs Carboplatin-Paclitaxel 2006-2007	Bevacizumab-Carboplatin-Paclitaxel vs Carboplatin-Paclitaxel 2002-2005	Bevacizumab-Carboplatin-Paclitaxel vs Carboplatin-Paclitaxel 2006-2007	Bevacizumab-Carboplatin-Paclitaxel vs Carboplatin-Paclitaxel 2002-2005
Unadjusted model	318 and 1182	318 and 2664	1.00 (0.88-1.14)	0.95 (0.84-1.08)
Multivariable-adjusted model ^a	318 and 1182	318 and 2664	1.01 (0.88-1.15)	0.94 (0.83-1.06)
Propensity score-adjusted model ^b				
Stratification	318 and 1153	311 and 2512	1.01 (0.89-1.16)	0.93 (0.83-1.06)
Within-propensity score quintile				
1 (Lowest propensity)	33 and 261	51 and 716	0.75 (0.51-1.10)	0.97 (0.72-1.30)
2	69 and 337	84 and 795	1.06 (0.80-1.39)	0.92 (0.73-1.17)
3	110 and 361	107 and 640	1.06 (0.85-1.32)	0.97 (0.79-1.20)
4	78 and 151	52 and 299	1.16 (0.87-1.55)	0.81 (0.59-1.11)
5 (Highest propensity)	28 and 43	17 and 62	0.84 (0.48-1.47)	1.07 (0.61-1.91)
Regression adjustment	318 and 1153	311 and 2512	1.01 (0.89-1.16)	0.94 (0.83-1.06)
Weighting (stabilized IPTW)	318 and 1153	311 and 2512	0.99 (0.87-1.13)	0.93 (0.82-1.06)
Matching 1:1	318 and 318	318 and 318	0.99 (0.79-1.23)	0.90 (0.72-1.13)

- HR, 1.01 95%CI[0.89-1.16]

多変量解析

- モデル(Cox回帰モデルやロジスティック回帰モデル)に、治療法と傾向スコアを共変量として入れた解析
- モデルを用いた解析ならではの仮定が必要
 - 傾向スコアが $0 \rightarrow 0.1$ に増加するのと、 $0.2 \rightarrow 0.3$ に増加するのは同じ効果という仮定
 - 治療法と傾向スコアの関係は足し算であるという仮定
 - 線形（二乗だったりしない）であるという仮定
- 利点
 - 全対象を解析に含めることができる
- 欠点
 - Cox回帰やロジスティック回帰モデルの妥当性の問題

多変量解析の結果

Table 2. Effect of Bevacizumab Added to Carboplatin and Paclitaxel Chemotherapy on Hazard Ratios for Overall Survival

Models	Sample Sizes, No.		Hazard Ratio (95% CI)	
	Bevacizumab-Carboplatin-Paclitaxel vs Carboplatin-Paclitaxel 2006-2007	Bevacizumab-Carboplatin-Paclitaxel vs Carboplatin-Paclitaxel 2002-2005	Bevacizumab-Carboplatin-Paclitaxel vs Carboplatin-Paclitaxel 2006-2007	Bevacizumab-Carboplatin-Paclitaxel vs Carboplatin-Paclitaxel 2002-2005
Unadjusted model	318 and 1182	318 and 2664	1.00 (0.88-1.14)	0.95 (0.84-1.08)
Multivariable-adjusted model ^a	318 and 1182	318 and 2664	1.01 (0.88-1.15)	0.94 (0.83-1.06)
Propensity score-adjusted model ^b				
Stratification	318 and 1153	311 and 2512	1.01 (0.89-1.16)	0.93 (0.83-1.06)
Within-propensity score quintile				
1 (Lowest propensity)	33 and 261	51 and 716	0.75 (0.51-1.10)	0.97 (0.72-1.30)
2	69 and 337	84 and 795	1.06 (0.80-1.39)	0.92 (0.73-1.17)
3	110 and 361	107 and 640	1.06 (0.85-1.32)	0.97 (0.79-1.20)
4	78 and 151	52 and 299	1.16 (0.87-1.55)	0.81 (0.59-1.11)
5 (Highest propensity)	28 and 43	17 and 62	0.84 (0.48-1.47)	1.07 (0.61-1.91)
Regression adjustment	318 and 1153	311 and 2512	1.01 (0.89-1.16)	0.94 (0.83-1.06)
Weighting (stabilized IPTW)	318 and 1153	311 and 2512	0.99 (0.87-1.13)	0.93 (0.82-1.06)
Matching 1:1	318 and 318	318 and 318	0.99 (0.79-1.23)	0.90 (0.72-1.13)

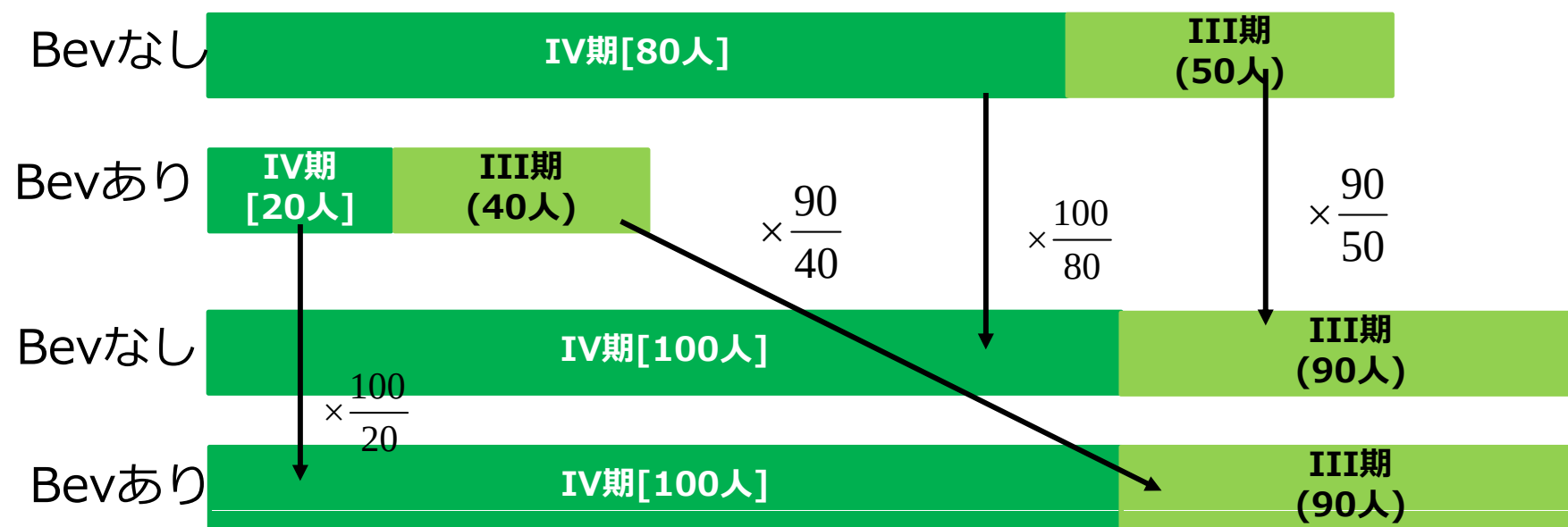
- HR, 1.01 95%CI[0.89-1.16]

IPTW解析

- 患者毎の傾向スコア（PS）の逆数で重み付けした解析
 - Bev群には $1/PS$ 、Bevなし群には $1/(1-PS)$ で重み付け
 - Cox回帰で治療法だけ共変量に入れ、PSを重みに指定
- 各群における傾向スコアの分布を揃えることで、比較可能にする
- 利点
 - 全対象を解析に含めることができる
 - 回帰モデルの誤特定の影響を受けにくい
- 欠点
 - 傾向スコアが0や1に近い場合、極端な重み付けをしてしまう

IPTWの考え方の概略図

- 簡単のため集団がStageのみで分類出来ると仮定
 - IV期の患者に対するBevありの傾向スコア=20/100
 - III期の患者に対するBevありの傾向スコア=40/90



IPTW解析の結果

Table 2. Effect of Bevacizumab Added to Carboplatin and Paclitaxel Chemotherapy on Hazard Ratios for Overall Survival

Models	Sample Sizes, No.		Hazard Ratio (95% CI)	
	Bevacizumab-Carboplatin-Paclitaxel vs Carboplatin-Paclitaxel 2006-2007	Bevacizumab-Carboplatin-Paclitaxel vs Carboplatin-Paclitaxel 2002-2005	Bevacizumab-Carboplatin-Paclitaxel vs Carboplatin-Paclitaxel 2006-2007	Bevacizumab-Carboplatin-Paclitaxel vs Carboplatin-Paclitaxel 2002-2005
Unadjusted model	318 and 1182	318 and 2664	1.00 (0.88-1.14)	0.95 (0.84-1.08)
Multivariable-adjusted model ^a	318 and 1182	318 and 2664	1.01 (0.88-1.15)	0.94 (0.83-1.06)
Propensity score-adjusted model ^b				
Stratification	318 and 1153	311 and 2512	1.01 (0.89-1.16)	0.93 (0.83-1.06)
Within-propensity score quintile				
1 (Lowest propensity)	33 and 261	51 and 716	0.75 (0.51-1.10)	0.97 (0.72-1.30)
2	69 and 337	84 and 795	1.06 (0.80-1.39)	0.92 (0.73-1.17)
3	110 and 361	107 and 640	1.06 (0.85-1.32)	0.97 (0.79-1.20)
4	78 and 151	52 and 299	1.16 (0.87-1.55)	0.81 (0.59-1.11)
5 (Highest propensity)	28 and 43	17 and 62	0.84 (0.48-1.47)	1.07 (0.61-1.91)
Regression adjustment	318 and 1153	311 and 2512	1.01 (0.89-1.16)	0.94 (0.83-1.06)
Weighting (stabilized IPTW)	318 and 1153	311 and 2512	0.99 (0.87-1.13)	0.93 (0.82-1.06)
Matching 1:1	318 and 318	318 and 318	0.99 (0.79-1.23)	0.90 (0.72-1.13)

- HR, 0.99 95%CI[0.87-1.13]

どの解析法が最も良い？

統計家の中でもディスカッション中

- 条件にもよるがIPTWが最も推定値のバイアスが少ない傾向あり

Austin PC Int J Biostat. 2009 Apr 14; 5(1):Article 13.他

- 多変量解析は唯一解析のモデルの妥当性が問われるので慎重に使うべき。ただし、解析のモデルが妥当であれば良い方法

Multivariate Behav Res. 2011 May; 46(3): 399–424.

- マッチングは相対的にバイアスが少ないので、まずマッチングを行い、次に一般化可能性を高めるため層別解析、多変量解析などを用いるのが良い

Katz, Mitchell H. Cambridge University Press, 2010.

傾向スコア解析のメリット(vs.多変量解析)

1. 推定した傾向スコアが妥当かを評価することが容易

- 予後に関係している因子を全て特定することは不可能だが、治療選択に関係している因子は特定可
- 群間で患者別の背景因子の分布の確認
- 背景因子がほとんど揃わなければ、揃っている対象だけで解析 or 調整した解析は不能と判断できる

2. 傾向スコアはアウトカムと用いずに推定可能

3. アウトカムがレアな時でも柔軟に対応可能

- 目安として多変量解析では1つの変数に対し10イベント必要
- モデルを使った多変量解析の場合も多くの変数を入れなくて良い

4. 傾向スコア推定のためのモデルの誤特定に対し頑健

- 条件を限定すれば、傾向スコア推定のためのモデルが間違っているとしても治療効果の推定値にバイアスが入りにくいと言われている

傾向スコアの解析をしないとマズイか？

- 過去の発表された論文をレビューすると、従来の手法(傾向スコアを用いずに、多変量Cox回帰や多変量ロジスティック回帰を行う)と傾向スコアを用いた解析は、ほとんど結果が変わらない

Shah et al. Journal of Clinical Epidemiology 58(2005) 550-559.
Sturmer et al. Journal of Clinical Epidemiology 59(2006) 437-447.

- 一般の単純な解析で目的が十分果たされるなら、従来の手法を用いていれば十分

Katz, Mitchell H. Cambridge University Press, 2010.

- 結局の所、傾向スコアの解析・傾向スコアを用いない多変量解析を全て行って（感度解析）結果の頑健性を確認すべし

Katz, Mitchell H. Cambridge University Press, 2010.

傾向スコアを用いない多変量解析の結果

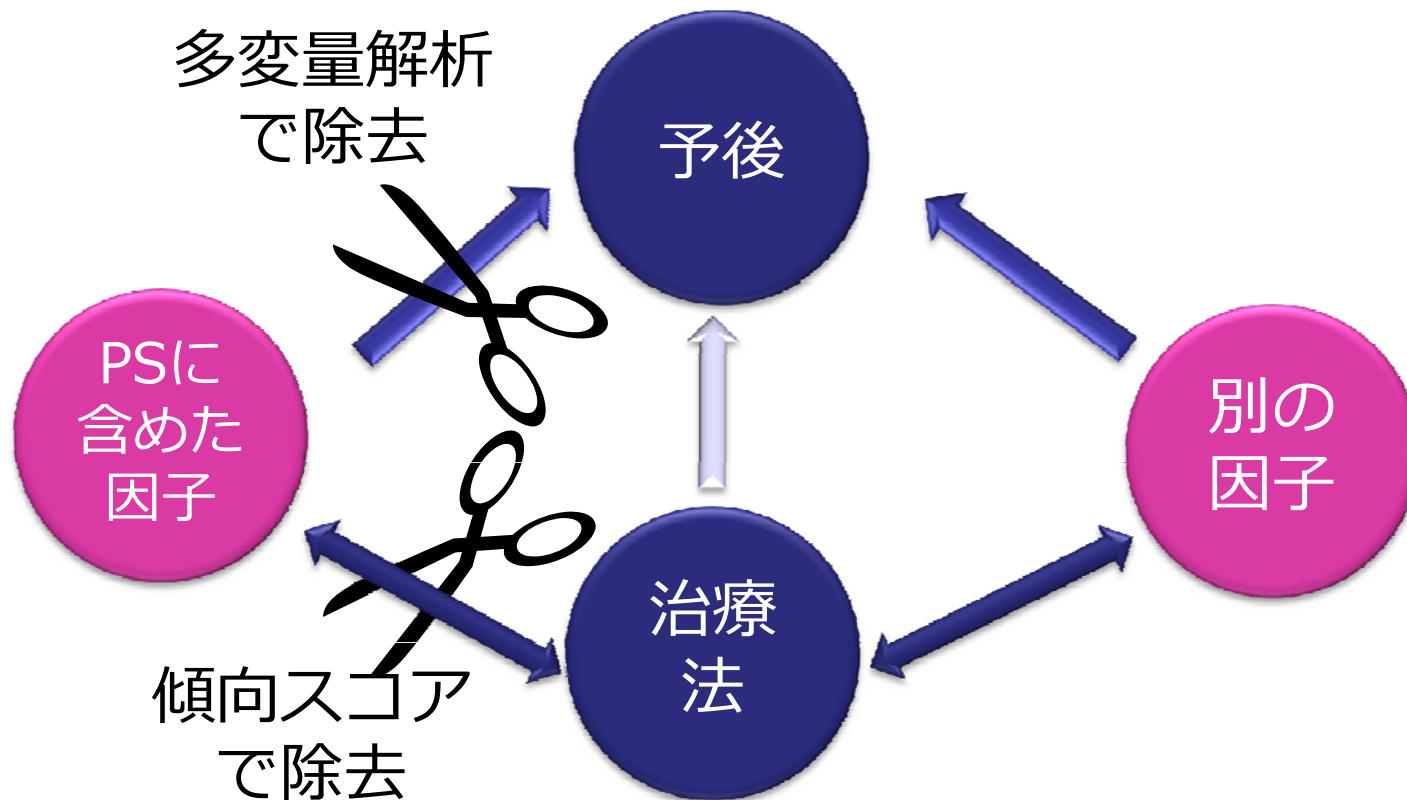
Table 2. Effect of Bevacizumab Added to Carboplatin and Paclitaxel Chemotherapy on Hazard Ratios for Overall Survival

Models	Sample Sizes, No.		Hazard Ratio (95% CI)	
	Bevacizumab-Carboplatin-Paclitaxel vs Carboplatin-Paclitaxel 2006-2007	Bevacizumab-Carboplatin-Paclitaxel vs Carboplatin-Paclitaxel 2002-2005	Bevacizumab-Carboplatin-Paclitaxel vs Carboplatin-Paclitaxel 2006-2007	Bevacizumab-Carboplatin-Paclitaxel vs Carboplatin-Paclitaxel 2002-2005
Unadjusted model	318 and 1182	318 and 2664	1.00 (0.88-1.14)	0.95 (0.84-1.08)
Multivariable-adjusted model ^a	318 and 1182	318 and 2664	1.01 (0.88-1.15)	0.94 (0.83-1.06)
Propensity score-adjusted model ^b				
Stratification	318 and 1153	311 and 2512	1.01 (0.89-1.16)	0.93 (0.83-1.06)
Within-propensity score quintile				
1 (Lowest propensity)	33 and 261	51 and 716	0.75 (0.51-1.10)	0.97 (0.72-1.30)
2	69 and 337	84 and 795	1.06 (0.80-1.39)	0.92 (0.73-1.17)
3	110 and 361	107 and 640	1.06 (0.85-1.32)	0.97 (0.79-1.20)
4	78 and 151	52 and 299	1.16 (0.87-1.55)	0.81 (0.59-1.11)
5 (Highest propensity)	28 and 43	17 and 62	0.84 (0.48-1.47)	1.07 (0.61-1.91)
Regression adjustment	318 and 1153	311 and 2512	1.01 (0.89-1.16)	0.94 (0.83-1.06)
Weighting (stabilized IPTW)	318 and 1153	311 and 2512	0.99 (0.87-1.13)	0.93 (0.82-1.06)
Matching 1:1	318 and 318	318 and 318	0.99 (0.79-1.23)	0.90 (0.72-1.13)

- HR, 1.01 95%CI[0.88-1.15]

傾向スコア & 多変量解析の欠点

- 傾向スコア(PS)の解析で調整していない因子の影響が除去できない
 - 調整していない因子が影響していたら、バイアスが入ったまま



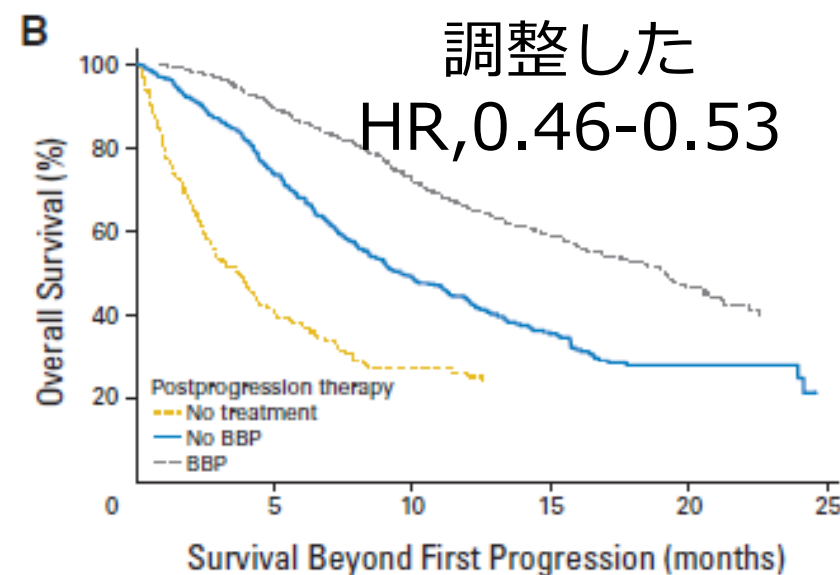
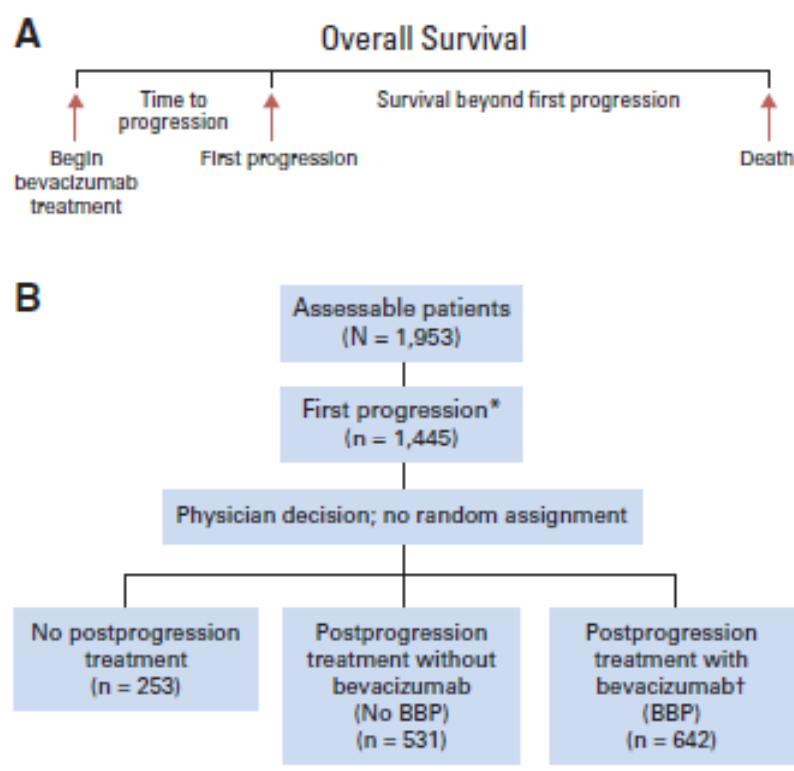
ランダム化の代わりにはならない！

- 調整していない変数が交絡因子になっていない保証はできないのでランダム化の代わりになることはない
- 残念ながら、多くの（？）先生方が考えているような魔法ではない
- Coxの多変量解析が流行りだした頃に、当時の臨床家が「Cox回帰を使えばランダム化は不要」と騒ぎ出した状況に近い

NEJM 295:74-80,1976

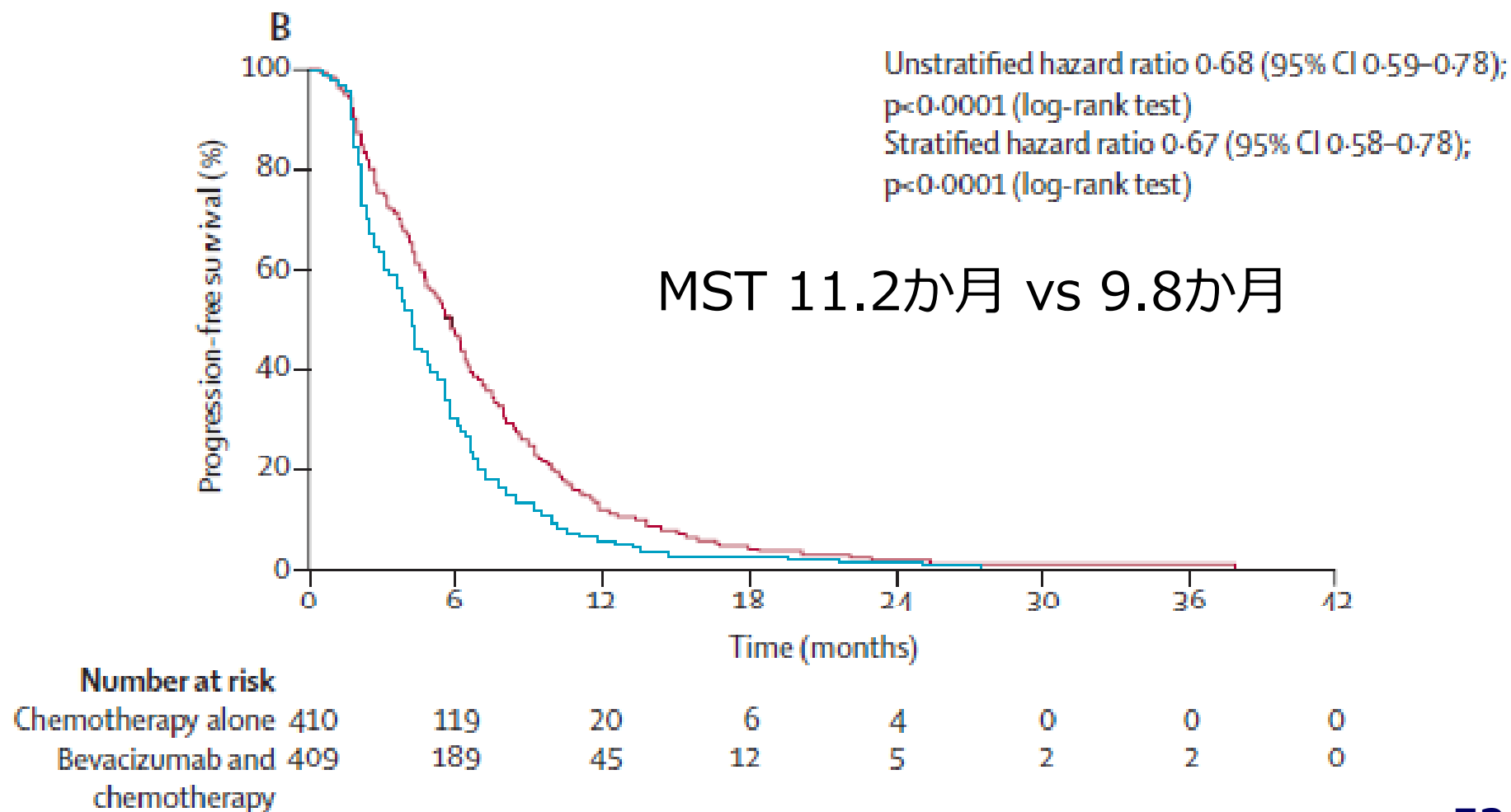
傾向スコアがあればRCT不要・・・？

- ・ 転移性大腸がんに対するBevのBeyond PDの効果を検討した観察研究（BRiTE study）
- ・ 通常のCox多変量解析以外にも傾向スコアを用いた解析も行っている



	All patients (N = 1,953)	No post- progression treatment (n = 253)	No BBP (n = 531)	BBP (n = 642)
Number of deaths	932	168	305	260
Percent	47.7	66.4	57.8	40.5
1-year survival rate, %	74.7	52.5	77.3	87.7
95% CI	72.7 to 76.7	46.2 to 58.8	73.7 to 80.9	85.2 to 90.3
Median survival beyond first progression, months	12.0	3.6	9.5	19.2
95% CI	11.1 to 13.3	2.7 to 4.3	8.4 to 11.2	16.8 to 20.7

RCTの結果



Take home message

- 「ある背景因子を持つ患者が一方の治療を受ける確率」のことを傾向スコアという
- 多変量解析よりも傾向スコアを用いた解析の方がモデルの間違っているかどうかを検討しやすい
- 未知の交絡因子・測定していない交絡因子の影響を傾向スコアで調整することは不可能なので、ランダム化の代わりにはなりえない
- 統計家の間でも方法論はディスカッションの最中
- どうしても傾向スコアを用いた解析を行いたい時は、臨床医だけで安易に行わず、データ収集前に統計家に相談

Appendix

マッチングの方法

- Optimal法
 - 全てのペアの傾向スコアの差が最も小さくなるように
- 最近隣法(Nearest neighbor method : NNM)
 - ある治療患者と傾向スコアが最も近い無治療患者をペアに
- NNM+ Caliper
 - ある許容範囲（キャリパー）を超えたらマッチしない
 - 背景が異なる患者をペアにしないように
 - マッチングされない患者が多くなるデメリットあり
- NNMとOptimalは背景因子のバランスがほとんど変わらないので、どちらを使っても良さそう

層別解析を前提に個々の背景因子別に確認した例

The cohort was then divided into **5 strata** defined by quintiles of estimated propensity scores. Next, we used *P* values of the 2 test to assess whether patients' baseline characteristics were balanced across the 2 treatment groups within each stratum.

eTable 1. *P* Values for the Chi-Square Test Comparing Characteristics Between BCP and CP 2002-5 Elderly Advanced Non-Squamous NSCLC Patients, by Propensity Score Quintile

Characteristics	BCP and CP 2006-7 Cohort				
	Strata 1	Strata 2	Strata 3	Strata 4	Strata 5
Age at diagnosis	0.54	0.61	0.12	0.67	0.24
Gender	0.71	0.91	1.00	0.99	0.21
Race/ethnicity	1.00	0.41	0.26	1.00	1.00
Marital status	0.57	0.94	0.72	1.00	0.27
Median income (census tract quintile) ^a	0.14	0.25	0.81	0.21	0.36
College educated (census tract quintile) ^b	0.95	0.31	0.98	0.54	0.00
SEER region	0.54	0.05	0.57	0.68	0.10
Urban residency ^c	1.00	0.58	1.00	0.44	0.24
Modified Charlson comorbidity score ^d	0.56	0.93	0.55	0.73	0.74
Level of differentiation (Tumor grading)	0.46	1.00	0.86	0.45	0.94
AJCC stage	1.00	0.45	0.05	0.27	1.00