

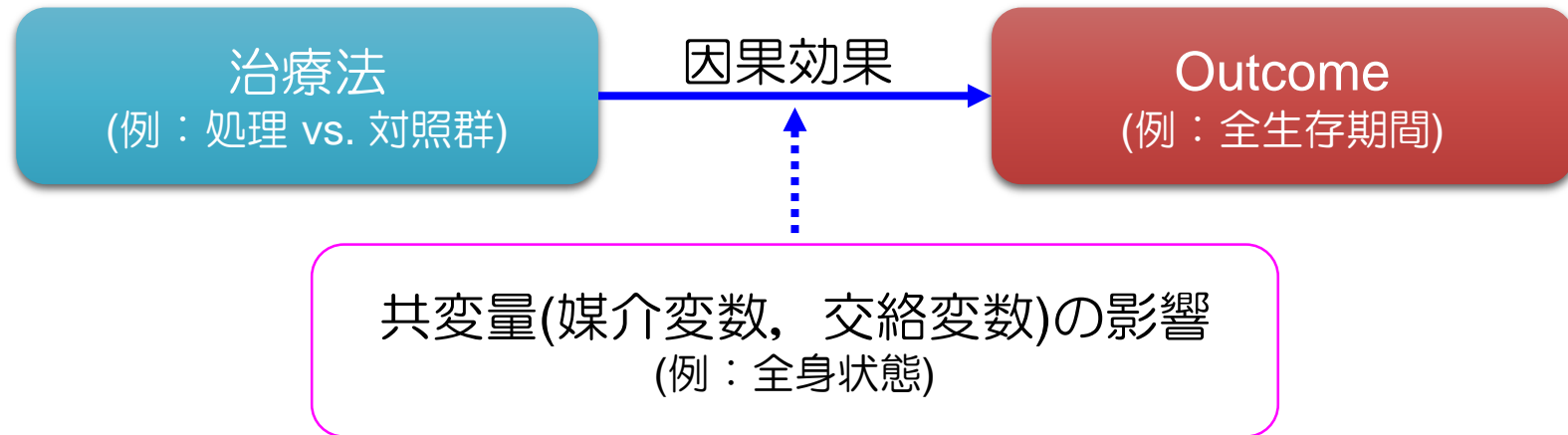
医学統計セミナー 第4回

- 傾向スコア分析 -

下川 敏雄

和歌山県立医科大学

□ 治療法の評価 = ある医学的介入による影響(因果効果)の検証



共変量が厄介な理由：コントロールできないものが多い

検証的な側面から共変量の影響を最小限にして因果効果を評価する

— 無作為化比較試験(RCT:Randomized Controlled Trial)

観察的な側面(観察研究)において(治療選択に対する)共変量の影響を最小限にして因果効果を評価する

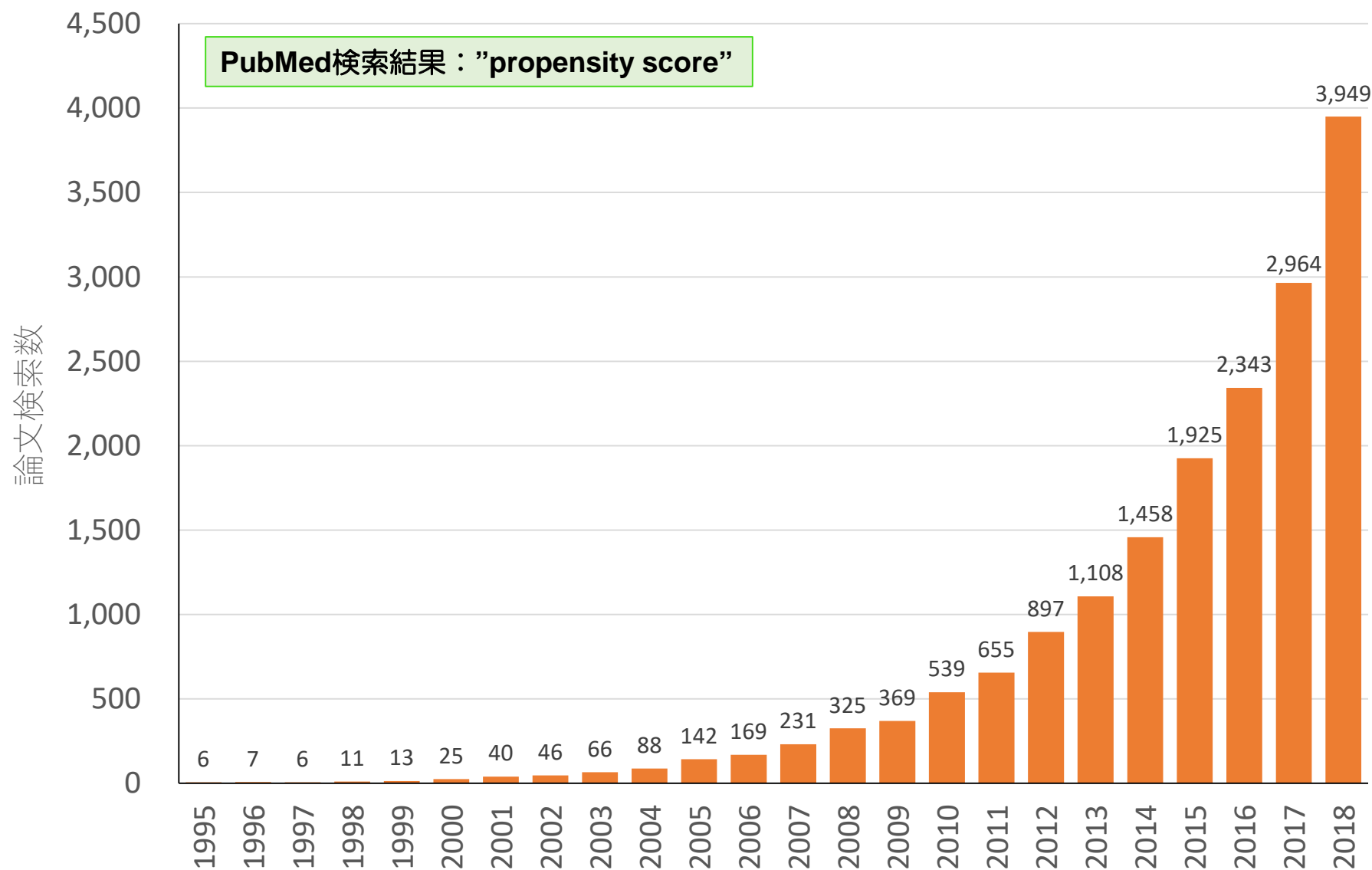
— 統計学的な調整法 (共変量とOutcomeの線形性などの制約)

— 傾向スコア(Propensity score) の利用

傾向スコア分析の論文数の推移

医学統計セミナー：傾向スコア
2019.2.5

2018年は12月末現在



観察研究における共変量を考慮した研究及び解析(星野・岡田, 2006)

研究デザイン

- (1) 均衡化：共変量の値が同じになるペアを作ることによって2つの群の被験者を構成
 - 完全に一致するペアを作ることとは事実上不可能
 - 多数の共変量を含めることができない
 - ペアは観測者が決めるため、主観的で恣意性を排除できない。
- (2) 恒常化・限定：共変量のある被験者のみに限定して解析を行う
 - 一部の被験者に限定するため、研究の一般可能性が低くなる。
 - 多数の共変量を含めることができない

統計

- (3) 統計的調整(多変量解析等)
 - 誤った統計的モデルを用いると誤った結果を導く可能性がある。

複数の共変量を一つの変数(ものさし)におきかえ，そのもとで均衡化や層別化などを行い，誤った統計モデルを用いても「頑健」な結果をえることができる方法



傾向スコア分析(propensity score)

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Perioperative Beta-Blocker Therapy and Mortality after Major Noncardiac Surgery

Peter K. Lindenauer, M.D., Penelope Pekow, Ph.D., Kaijun Wang, M.S.,
Dheeresh K. Mamidi, M.B., B.S., M.P.H., Benjamin Gutierrez, Ph.D.,
and Evan M. Benjamin, M.D.

ABSTRACT

BACKGROUND

Despite limited evidence from randomized trials, perioperative treatment with beta-blockers is now widely advocated. We assessed the use of perioperative beta-blockers and their association with in-hospital mortality in routine clinical practice.

METHODS

We conducted a retrospective cohort study of patients 18 years of age or older who underwent major noncardiac surgery in 2000 and 2001 at 329 hospitals throughout the United States. We used propensity-score matching to adjust for differences between patients who received perioperative beta-blockers and those who did not receive such therapy and compared in-hospital mortality using multivariable logistic modeling.

RESULTS

Of 782,969 patients, 663,635 (85 percent) had no recorded contraindications to beta-blockers, 122,338 of whom (18 percent) received such treatment during the first two hospital days, including 14 percent of patients with a Revised Cardiac Risk Index (RCRI) score of 0 and 44 percent with a score of 4 or higher. The relationship between perioperative beta-blocker treatment and the risk of death varied directly with cardiac risk; among the 580,665 patients with an RCRI score of 0 or 1, treatment was associated with no benefit and possible harm, whereas among the patients with an RCRI score of 2, 3, or 4 or more, the adjusted odds ratios for death in the hospital were 0.88 (95 percent confidence interval, 0.80 to 0.98), 0.71 (95 percent confidence interval, 0.63 to 0.80), and 0.58 (95 percent confidence interval, 0.50 to 0.67), respectively.

CONCLUSIONS

Perioperative beta-blocker therapy is associated with a reduced risk of in-hospital death among high-risk, but not low-risk, patients undergoing major noncardiac surgery. Patient safety may be enhanced by increasing the use of beta-blockers in high-risk patients.

From the Division of Healthcare Quality, Baystate Medical Center, Springfield, Mass. (P.K.L., P.P., E.M.B.); the Department of Medicine, Tufts University School of Medicine, Boston (P.K.L., E.M.B.); the School of Public Health and Health Sciences, University of Massachusetts at Amherst, Amherst (P.P., K.W., D.K.M.); and Premier Healthcare Informatics, Premier Inc., Charlotte, N.C. (B.G.) Address reprint requests to Dr. Lindenauer at the Division of Healthcare Quality, Baystate Medical Center, 759 Chestnut St. P-5928, Springfield, MA 01199, or at Peter.Lindenauer@bhs.org.

N Engl J Med 2005;353:349-61.

Copyright © 2005 Massachusetts Medical Society.

心臓以外の大手術において、β遮断薬を術前投与することの有用性について傾向スコア分析に基づいて評価している。

そこでは、β遮断薬投与群と非投与群に分けて、術前β遮断薬投与によって死亡リスクが提言するか否かを評価している。

Revised Cardiac Risk Index (RCRI)

心機能低下の評価指標として用いられている指標

- (1) 高リスク手術
- (2) 冠動脈疾患の既往
- (3) 心不全所見
- (4) 脳血管障害の既往
- (5) インスリン依存性糖尿病
- (6) 腎不全

を用いて得点化している。

リスク因子が少ない患者に対する β ブロッカーの投与は院内死亡を増加させる恐れがある。

一方で、高リスクな患者に対しては、 β ブロッカーの投与が院内死亡を減少させることに繋がる。

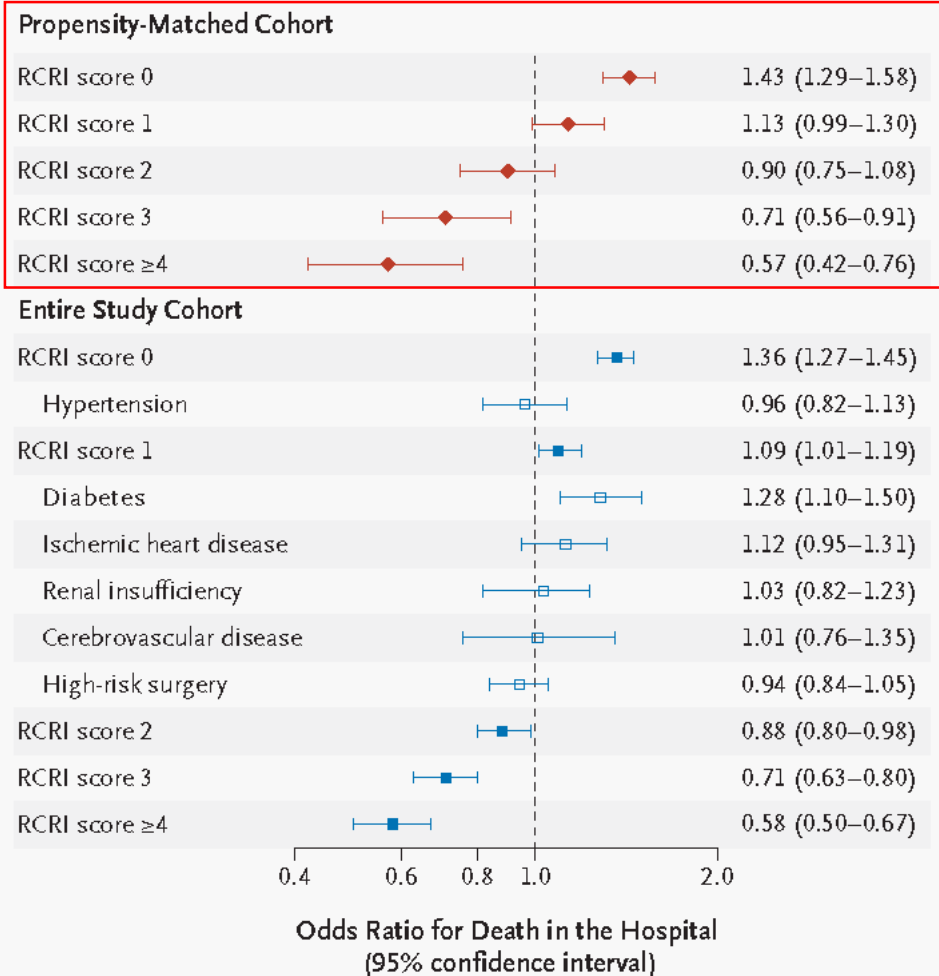
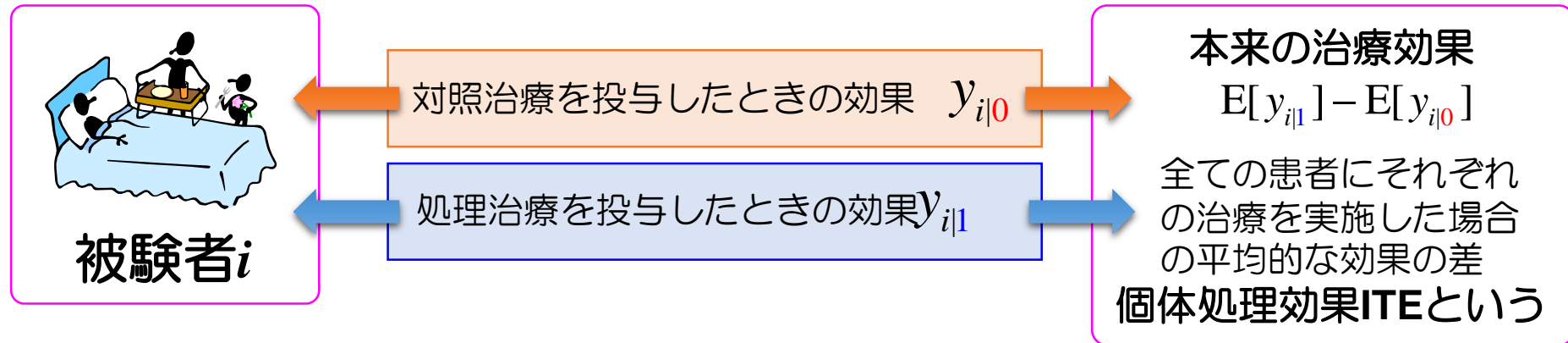


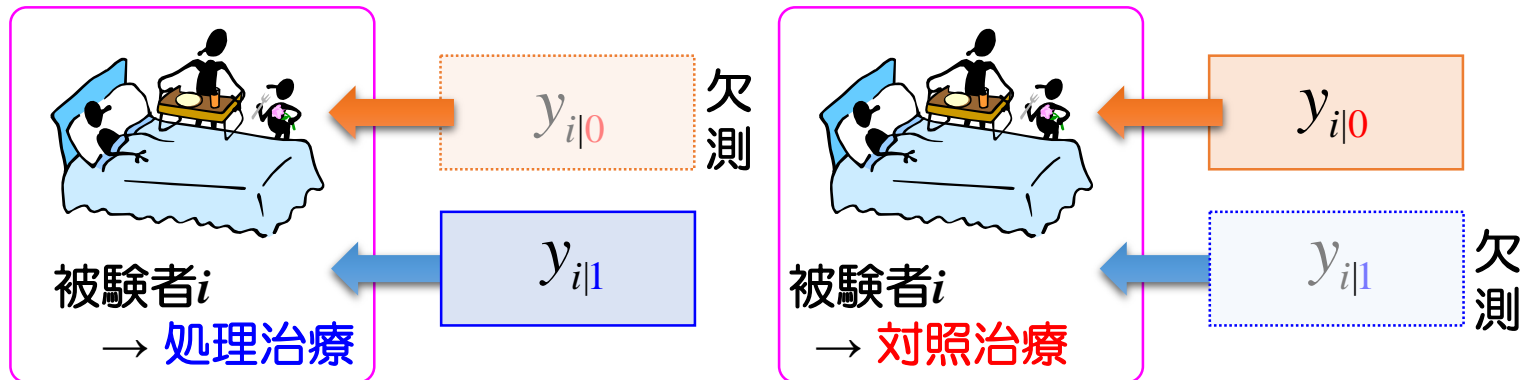
Figure 1. Adjusted Odds Ratio for In-Hospital Death Associated with Perioperative Beta-Blocker Therapy among Patients Undergoing Major Noncardiac Surgery, According to the RCRI Score and the Presence of Other Risk Factors in the Propensity-Matched Cohort and the Entire Study Cohort.

Open boxes represent patient subgroups within the listed RCRI category.

因果効果の定義：Neyman-Rubinの反事実的モデル



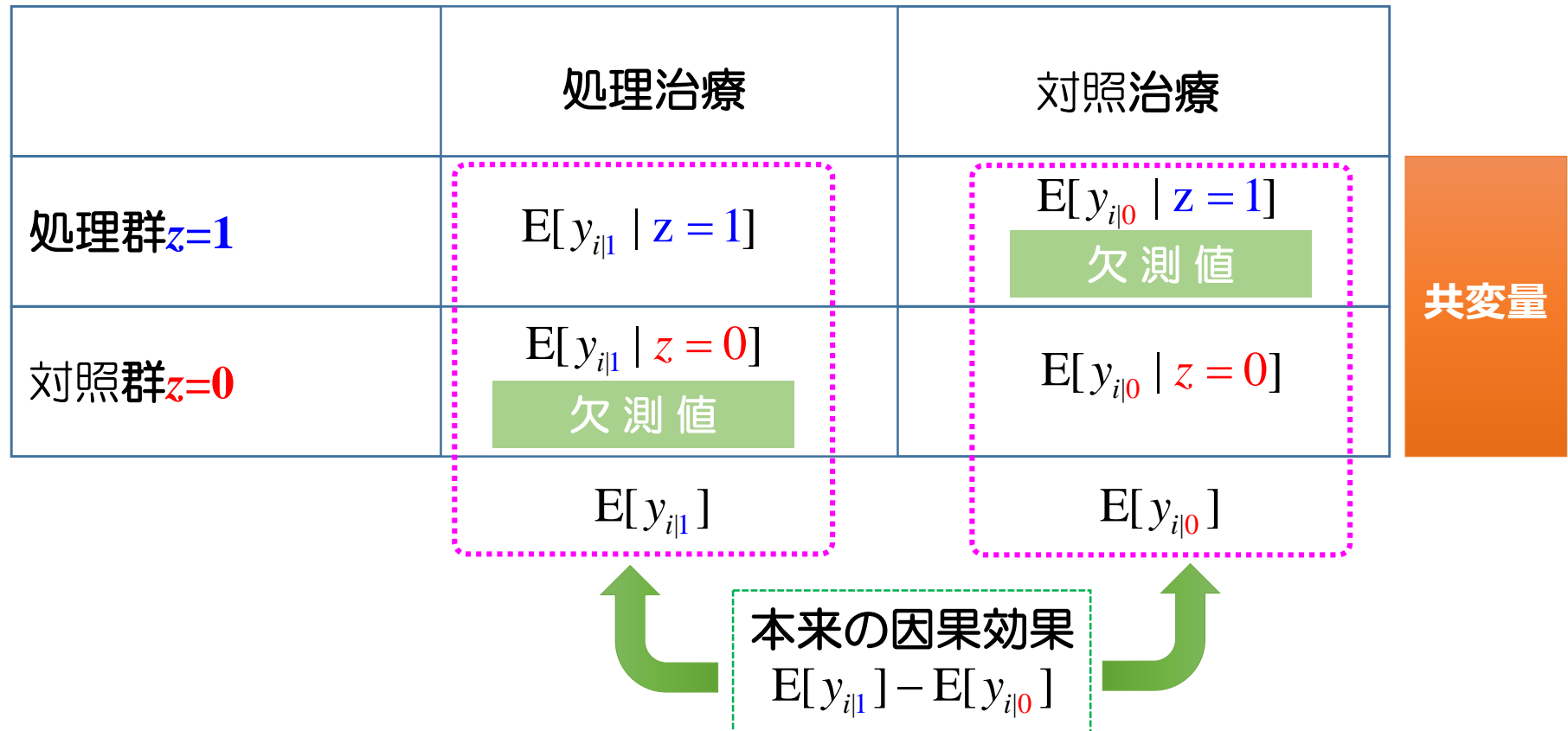
しかしながら、現実には...



となり、いずれか一方は欠測であると考えられる。

□ 無作為化比較試験の枠組みで考える

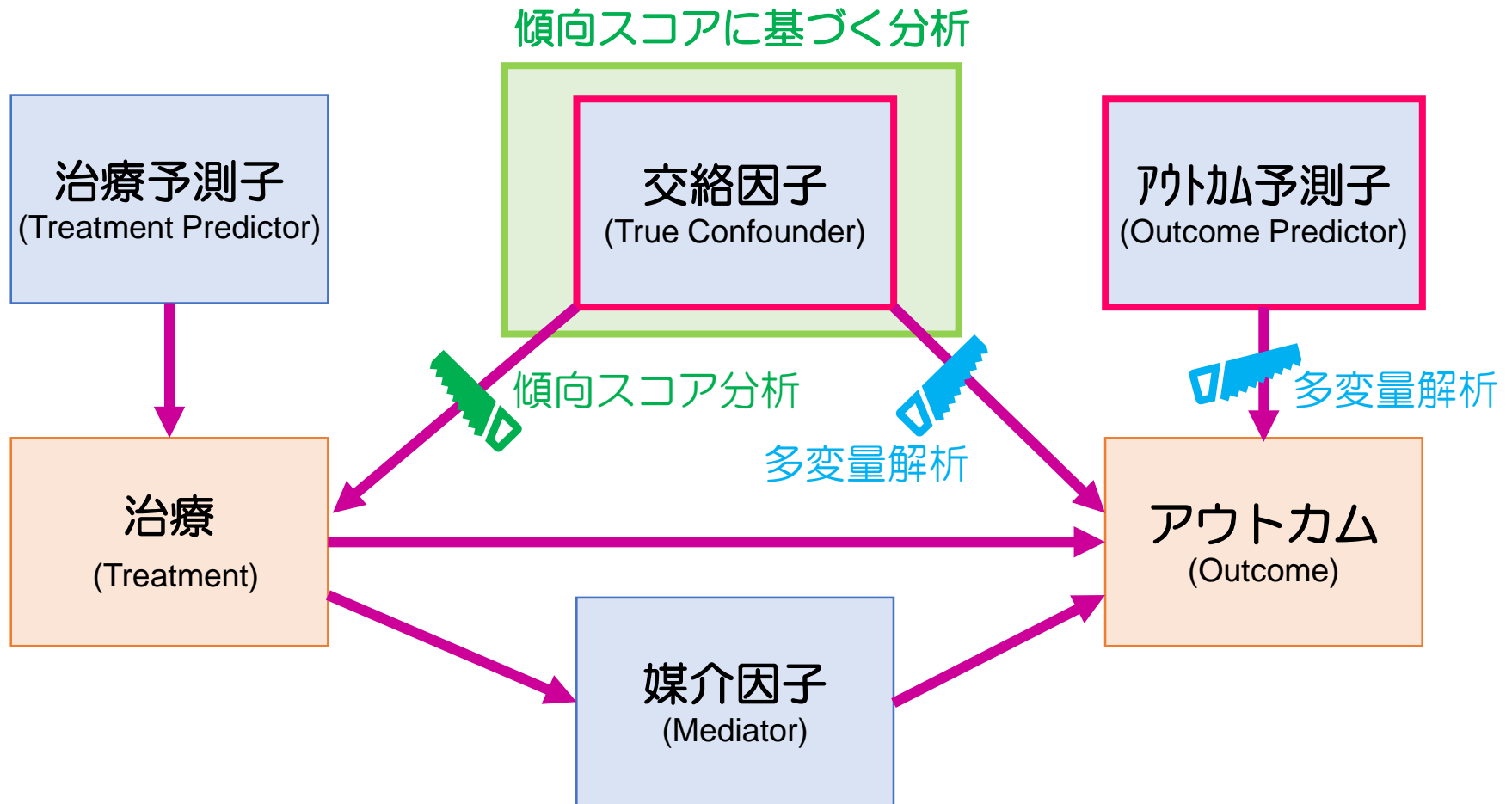
対照群 $z=0$ ，処理群 $z=1$ としたときの平均的な効果(z は割付変数)





となることから、欠測と捉えられていた部分を無作為化によって補うことで、本来の因果効果を評価することができる。

因果関係に関する模式図(Leite, 2017)

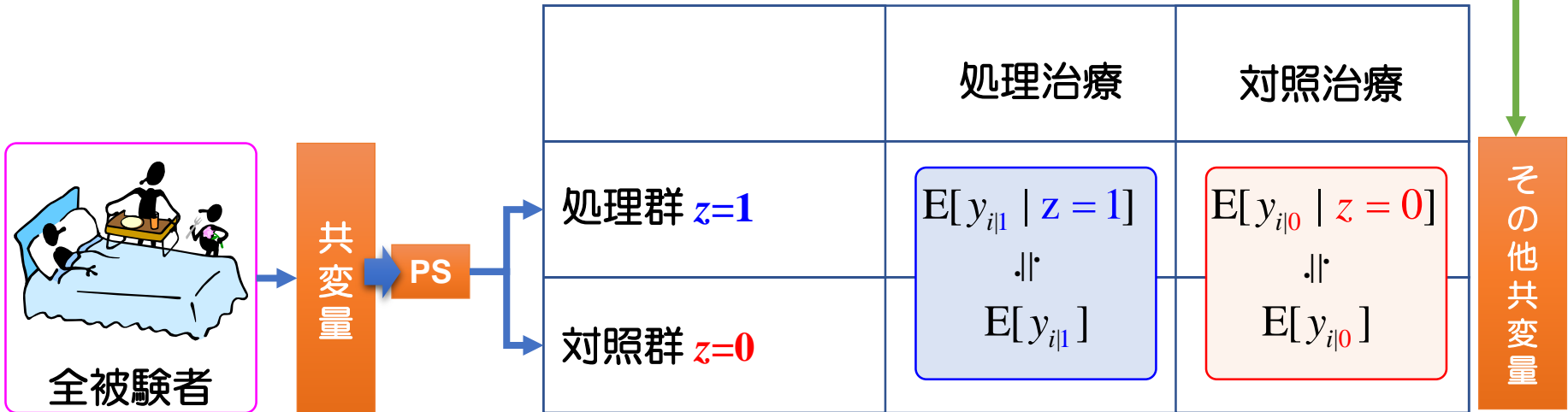


それぞれの因子の意味(Brookhart et al., 2006)

交絡因子：因果効果の偏りとバラツキの削減に繋がる。

アウトカム予測子：因果効果のバラツキの削減に繋がる(偏りはない)。

傾向スコアの計算に用いた共変量は事後の調整因子等に用いない。



傾向スコアを用いることで複数の共変量から1つの変数に集約できる (傾向スコアモデルに関する適切性の評価をc統計量、疑似決定係数で評価すること(評価していない論文が多いことが指摘されている))

共変量による投与群の選択性を揃えることで、処理の割付については偶発性のみに依存するようにする。これにより、群間の被験者層を揃えることができる。

すなわち、傾向スコアとは、観察研究でのインバランスな群間のバランスを行う「ものさし」となる。

	処理治療	対照治療
処理群 $z=1$	$E[y_{i1} z = 1]$ 	$E[y_{i0} z = 1]$ 欠測値
対照群 $z=0$	$E[y_{i1} z = 0]$ 欠測値	$E[y_{i0} z = 0]$
	$E[y_{i1}]$	$E[y_{i0}]$

- 平均処理効果(ATE: Average Treatment Effect) ・ 平均因果効果(ACE: Average Causal Effect)

母集団のすべての個体での処理でのアウトカムの期待値と対照でのアウトカムの期待値の差

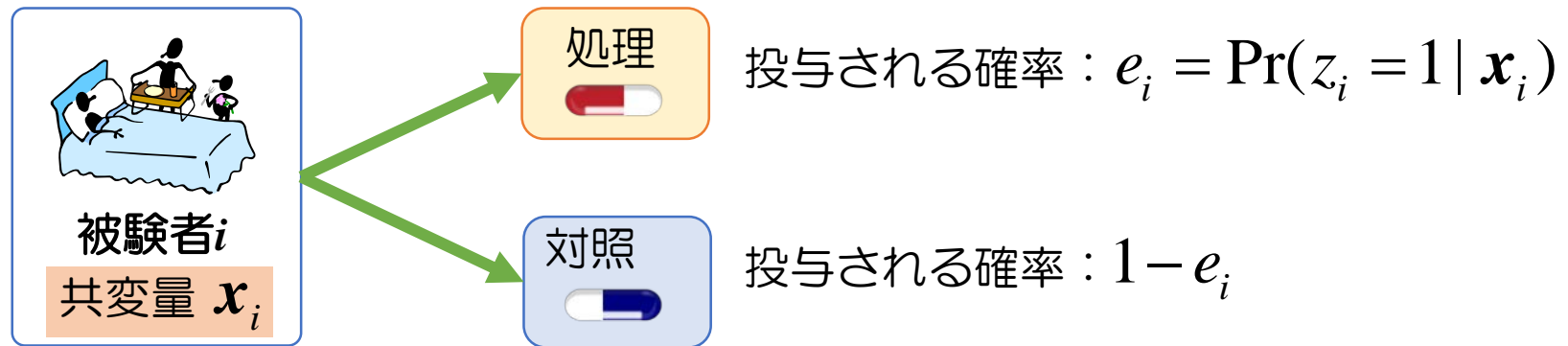
$$ATE = E[y_{i1}] - E[y_{i0}]$$

- 処理群の平均処理効果(ATT: Average Treatment effect for Treated)

母集団のうち処理群での処理のアウトカムの期待値と対照でのアウトカムの期待値の差

$$ATT = E[y_{i1} | Z = 1] - E[y_{i0} | Z = 1]$$

Step	目標	手順の例
1. 準備	解析可能な「完全な」データ集合を得る	<ul style="list-style-type: none">• 共変量選択• 欠測値に対する処理
2. 傾向スコアの推定	処理群と対照群に対して傾向スコアを推定する.	<ul style="list-style-type: none">• ロジスティック回帰• Random Forests• 一般化ブースティング樹木
3. 傾向スコア手順の実施	処理群と対照群の共変量の分布のバランスをとるための戦略を実施する.	<ul style="list-style-type: none">• 傾向スコア・マッチング• 傾向スコア層化• IPW(傾向スコア重み付け)
4. 共変量のバランスの評価	処理群と対照群のあいだの共分散の分布のバランスが達成された程度を評価する.	<ul style="list-style-type: none">• 標準化された平均距離の計算• 分散比の計算
5. 処置効果推定	処置効果とその標準誤差を推定する.	<ul style="list-style-type: none">• 重み付け平均距離• 一般化線形モデル
6. 感度分析	除外された共変量が処置効果の有意性検定をどの程度変化させるかについて評価する.	<ul style="list-style-type: none">• Rosenbaum(2002)の方法• Carnegie et al.(2016)の方法



このとき、処理が投与される確率 e_i は**傾向スコア(Propensity score)**と呼ばれる。

傾向スコアの推定方法として一般的に用いられる方法：**ロジスティック回帰分析**

傾向スコアの特徴

- 条件付き独立性：傾向スコアでの割付がアウトカムに影響しない
- バランシング：同じ傾向スコアの値に対応した共変量の分布は、それぞれの群で等しい。

ただし、傾向スコアを計算するのに用いた共変量の範囲でのことであり、観測されていない(あるいは、傾向スコアの計算に用いていない)共変量があり、それが交絡因子である場合には、その限りではない) したがって、傾向スコアを計算するうえで、**もちいる共変量は十分に吟味しなければならない。**

Annals of Surgery, 257(4), 640-646.

Outcomes After Laparoscopic or Open Distal Gastrectomy for Early-Stage Gastric Cancer

A Propensity-Matched Analysis

Hideo Yasunaga, MD, PhD,† Hiromasa Horiguchi, PhD,* Kazuaki Kuwabara, MD, DPH,‡
Shinya Matsuda, MD, PhD,§ Kiyohide Fushimi, MD, PhD,|| Hideki Hashimoto, MD, PhD,¶
and John Z. Ayanian, MD, MPP†*

早期胃癌患者に対する開腹切除術 vs. 腹腔鏡切除術の検討をしている

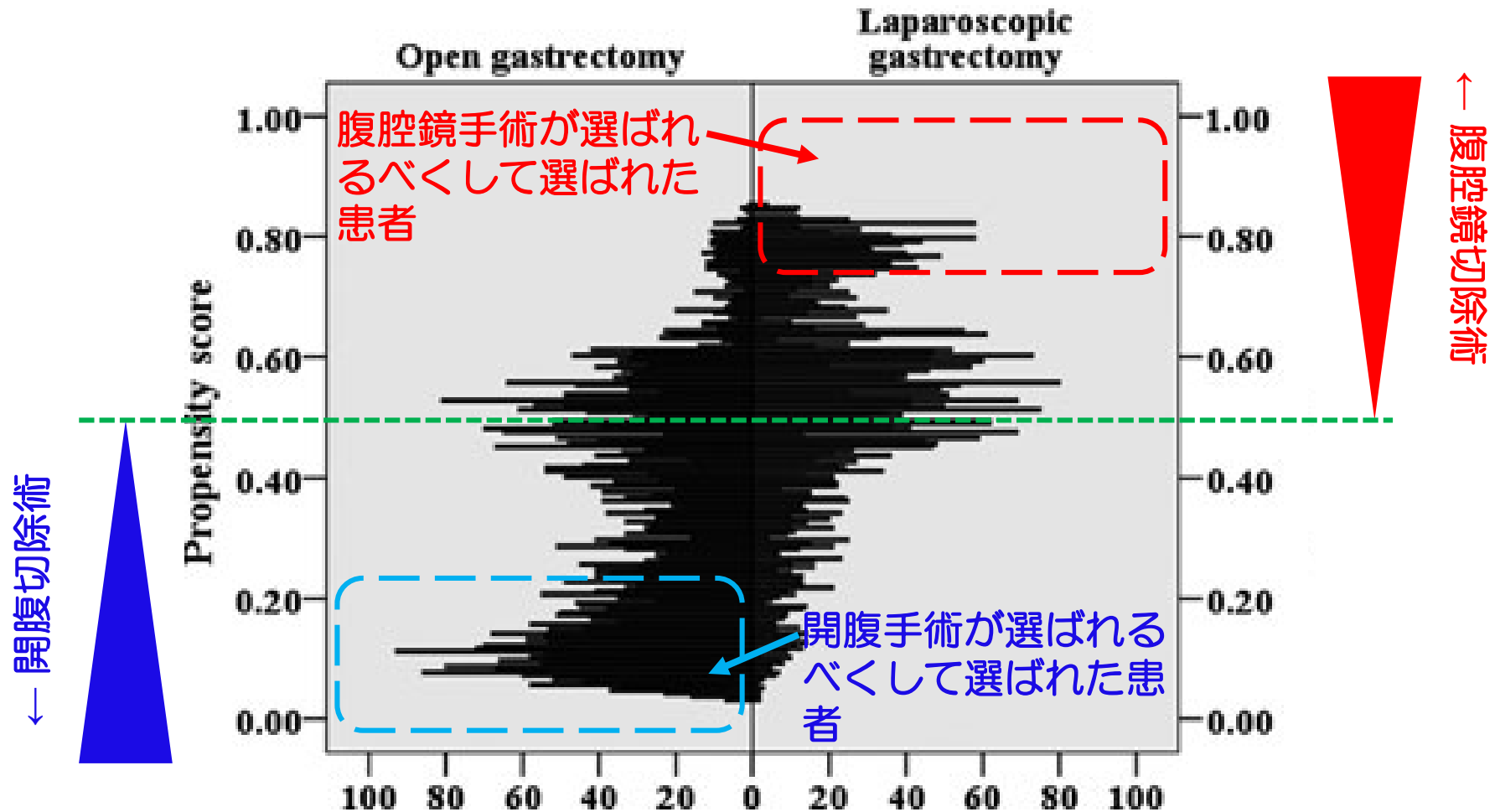
To estimate the propensity score, we fitted a **logistic regression model** for the receipt of laparoscopic gastrectomy as a function of patient demographic and hospital factors including **age, sex, Charlson comorbidity index, body mass index, smoking index, cancer stage (I or II), hospital volume category, and type of hospital (teaching or nonteaching)**. The C-statistic for evaluating the goodness of fit was calculated. Each patient who received laparoscopic gastrectomy was matched with a patient who received open gastrectomy with the closest estimated propensity on the logit scale within a specified range (≤ 0.6 of the pooled standard deviation of estimated logits) to reduce differences between treatment groups by at least 90%. (後略)

ロジスティック回帰を用いて傾向スコアを計算している

共変量

年齢, 性別, hospital volume(1年あたりの症例の手術件数)のカテゴリ,
BMI, 喫煙指数, ステイジ(I/II), 病院の種類(Teaching hospital or Nonteaching hospital)

先ほどの事例における傾向スコアの解釈



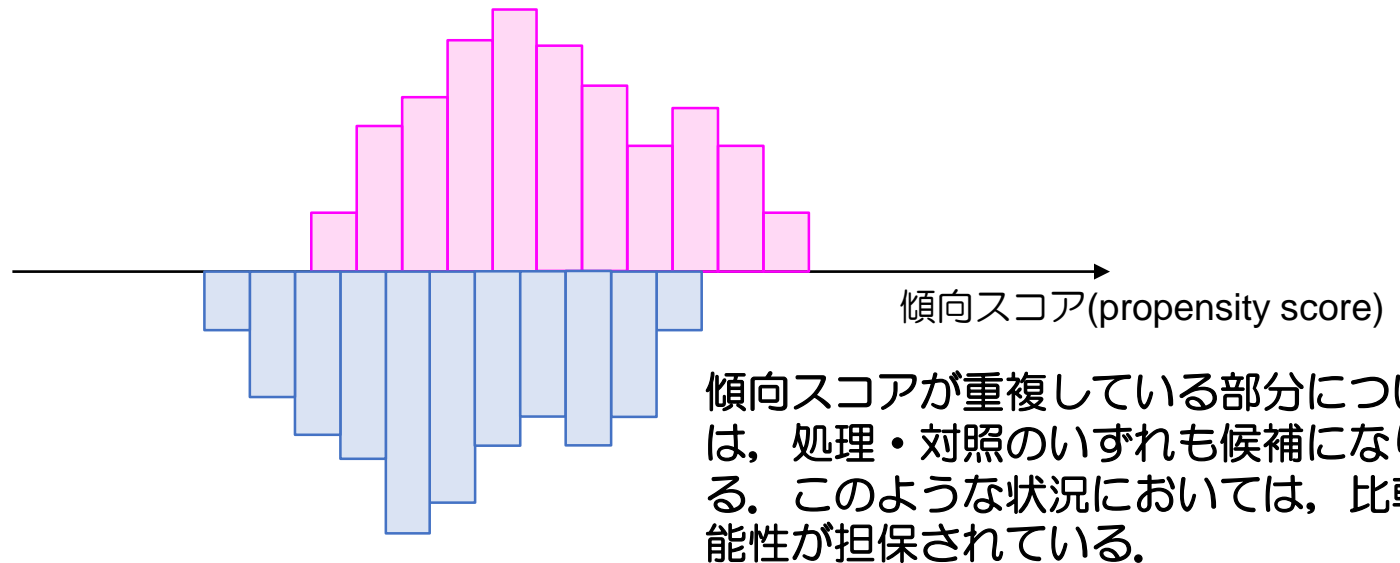
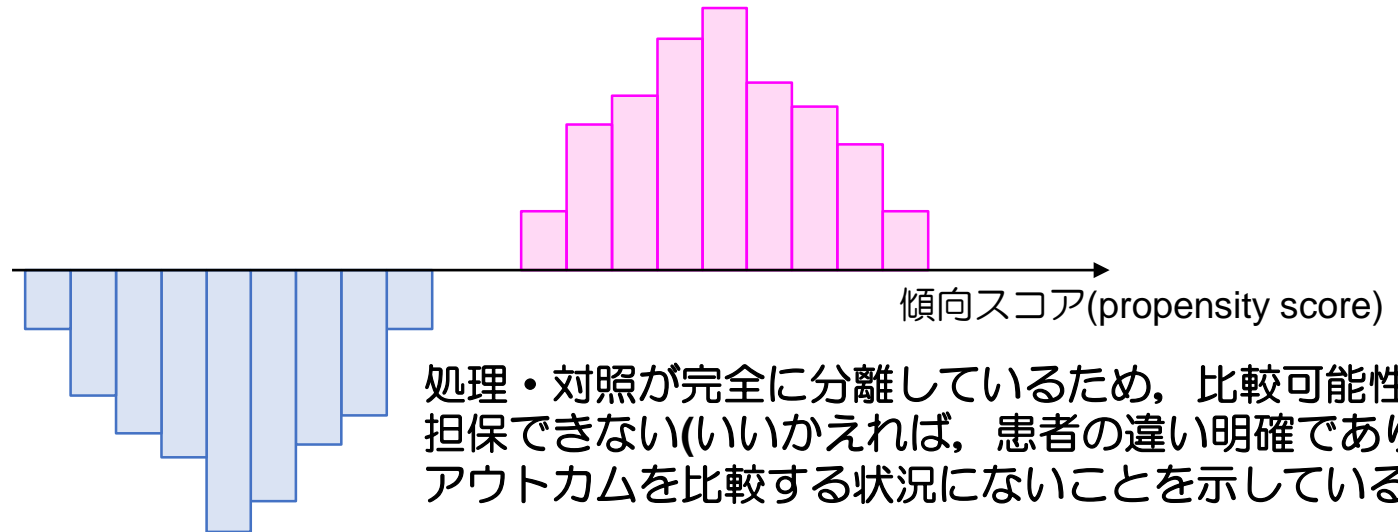
傾向スコアとは、処理が投与される確率で表される(つまり、傾向スコアは0~1の範囲をとる)。

処理群

対照群

処理群

対照群



治療選択の根拠となる共変量を選ぶ

傾向スコア計算のための共変量 x は、アウトカムへの影響を与える要因として考えるのではなく、**治療選択 z の根拠となるものを選択しなければならない**(治療選択に交絡がない状況(条件付独立性)が担保されなければならない)。

共変量は治療選択前のものを選ぶ

傾向スコアは、「処理投与(あるいは一方の治療法の選択)の確率」である。したがって、傾向スコアの計算には、治療選択前のものを選ぶ。

傾向スコア推定の精度評価

傾向スコア推定の制度は、処理群と対照群での共変量の分布が同じであるかどうかで行うものであって、両群がうまく分離されているかどうかは評価指標にならない。

- マッチングを用いた場合
 - マッチング後の個々の共変量の要約統計量の評価、効果量、検定(批判的な論文もある)による評価
 - マッチング後の分布の比較(例えば、ヴァイオリンプロット)

● マッチング

傾向スコアの一致した(あるいは極めて近い)個体同士を選択する。

● 層別

傾向スコアを用いて層別(5層程度)したうえで層別解析を行う。

● 逆確率重み付け(IPWE: Inverse Probability Weighting Estimator)

傾向スコアを用いて層別(5層程度)したうえで層別解析を行う。

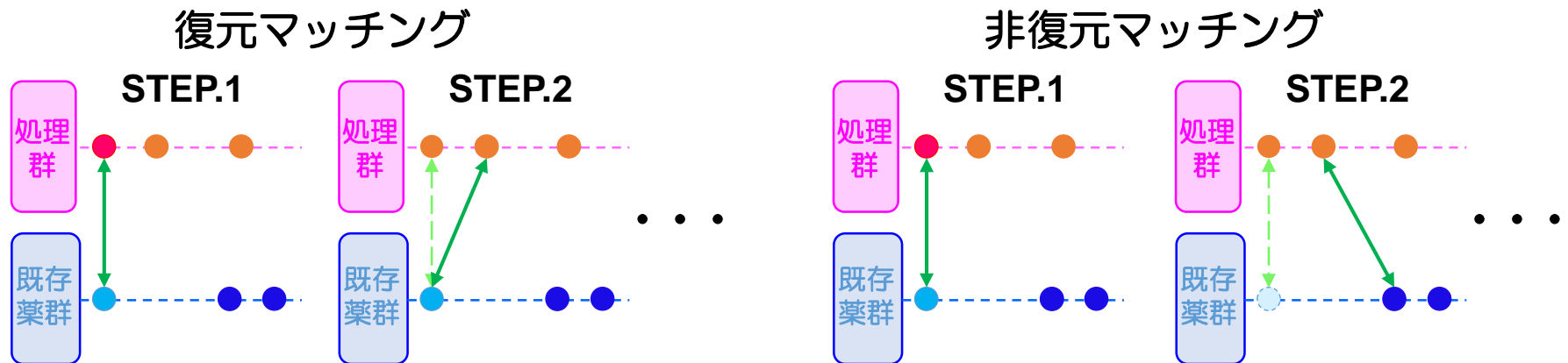
● 共分散分析

傾向スコアを共変量にとった共分散分析を実施する。

これらの手法はいずれかを使うということではなく、組み合わせて用いることもできる。例えば、マッチング後に共分散分析を行うことが考えられる。

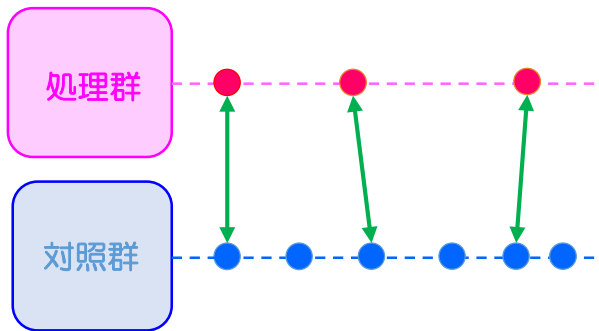
実際の応用分野では、マッチングあるいは逆確率重み付けを用いることが多いようである。

項目	説明
マッチングのアルゴリズム	<ul style="list-style-type: none"> 最近傍マッチング(nearest neighbor matching) 処理群の任意の個体に対して、傾向スコアが最も近い対照群の個体を逐次に探索する方法(マッチングの順番の影響を受ける) 最適マッチング(optimal matching) マッチング後の傾向スコアの距離の合計値が最小になるようにマッチングを行う。
マッチング後の対照群の個体の取扱い	<ul style="list-style-type: none"> 復元マッチング 処理群の異なる個体に対して同じ対照群の個体を対応させることを許容する。 非復元マッチング 処理群の異なる個体に対して同じ対照群の個体を対応させることを許容しない。
マッチングの比率	<ul style="list-style-type: none"> 1：1マッチング 1名の治療群と1名の対照群をマッチングを行う。 固定比マッチング(1：kマッチング) 1名の治療群とk名の対照群をマッチングを行う。 変動比マッチング 1名の治療群と複数(個体毎に変動、上限のみ設定)の非暴露群でマッチングを行う。
キャリパー(マッチングさせる許容領域)の設定	<p>キャリパーとは、マッチングさせる許容領域を表しており、マッチングされたペアの傾向スコアの距離がキャリパー以上であればマッチングしない。</p>

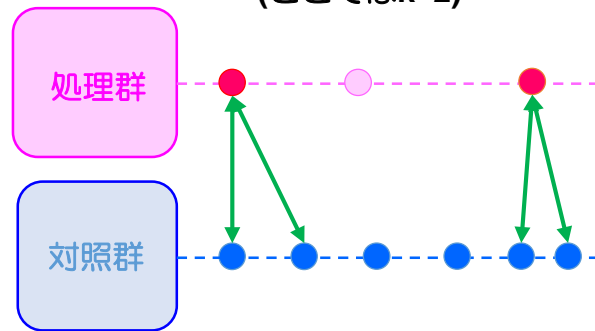


- 復元マッチングでは、処理群の異なる個体に対して同じ対照群の個体を対応させることを許容し、非復元マッチングでは許容しない。
- そのため、バイアス低減の点では、非復元マッチングのほうが優れているが、復元マッチングは殆ど用いられていない。
- 非復元マッチングでは選択する個体の順番の影響を受ける (強欲マッチング以外を利用した場合)。
- マッチング数(例数)が少ない場合には、復元マッチングが推奨される(Rosenbaum, 1989)。
- 一方で、マッチング数が多くなるにつれて、二つの差はほとんどなくなる。
- 岩崎(2015)では、非復元マッチングが推奨されている。

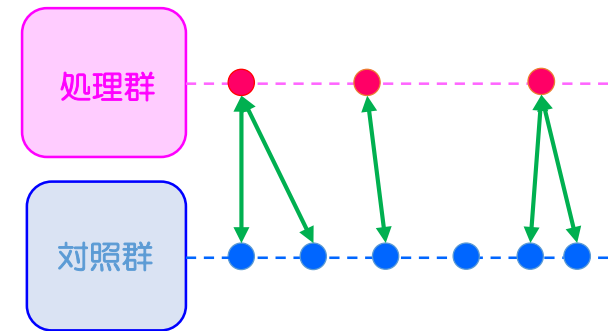
1:1マッチング



固定比マッチング (ここではk=2)



変動比マッチング



■ 1:1マッチング[one-to-one matching]

1名の処理群と1名の対照群でマッチングする方法である。

- 例数の減少が最も顕著であるが、群間の例数の不均衡が起こらない。したがって、検出力の低下が最も少ない(Cohen, 1988)。

■ 固定比(1:k) マッチング[fixed rate matching, one-to-k matching]

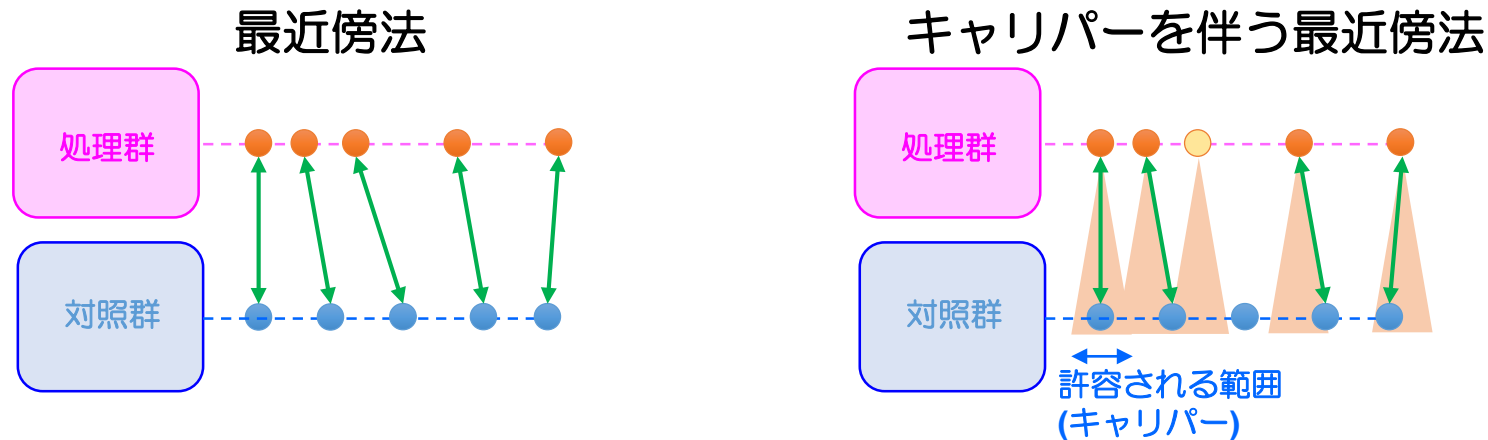
1名の処理群とk名の対照群でマッチングする方法である。

- 最も推奨されないマッチング法である(Leite, 2017)。

■ 変動比マッチング[variable rate matching, one-to-many matching]

1名の処理群と複数(個体ごとに変動, 上限のみ設定)の対照群でマッチングする方法

- 処理群の例数が対照群の例数よりかなり少ない場合には有効(Leite, 2017)。
- 変動比マッチングは1:1マッチングに比べてバイアス除去に優れている(Cepeda et al, 2003; Gu & Rosenbaum, 1993; Ming & Rosenbaum, 2000)



■ 最近傍マッチング [Nearest neighbor]

処理群に一番近い傾向スコアを持つ対照群をマッチングする方法である。

■ キャリパーを伴う最近傍マッチング[Nearest neighbor within caliper]

マッチングさせる許容領域(キャリパー)を設定した最近傍マッチング(キャリパーを大きくするほどマッチングされる個体は増加する)。

— Rosenbaum & Rubin(1985)は $0.25 \times SD$ をキャリパーに設定することを推奨している(0.20を推奨する報告もあるため、0.25~0.20を推奨している)。

キャリパーを設定することが一般的である。

最適マッチング

ID	PS	ID	PS
T1	0.46	C1	0.28
T2	0.92	C2	0.12
T3	0.89	C3	0.68
T4	0.63	C5	0.26
T5	0.72	C5	0.51
T6	0.32	C6	0.31
T7	0.28	C7	0.48

ID	PS	ID	PS	距離
T7	0.28	C2	0.12	0.16
T6	0.32	C5	0.26	0.06
T1	0.46	C1	0.28	0.18
T4	0.63	C6	0.31	0.32
T5	0.72	C7	0.48	0.24
T3	0.89	C5	0.51	0.38
T2	0.92	C3	0.68	0.24

合計=1.58

最適マッチング(Optimal matching)とは、PSの距離の**合計値**が最小になるようにペアを作る

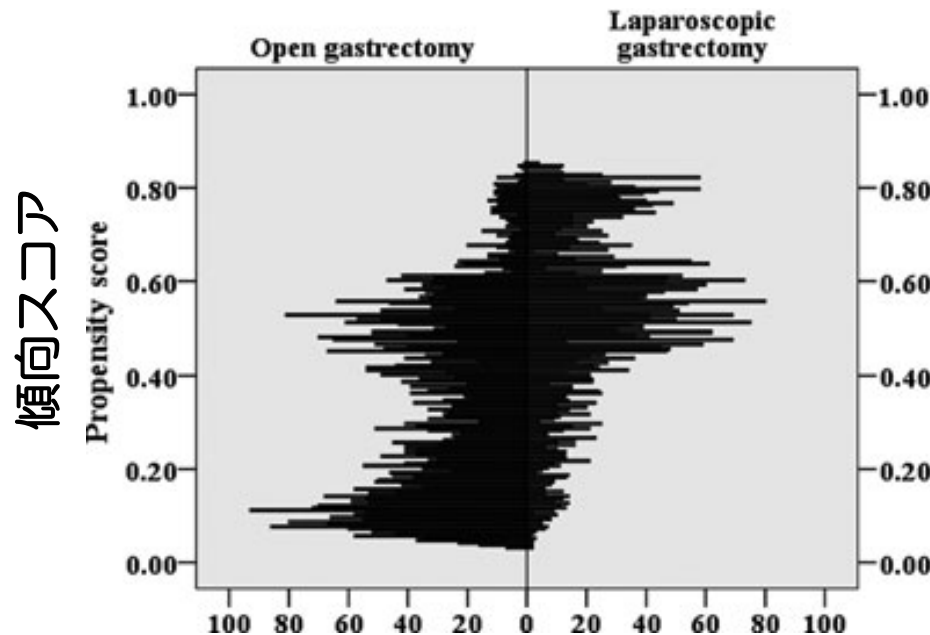
- 処理群と対照群の標本サイズに違いが大きいとき、強欲マッチングに比べて顕著に優れている。
- 標本サイズが大きい場合には、強欲マッチングとの差は小さい。

その他のマッチングの方法としては、遺伝的マッチング(Genetic matching)、フルマッチング(full matching)などがある。

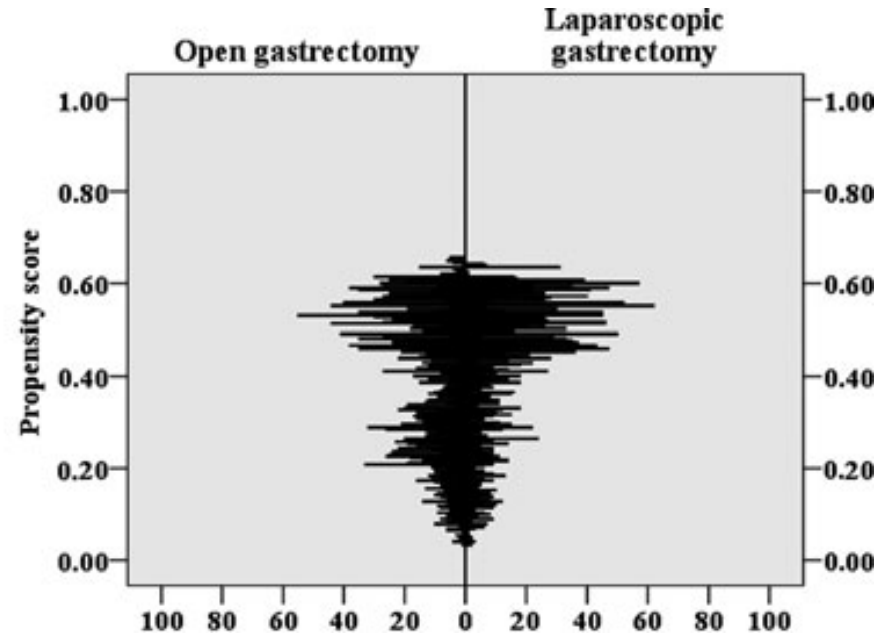
論文の記載

We performed a one-to-one matching analysis between laparoscopic and open distal gastrectomy groups on the basis of estimated propensity scores of each patient. この論文ではマッチング方法が未記載

マッチング前



マッチング後



マッチング後に共変量の分布が揃っていることがわかる

マッチング前

マッチング後

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	

マッチング後に各共変量のp値が有意ではない。

Short communication

Lung Cancer, 120, 88-90, 2018

Radiosurgery alone is associated with favorable outcomes for brain metastases from small-cell lung cancer

小細胞肺癌患者の脳転移に対する全脳照射療法(WBRT)と定位放射線手術(SRS)の比較

Tyler P. Robin^a, Bernard L. Jones^a, Arya Amini^b, Matthew Koshy^{c,d}, Laurie E. Gaspar^a, Arthur K. Liu^a, Sameer K. Nath^a, Brian D. Kavanagh^a, D. Ross Camidge^e, Chad G. Rusthoven^{a,*}

論文の記載

(前略) Propensity score matching (PSM) was also performed. Ten-to-one matching without replacement was completed using the nearest neighbor match on the logit of the propensity score for treatment approach with caliper width set to 0.05 times the standard deviation of the logit of the propensity score. (後略)

Variable	WBRT	SRS	p-value
Total patients	N=5,752	N=200	
Age			
<65	3,013 (52.4%)	102 (51.0%)	0.701
≥65	2,739 (47.6%)	98 (49.0%)	
Sex			
Male	2,926 (50.9%)	96 (48.0%)	0.425
Female	2,826 (49.1%)	104 (52.0%)	
Race/Ethnicity			
White (non-Hispanic)	5,030 (87.4%)	172 (86.0%)	0.008
Black	476 (8.3%)	14 (7.0%)	
Hispanic	116 (2.0%)	11 (5.5%)	
Other/unknown	130 (2.3%)	3 (1.5%)	
CDCC			
0	3,423 (59.5%)	125 (62.5%)	0.437
1	1,617 (28.1%)	48 (24.0%)	
2+	712 (12.4%)	27 (13.5%)	
Extracranial metastases			
No	3,069 (53.4%)	138 (69.0%)	<.001
Yes	2,683 (46.6%)	62 (31.0%)	

Supplemental table 1. Patient characteristics.

CDCC=Charlson/Deyo combined comorbidity score;

WBRT=whole brain radiation therapy; SRS=stereotactic radiosurgery.

Variable	WBRT	SRS	p-value
Total patients	N=1,930	N=193	
Age			
<65	960 (49.7%)	99 (51.3%)	0.680
≥65	970 (50.3%)	94 (48.7%)	
Sex			
Male	934 (48.4%)	92 (47.7%)	0.848
Female	996 (51.6%)	101 (52.3%)	
Race/Ethnicity			
Caucasian (non-Hispanic)	1,718 (89.0%)	172 (89.1%)	0.999
Black	141 (7.3%)	14 (7.3%)	
Hispanic	50 (2.6%)	5 (2.6%)	
Other/unknown	21 (1.1%)	2 (1.0%)	
CDCC			
0	1,232 (63.8%)	121 (62.7%)	0.952
1	437 (22.7%)	45 (23.3%)	
2	261 (13.5%)	27 (14.0%)	
Extracranial metastases			
No	1,344 (69.6%)	135 (69.9%)	0.929
Yes	586 (30.4%)	58 (30.1%)	

Supplemental table 2. Propensity score matched patient groups.

CDCC=Charlson/Deyo combined comorbidity score; WBRT=whole

brain radiation therapy; SRS=stereotactic radiosurgery.

マッチドペア解析

連続値

対応のあるt検定

2値応答

McNemar検定, 条件付きロジスティック回帰

生存時間

層別ログランク検定, 層別比例ハザードモデル

VS

独立2標本

連続値

2標本t検定

2値応答

カイ2乗検定, ロジスティック回帰

生存時間

ログランク検定, 比例ハザードモデル

マッチングを反映した(マッチドペア解析によって)解析を行うほうが良い(Austin, 2008).

— マッチングを反映させたほうが有意になりやすい(水澤, ICR-Web).

傾向スコアは両群における共変量の分布が等しいことを保証するものであり, 個体レベルまで近いとことまでは保証していない(Hill & Stuart, 2008; Ho et al., 2007, Schafer and Kang(2998), 岩崎(2015)).

— マッチドペアで取り扱うことは推奨されない.

結局のところ, 問題は解決していない(下川は独立2標本で解析している)

マッチング後の解析の例示

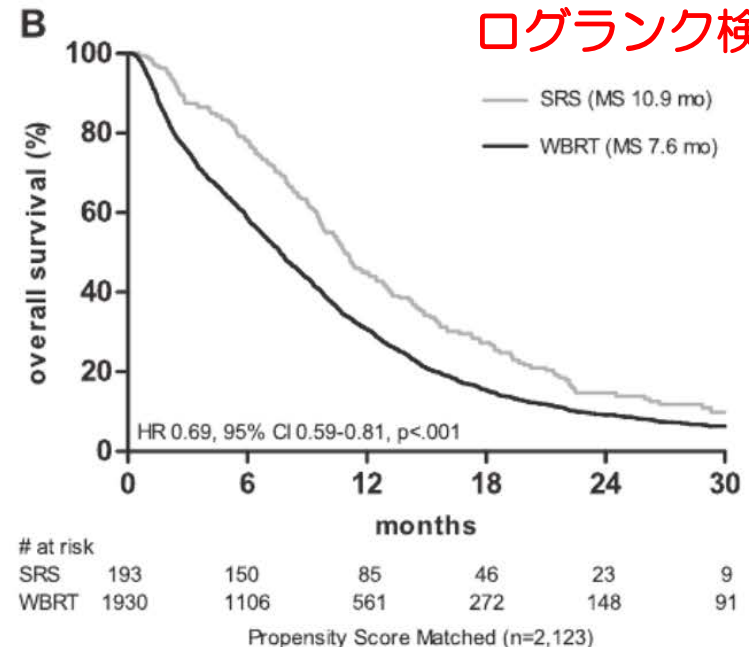
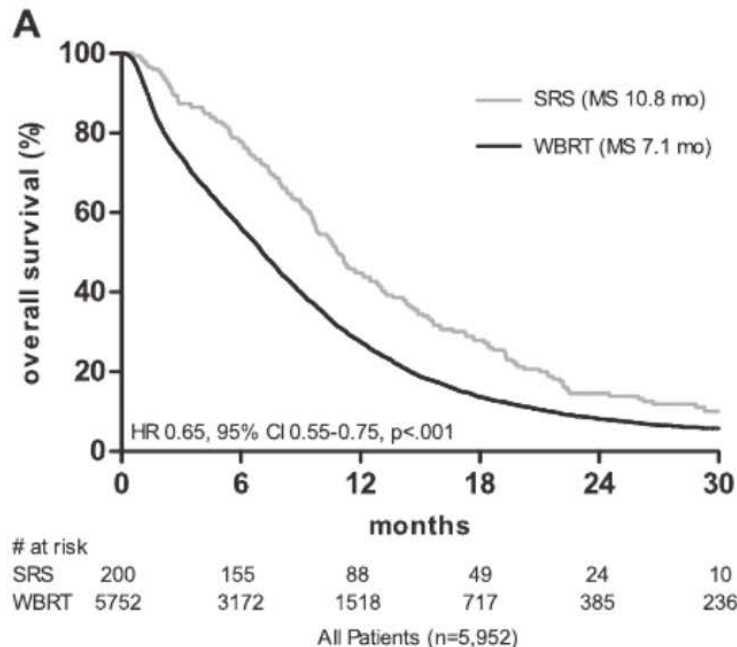
■ 早期胃がん患者に対する開腹切除術 vs. 腹腔鏡切除術の検討

Wilcoxon符号付順位検定

	Median (Interquartile Range)	<i>P</i>	Coefficient	95% Confidence Interval	<i>P</i>
Duration of anesthesia, min					
Laparoscopic	345 (289–415)	<0.001	90	85–96	<0.001
Open	262 (213–313)		Reference		
Postoperative length of stay, d					
Laparoscopic	13 (10–17)	<0.001	–2.0	–2.7 to –1.3	<0.001
Open	15 (12–20)		Reference		
Total costs, US \$					
Laparoscopic	21,510 (19,530–24,379)	0.002	560	210–909	0.002
Open	21,024 (18,917–25,139)		Reference		

■ 小細胞肺癌患者の脳転移に対する全脳照射療法(WBRT)と定位放射線手術(SRS)の比較

ログラंक検定



各共変量の分布のバランスの確認

群間のバランスをp値で評価するのは適切でない(Yang & Dalton, 2012)

標準化差スコア(Standardized difference score)の利用

■ 量的変数

$$d_{\text{cont}} = \frac{\bar{x}_1 - \bar{x}_0}{\sqrt{(s_1^2 + s_0^2)/2}} \quad \left[\begin{array}{ll} \bar{x}_1: \text{処理群における共変量の平均値} & \bar{x}_0: \text{対照群における共変量の平均値} \\ s_1^2: \text{処理群における共変量の分散} & s_0^2: \text{対照群における共変量の分散} \end{array} \right]$$

■ 2値変数

$$d_{\text{bin}} = \frac{\hat{p}_1 - \hat{p}_0}{\sqrt{(\hat{p}_1(1 - \hat{p}_1) + \hat{p}_0(1 - \hat{p}_0))/2}} \quad \left[\begin{array}{l} \hat{p}_1: \text{処理群における割合} \\ \hat{p}_0: \text{対照群における割合} \end{array} \right]$$

標準化差スコアの95%信頼区間は以下のように計算される

$$d \pm 1.96 \sqrt{\frac{n_1 + n_0}{n_1 \times n_0} + \frac{d^2}{2(n_1 + n_2)}}$$

J Vasc Surg Venous Lymphat Disord. 5(2),171-176. 2017

Systemic thrombolysis increases hemorrhagic stroke risk without survival benefit compared with catheter-directed intervention for the treatment of acute pulmonary embolism

Nathan L. Liang, MD, MS, Efthymios D. Avgerinos, MD, Michael J. Singh, MD, Michel S. Makaroun, MD, and Rabi A. Chaer, MD, MSc, Pittsburgh, Pa

急性肺血栓症の治療において、抗血栓療法はカテーテル療法に比べて、院内死亡率を改善するに至らない一方で、出血性脳卒中のリスクを増大させることを示したデータベース研究である。

マッチング前

	Total (N = 5727)	ST (n = 4272)	CDI (n = 1455)	P
Age, years	57.13 ± 16.7	57.41 ± 16.7	56.32 ± 16.7	.032
Female	2849 (49.8)	2195 (51.4)	654 (44.9)	<.001
Hypertension	2890 (50.5)	2198 (51.5)	692 (47.6)	.010
Congestive heart failure	776 (13.5)	621 (14.5)	155 (10.7)	<.001
Diabetes mellitus	1185 (20.7)	932 (21.8)	253 (17.4)	<.001
Chronic renal failure	516 (9.0)	421 (9.9)	95 (6.5)	<.001
Emphysema	1029 (18.0)	777 (18.2)	252 (17.3)	.46
Cancer	618 (10.8)	446 (10.4)	172 (11.8)	.14
Respiratory failure	1500 (26.2)	1192 (27.9)	308 (21.2)	<.001
Hypotension	953 (16.6)	779 (18.2)	174 (12.0)	<.001
Deep venous thrombosis	2198 (38.4)	1530 (35.8)	668 (45.9)	<.001
Vena cava filter placement	2158 (37.7)	1328 (31.1)	830 (57.0)	<.001

CDI, Catheter-directed intervention; ST, systemic thrombolysis.
All values are mean ± standard deviation or number (%).

P値は標本サイズに依存するため、大規模な観察研究ほどマッチング後も有意になりやすい。これに対して、標準化差スコアでは標本サイズに依存しない。標準化スコアの絶対値が0.1未満であればバランスがとれていると判断する。

マッチング後

	ST (n = 1430)	CDI (n = 1430)	P	Std diff
Age, years	55.80 ± 16.5	56.39 ± 16.7	.34	0.02
Female	647 (45.2)	641 (44.8)	.82	0.01
Hypertension	671 (46.9)	683 (47.8)	.65	0.04
Congestive heart failure	121 (8.5)	153 (10.7)	.042	0.03
Diabetes mellitus	227 (15.9)	247 (17.3)	.31	0.04
Chronic renal failure	88 (6.2)	94 (6.6)	.65	0.01
COPD	242 (16.9)	248 (17.3)	.77	0.01
Any cancer	170 (11.9)	170 (11.9)	1.00	0.01
Respiratory failure	293 (20.5)	302 (21.1)	.68	0.02
Hypotension	161 (11.3)	170 (11.9)	.60	0.07
Deep venous thrombosis	655 (45.8)	658 (46.0)	.91	0.01
IVC filter placement	804 (56.2)	819 (57.3)	.57	0.03

COPD, Chronic obstructive pulmonary disease; IVC, inferior vena cava.
All values are mean ± standard deviation or number (%). A standardized difference (Std diff) of < 0.1 suggests adequate variable balance after propensity matching.

A standardized difference (Std diff) of < 0.1 suggests adequate variable balance after propensity matching

■ 利点

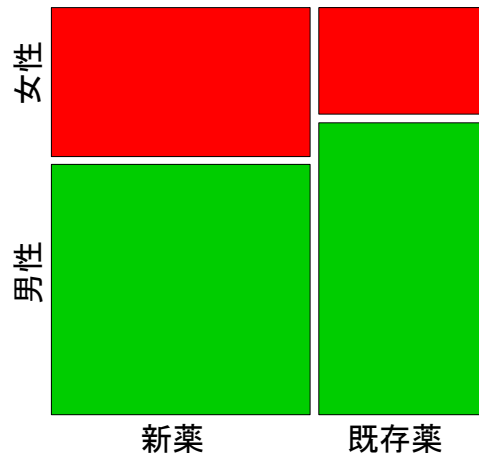
- (1) わかりやすく，統計の知識がなくても使える.
- (2) 他の手法に比べてバイアスが入りにくい.
- (3) マッチング後の解析が容易である.

■ 欠点

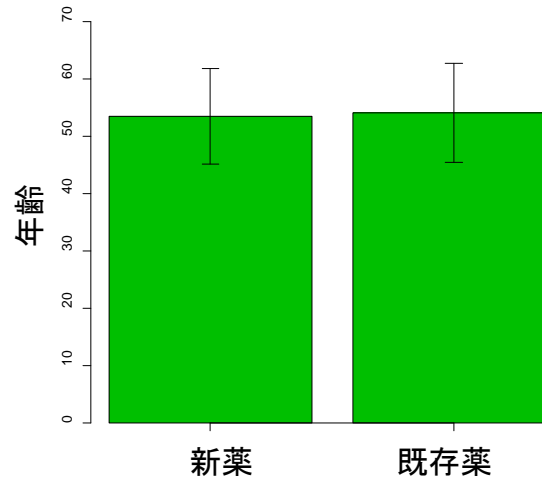
- (1) マッチングにコストがかかる．また，マッチングを行うと標本サイズが必ず少なくなる.
- (2) 標本サイズの減少に伴い，標準誤差(標準偏差を標本サイズの平方根で割ったもの)が増加するため，推定効率が下がる場合がある.
- (3) 他の方法(事後的な共分散分析や事後層別)に比べて必ずしも効率が良い方法ではない.

層別化のお話の前に：仮想例

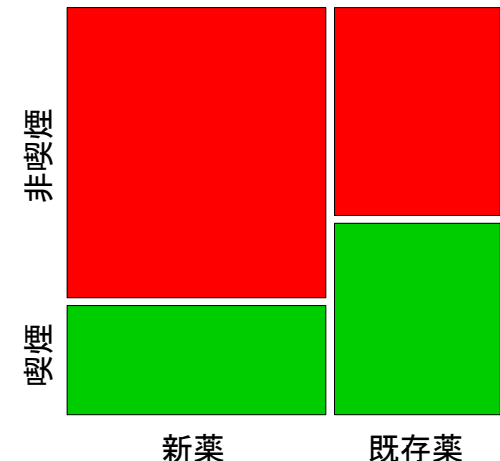
これは、新薬を投与した110例と既存薬を投与した71例を調査した後ろ向き研究のデータである。共変量としては、性別、年齢、喫煙の有無、BMI、重症度スコアが計測されている。



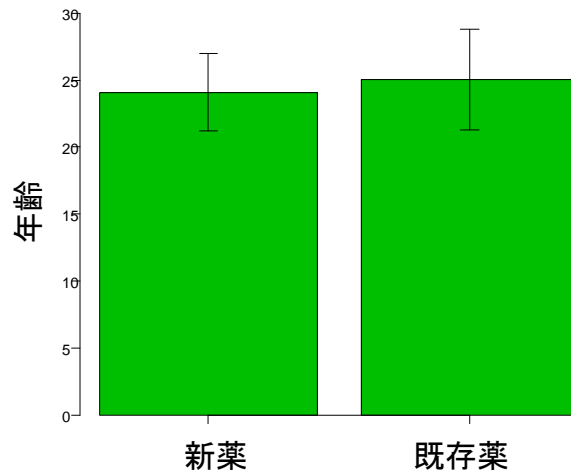
(a)性別 [p値=0.1500]



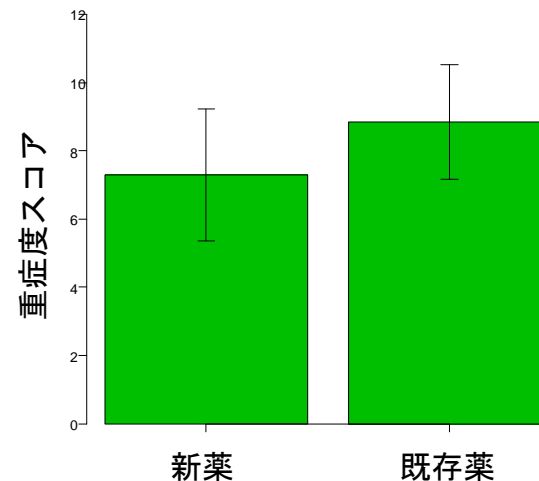
(b)年齢 [p値=0.6040]



(c)喫煙 [p値=0.0048]

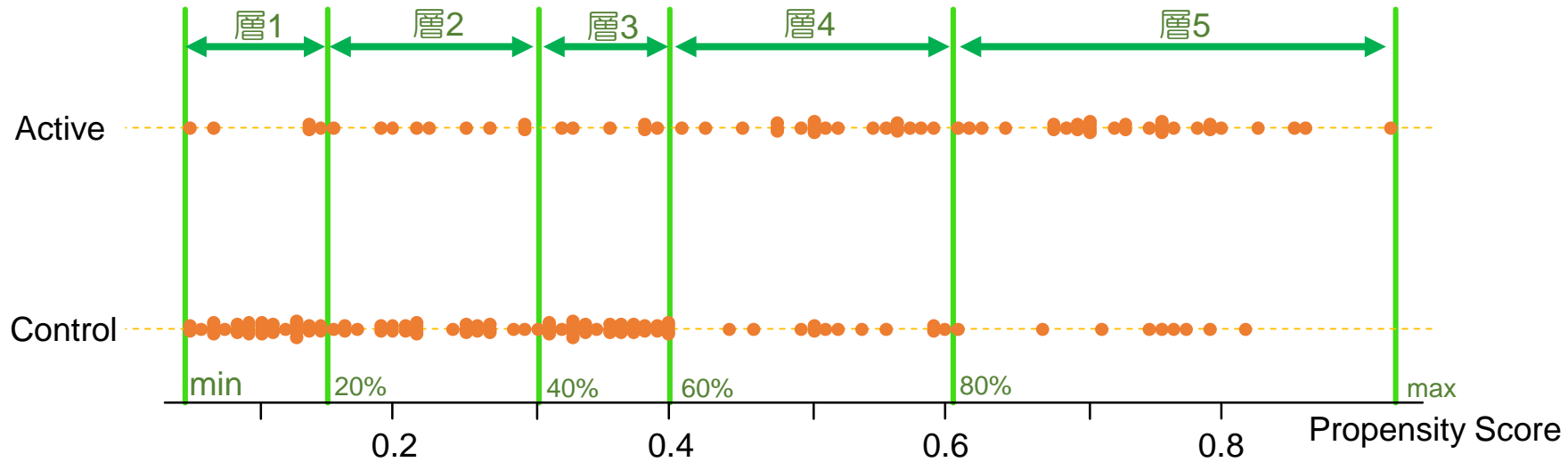


(d)BMI [p値=0.0632]



(e)重症度スコア [p値<0.0001]

傾向スコアで層に分ける(Stratify)ことにより，共変量の影響を除去する方法を傾向スコア層別化(propensity score stratification)という。



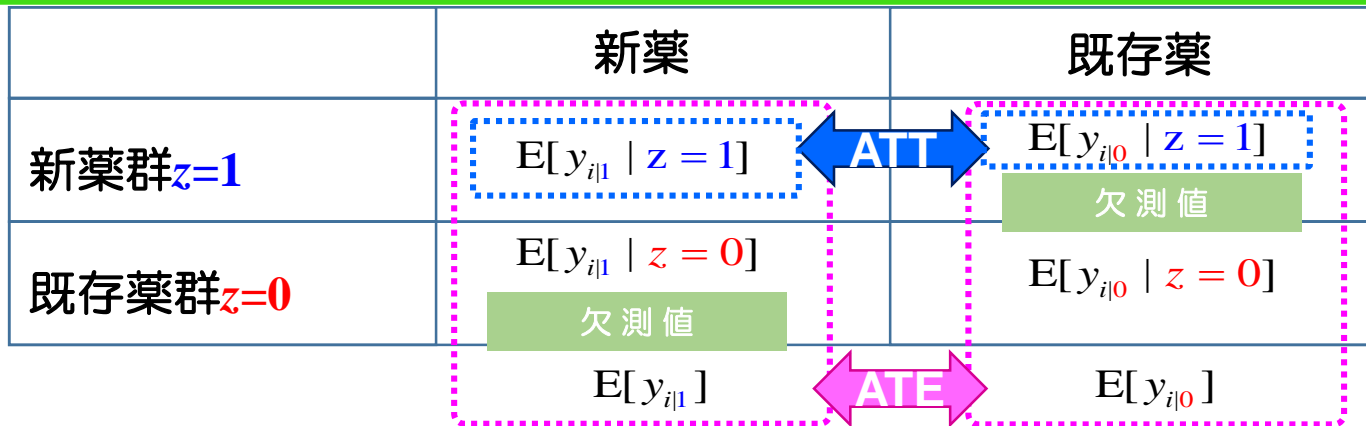
各層の症例数

	層1	層2	層3	層4	層5	合計
Active (新薬)	5	9	6	23	28	71
Control (既存薬)	32	27	30	13	8	110
合計	37	36	36	36	36	181



症例の割合で加重平均をとる方法が傾向スコア層別化である。

傾向スコア層別化：通常法



各層の症例数

	層1	層2	層3	層4	層5	合計
Active (新薬)	5	9	6	23	28	71
Control (既存薬)	32	27	30	13	8	110
合計	37	36	36	36	36	181

各層の有効率

Active (新薬)	0.40	0.56	0.17	0.39	0.46
Control (既存薬)	0.16	0.07	0.27	0.25	0.25
処理効果 (A - C)	0.24	0.48	-0.10	0.14	0.21

層 k の ATE での重み = $\frac{(\text{層 } k \text{ の症例数})}{\text{全症例数}}$

層 k の ATT での重み = $\frac{\text{層 } k \text{ の Active 群 での症例数}}{\text{Active 群 の全症例数}}$

ATE の重み	0.20	0.20	0.20	0.20	0.20
ATT の重み	0.07	0.13	0.08	0.32	0.39

$$\text{推定値} : \Delta = \sum_{\text{層の個数}} (\text{層}k\text{の重み}) \times (\text{層}k\text{の治療効果})$$

各層の有効率	Active (新薬)	0.40	0.56	0.17	0.39	0.46
	Control (既存薬)	0.16	0.07	0.27	0.25	0.25
	処理効果 (A - C)	0.24	0.48	-0.10	0.14	0.21
各層の重み	ATEの重み	0.20	0.20	0.20	0.20	0.20
	ATTの重み	0.07	0.13	0.08	0.32	0.39
処理効果 × ATEの重み		0.048	0.096	-0.020	0.028	0.042
処理効果 × ATTの重み		0.017	0.062	-0.008	0.045	0.082

■ 平均処理効果(ATE:Average Treatment Effect) **0.194**

■ 新薬群の平均処理効果(ATT: Average Treatment effect for Treated) **0.198**

調整なしの場合= 0.241 (新薬群：0.423, 既存薬群：0.182)

層別化解析における周辺平均重み付け

MMWS (Marginal Mean Weighting through Stratification; Hong, 2012)

医学統計セミナー：傾向スコア
2019.2.5

通常法では、層に対して群に関係なく同じ重みを設定する。これに対して、群毎に異なる重みを設定する方法が**MMWS**である。

	層1	層2	• • •	層S
処理群(z=1)	n_{11}	n_{21}	• • •	n_{S1}
対照群 (z=0)	n_{10}	n_{20}	• • •	n_{S0}
合 計	n_1	n_2	• • •	n_S

	割合
処理群(z=1)	p_1
対照群 (z=0)	p_0

ATE：層sの群zでの重み $w_{sz} = \frac{n_s \times p_z}{n_{sz}}$

ATT：層sの群zでの重み $w_{sz} = z + (1 - z) \frac{n_{1s} \times p_0}{n_{0s} \times p_1}$

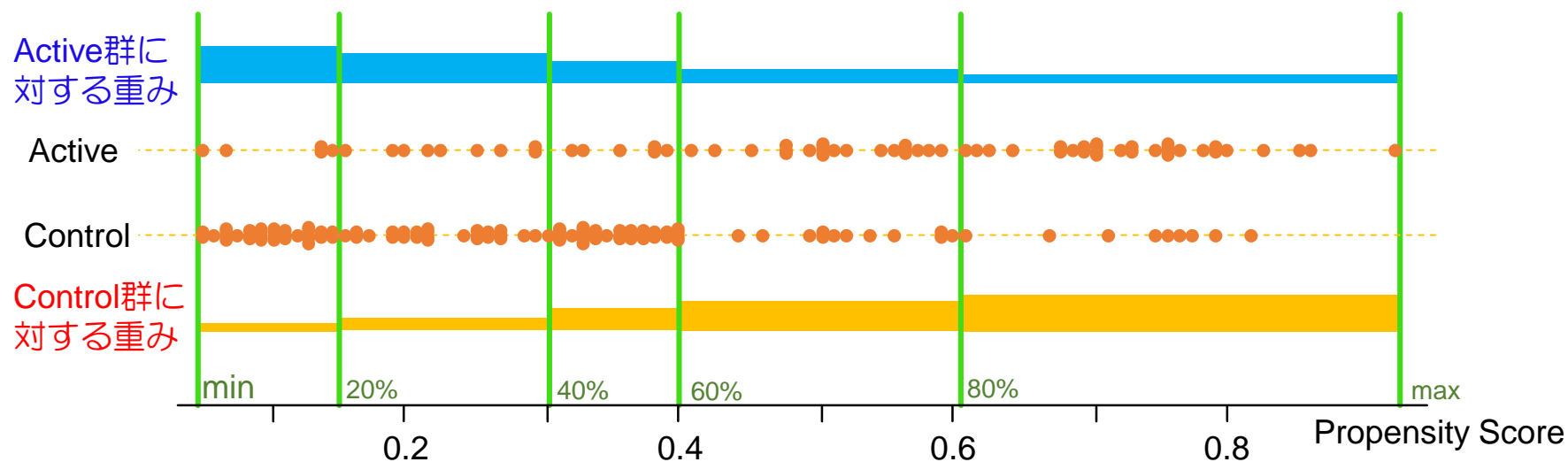
		層1	層2	層3	層4	層5	合計	割合
新 薬 (処理群)		5	9	6	23	28	71	0.392
既存薬 (対照群)		32	27	30	13	8	110	0.608
合 計		37	36	36	36	36	181	
ATE	新薬での重み	2.90	1.57	2.35	0.61	0.50		
	既存薬での重み	0.70	0.81	0.73	1.68	2.74		
ATT	新薬での重み	1.00	1.00	1.00	1.00	1.00		
	既存薬での重み	0.24	0.52	0.31	2.74	5.43		

層別化解析における周辺平均重み付け

MMWS (Marginal Mean Weighting through Stratification; Hong, 2012)

医学統計セミナー：傾向スコア
2019.2.5

		層1	層2	層3	層4	層5
新 薬 (処理群)		5	9	6	23	28
既存薬 (対照群)		32	27	30	13	8
合 計		37	36	36	36	36
ATE	新薬での重み	2.90	1.57	2.35	0.61	0.50
	既存薬での重み	0.70	0.81	0.73	1.68	2.74



MMWSは、IPWの代替として用いることができる(Linden, 2017).

Horovitz-Thompson型推定量

$$\Delta = \sum_{s=1}^{\text{層の個数}} \frac{(\text{層}sでのz=1の重み) \times (\text{層1での}z=1\text{のアウトカム})}{z=1の重みの総和} - \sum_{s=1}^{\text{層の個数}} \frac{(\text{層}sでのz=0の重み) \times (\text{層1での}z=0\text{のアウトカム})}{z=0の重みの総和}$$

		層1	層2	層3	層4	層5	合計
ATE	新薬での重み	2.90	1.57	2.35	0.61	0.50	7.94
	既存薬での重み	0.70	0.81	0.73	1.68	2.74	6.66
ATT	新薬での重み	1.00	1.00	1.00	1.00	1.00	5.00
	既存薬での重み	0.24	0.52	0.31	2.74	5.43	9.24

Active (新薬)	0.40	0.56	0.17	0.39	0.46
Control (既存薬)	0.16	0.07	0.27	0.25	0.25

ATE	新薬での有効率	0.15	0.11	0.05	0.03	0.03	0.37
	既存薬での有効率	0.02	0.01	0.03	0.06	0.10	0.22
ATT	新薬での有効率	0.08	0.11	0.03	0.08	0.09	0.40
	既存薬での有効率	0.00	0.00	0.01	0.07	0.15	0.24

■ 平均処理効果(ATE: Average Treatment Effect) : $\Delta_{ATE} = 0.15$

■ 新薬群の平均処理効果(ATT: Average Treatment effect for Treated) : $\Delta_{ATT} = 0.16$

Journal of Surgical Research, 210,204-212, 2017.

Ileostomy creation in colorectal cancer surgery: risk of acute kidney injury and chronic kidney disease

大腸がん患者における人工肛門造設術の急性腎疾患及び慢性腎臓病のリスクに関する観察研究

Linda Li, MD,^a Kelsey S. Lau, MD,^a Venkat Ramanathan, MD,^b
Sonia T. Orcutt, MD,^c Shubhada Sansgiry, PhD,^{b,d} Daniel Albo, MD, PhD,^a
David H. Berger, MD, MHCM,^{a,d} and Daniel A. Anaya, MD^{a,c,d,*}

^aMichael E. DeBakey Department of Surgery, Baylor College of Medicine, Houston, Texas

^bDepartment of Medicine, Baylor College of Medicine, Houston, Texas

^cDepartment of GI Oncology, H. Lee Moffitt Cancer Center & Research Institute, Tampa, Florida

^dCenter for Innovations in Quality, Effectiveness and Safety (IQES), Michael E. DeBakey VA Medical Center, Houston, Texas

To account for selection bias inherent to a retrospective study, a propensity score was calculated for each patient, which assessed the probability of ileostomy creation conditional on the observed covariates. The covariates used in building the propensity score were age, gender, race, marital status, BMI, Charlson comorbidity index, history of diabetes, baseline creatinine, cancer stage, and surgical approach. All patients were stratified into quintiles based on their corresponding propensity score and included in the analyses. We assessed the adequacy of the propensity score specification by comparing the standardized difference (with their corresponding P values) in baseline covariates after stratification, between those in the ileostomy and no ileostomy groups. A standardized mean difference of <0.1 or $P > 0.05$ was used to indicate no significant difference between baseline covariates.

共変量：年齢，性別，人種，婚姻状況，BMI，Charlson合併症指数，糖尿病歴，ベースラインでのクレアチニン，癌のステージ，外科的アプローチ

層別化：5層に層別化

人工肛門造設術による腎障害リスク

再入院を必要とする
急性腎障害 (OR)

Model 1A: Ileostomy vs. No Ileostomy
Unadjusted

9.5 (4.1-21.7)

PSなし

Model 1B: Ileostomy vs. No Ileostomy
*Adjusted for: socio-demographic characteristics,
pre-operative characteristics, tumor characteristics
and surgical approach.*

10.3 (3.9-27.2)

PS調整

術後12カ月での重度
腎疾患 (OR)

Model 2A: Ileostomy vs. No Ileostomy
Unadjusted

2.9 (1.2-7.0)

PSなし

Model 2B: Ileostomy vs. No Ileostomy
Adjusted using propensity score stratification

4.1 (1.4-11.9)

PS調整

重度腎疾患の経時的発
症リスク (HR)

Model 3A: Ileostomy vs. No Ileostomy*
Unadjusted

2.3 (1.3-4.1)

PSなし

Model 3B: Ileostomy vs. No Ileostomy*
Adjusted using propensity score stratification

4.2 (2.3-7.6)

PS調整

0 5 10 15 20 25 30
Ratio Measure (95% CI)

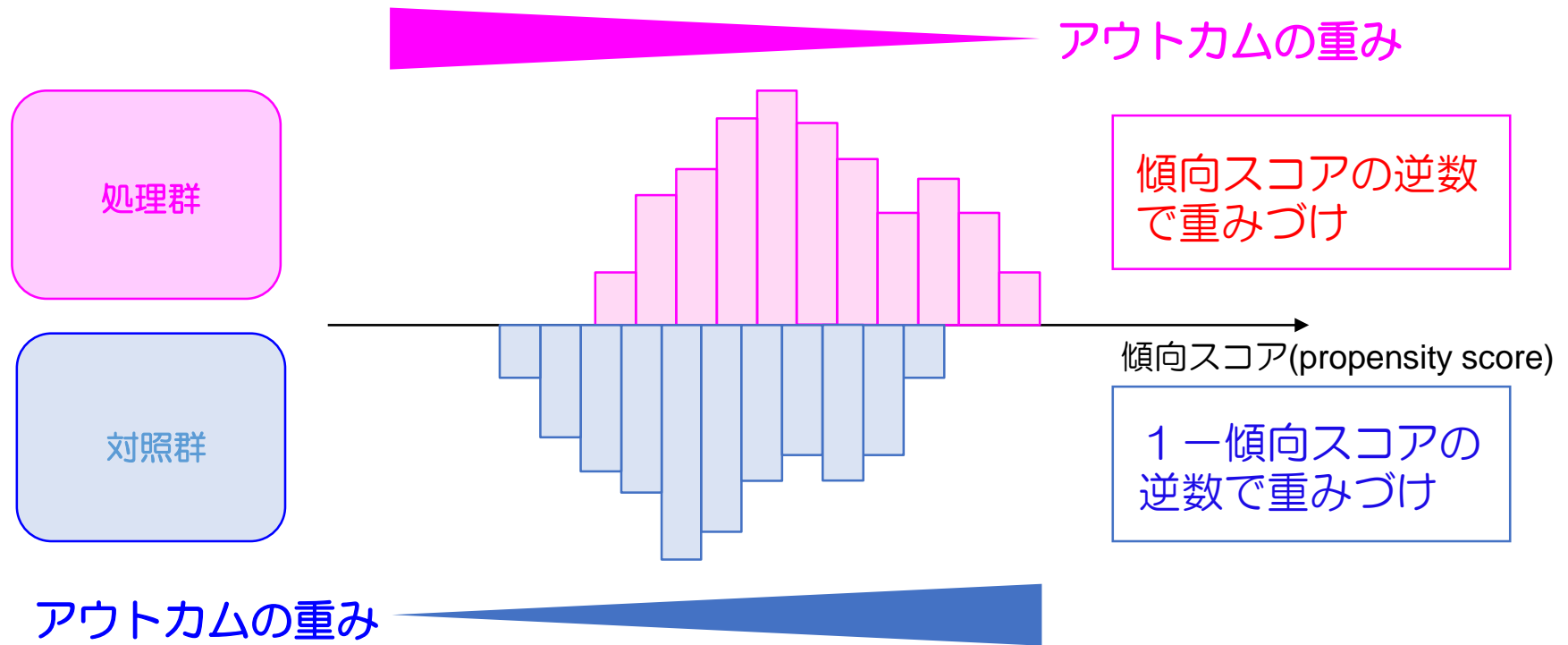
大腸癌手術における人工肛門造設術は、重要な手術であるものの、その後の腎障害への管理は重要であることを指摘している。

■ 層化の方法

- (1) 両群の観測値の標本サイズ $m+n$ をもとに各層のパーセンテージを定める.
 - 平均処理効果(ATE: Average Treatment Effect)の推定を想定する場合
- (2) 処理群の標本サイズ m をもとに各層のパーセンテージを決める.
 - 処理群の平均処理効果(ATT: Average Treatment effect for Treated)の推定を想定する場合
- (3) 対照群の標本サイズ n をもとに各層のパーセンテージを決める.
 - 対照群の平均処理効果(ATC: Average Treatment effect for Controlled)の推定を想定する場合

各層で両群の傾向スコアの分布の重なり具合が大きくなないと、適切な処理効果の推定ができない惧れがある (岩崎, 2015).

平均処理効果(ATE: Average Treatment Effect)のイメージ



逆数重み付け法の場合とは、傾向スコアにおいて処理群、対照群が選択された必然性が小さい個体の影響を強く、必然性が大きい個体の影響を弱くするように重みを付ける方法である。

逆確率重み付け(IPW: Inverse Probability Weighting)法をIPTW(Inverse Probability Treatment Weighting)と記載している文献もある。

■ 平均処理効果(ATE: Average Treatment Effect)

$$\text{被験者 } i \text{ の ATE での重み} = \frac{z_i}{e_i} + \frac{1 - z_i}{1 - e_i}$$

z_i : 1 のとき処理群, 0 のとき対照群
 e_i : 傾向スコア

処理群では $1/(\text{傾向スコア})$ の重み
対照群では $1/(1 - \text{傾向スコア})$ の重み

■ 暴露群の平均処理効果(ATT: Average Treatment effect for Treated)

$$\text{被験者 } i \text{ の ATE での重み} = z_i + (1 - z_i) \frac{e_i}{1 - e_i}$$

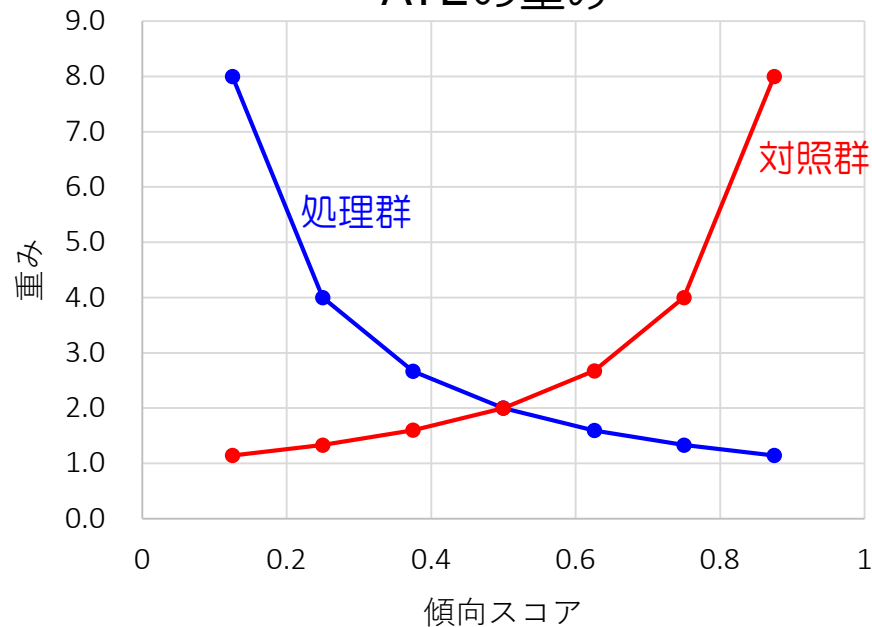
z_i : 1 のとき処理群, 0 のとき対照群
 e_i : 傾向スコア

処理群の重みは 1
対照群の重みは オッズ比

IPWにおける重み：数値例

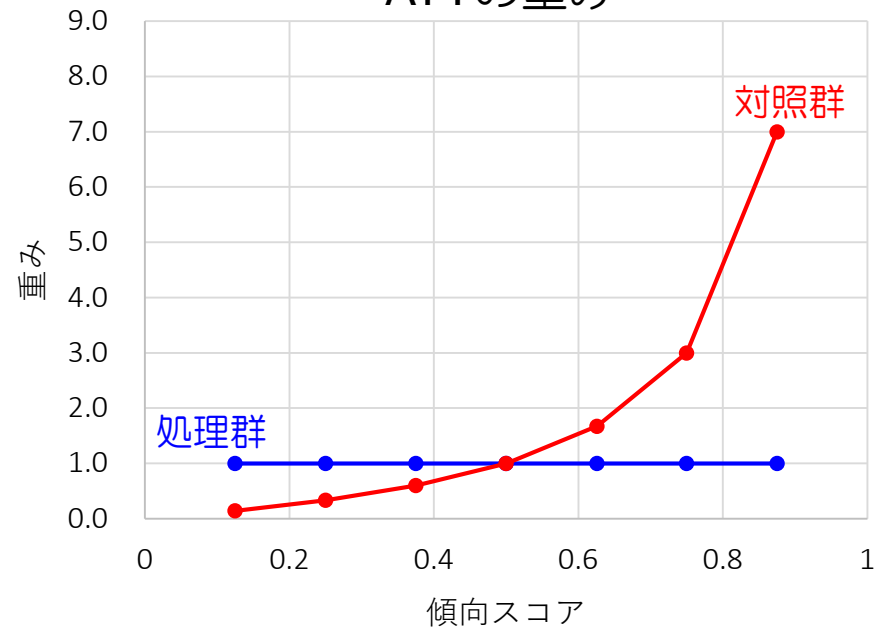
傾向スコア e	ATE		ATT	
	処理群	対照群	処理群	対照群
0.125	8.000	1.143	1.000	0.143
0.250	4.000	1.333	1.000	0.333
0.375	2.667	1.600	1.000	0.600
0.500	2.000	2.000	1.000	1.000
0.626	1.597	2.674	1.000	1.674
0.750	1.333	4.000	1.000	3.000
0.875	1.143	8.000	1.000	7.000

ATEの重み



← 対照群確率大 処理群確率大 →

ATTの重み



← 対照群確率大 処理群確率大 →

Horovitz-Thompson型推定量

$$\Delta = \frac{\sum_{i=1}^{\text{暴露群の被験者数}} (\text{被験者}i\text{の重み}) \times (\text{被験者}i\text{のアウトカム})}{\text{暴露群の重みの総和}} - \frac{\sum_{j=1}^{\text{非暴露群の個数}} (\text{被験者}j\text{の重み}) \times (\text{被験者}j\text{のアウトカム})}{\text{非暴露群の重みの総和}}$$

傾向スコア e	ATE		アウトカム		重み付きアウトカム	
	暴露群	非暴露群	暴露群	非暴露群	暴露群	非暴露群
0.125	8.000	1.143	35	10	13.50	0.55
0.250	4.000	1.333	40	15	7.71	0.96
0.375	2.667	1.600	45	20	5.79	1.54
0.500	2.000	2.000	50	25	4.82	2.41
0.626	1.597	2.674	55	30	4.24	3.87
0.750	1.333	4.000	60	35	3.86	6.75
0.875	1.143	8.000	65	40	3.58	15.42
			合計		43.5	31.5
			治療効果		12.0	

傾向スコアを用いない場合が50.0-25.0=25.0であることから、IPWでの重み付け推定量の場合のほうが、減少している。

IPWでは傾向スコアの逆数で重み付けを行う。そのため、傾向スコアが極端に小さい(あるいは大きい)場合には、重みが非常に大きくなる。たとえば、

(対照群に属する被験者のPS) = 0.999 → (ATEにおける重み)=1000

Leite (2017)は、この問題に対応する方策として、次を挙げている：

(1) 傾向スコアを計算するモデルの誤特定に起因していないか点検する。

(2) 傾向スコアの方法を変更する(例えば、層別化法など)

(3) 重みの打ち切り(weight truncation)を行う

Lee et al.(2011)は、任意のパーセント点をそれ以上の重みをもつ個体に割り当てることを提案しており、Gurel & Leite(2012)はその基準として99%点の適用を推奨している。

(4) 安定化重み(stabilized weights)を用いる

Harder et al.(2010)は、群毎に安定化を行うための定数を掛け合わせることで、上記の問題に対処している。

Brief Correspondence

European Urology, 71, 714-718, 2017

Efficacy of Systemic Chemotherapy Plus Radical Nephroureterectomy for Metastatic Upper Tract Urothelial Carcinoma

Thomas Seisen^a, Tarun Jindal^a, Patrick Karabon^a, Akshay Sood^a, Joaquim Bellmunt^b, Morgan Rouprêt^c, Jeffrey J. Leow^d, Malte W. Vetterlein^a, Maxine Sun^d, Shaheen Alanee^a, Toni K. Choueiri^b, Quoc-Dien Trinh^d, Mani Menon^a, Firas Abdollah^{a,*}

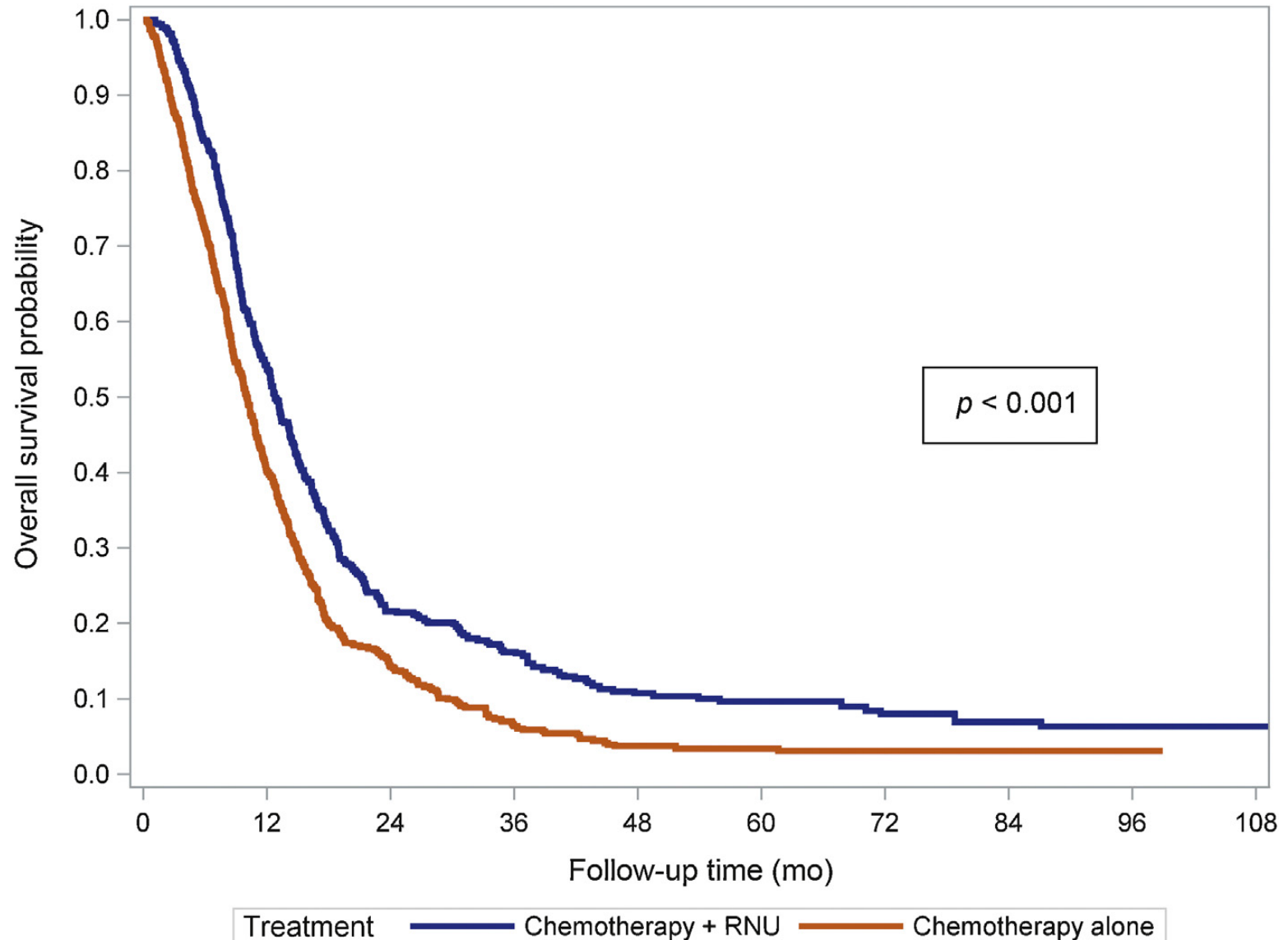
転移性上部尿路上皮癌に対する全身化学療法＋根治的腎盂腎切除術 vs. 全身化学療法の後ろ向き観察研究の結果である。

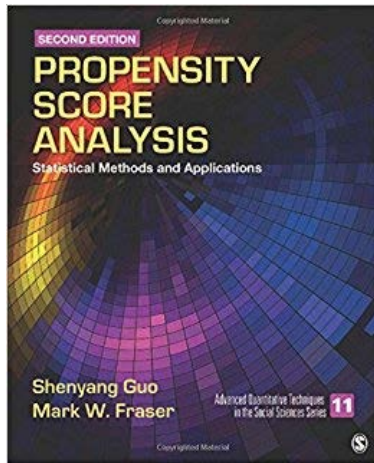
To account for potential selection bias, observed differences in baseline characteristics between patients who received chemotherapy plus RNU and those who received chemotherapy alone were controlled for with **an inverse probability of treatment weighting (IPTW)–adjusted analysis**. Balance in covariates between treatment groups before and after IPTW adjustment was assessed using the standardized difference approach. **IPTW-adjusted Kaplan-Meier curves and log-rank test were used to compare OS between patients who received chemotherapy plus RNU and those who received chemotherapy alone.**

IPTWに基づいて調整を行い、重み付きKaplan-Meier曲線及びログランク検定により評価している。

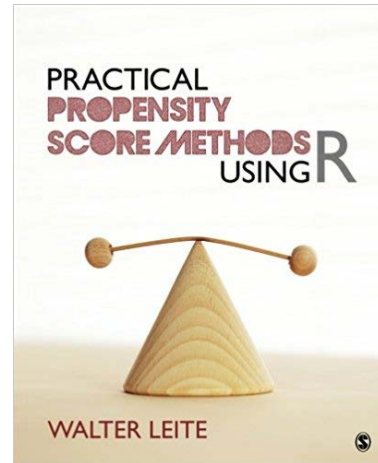
IP(T)Wを用いた文献例の結果

IP(T)Wにおける調整Kaplan-Meier推定量は、リスク集合を構成する被験者、及びイベントがあった被験者に関する重み付きによって計算される。

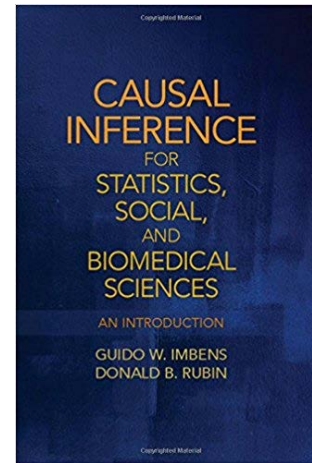




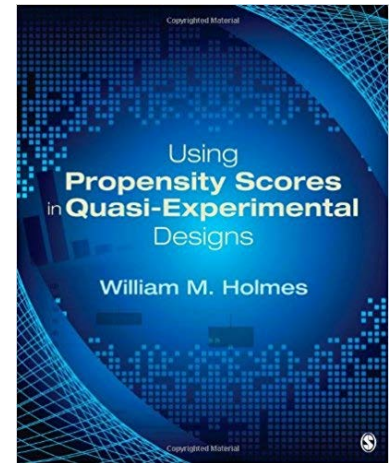
Guo, S. Fraster, M.W. SAGE Publications, 2014.



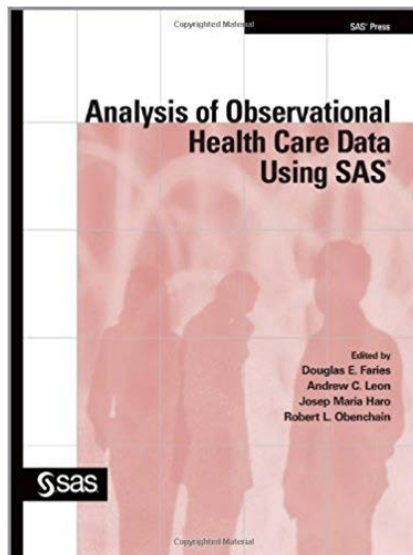
Leite, W. SAGE Publications, 2017.



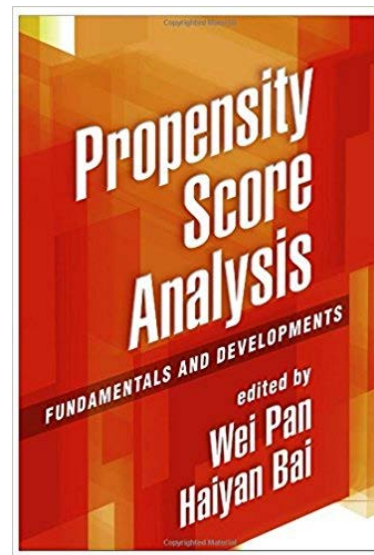
Imbens, G.W, Rubin, D.B. Cambridge Univ. Press., 2015.



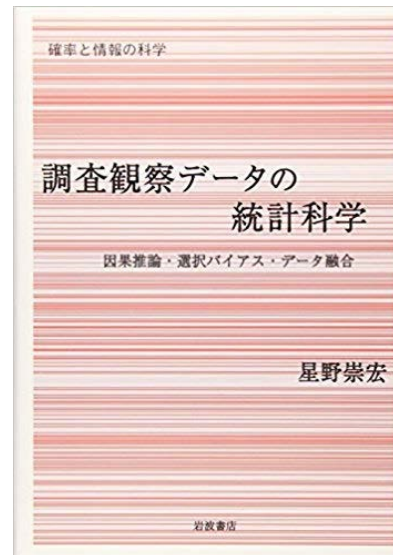
Holmes, W.M. SAGE Publications, 2013.



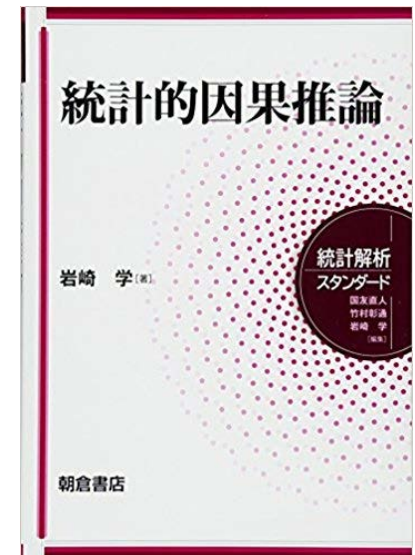
Faries, D. et al. SAS Inst., 2010



Pan, W., Bai, H. Guilford Pubn, 2015.



星野崇宏, 岩波書店, 2009.



岩崎学, 朝倉書店, 2015.

ご清聴ありがとうございます



shimokaw@wakayama-med.ac.jp

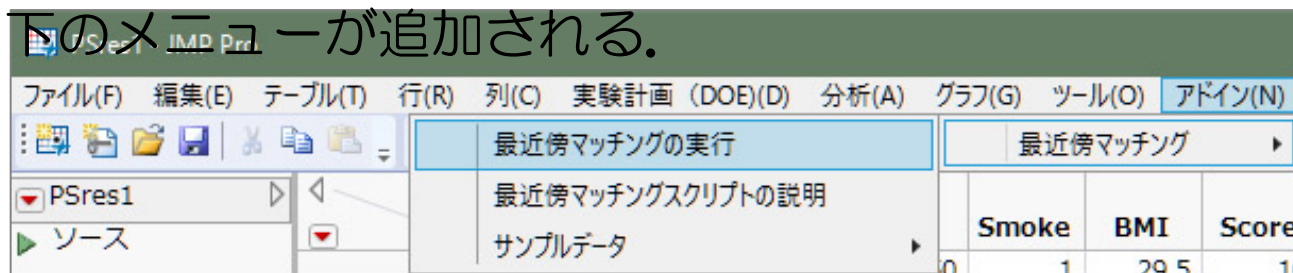
JMPでは、キャリパーを伴う1：1最近傍マッチングの方法をアドインとして公開している。

<http://www.jmp.com/content/dam/jmp/documents/jp/support/propensityaddin.zip>

[STEP.1] Propensity scoreを計算 (ここではロジスティック回帰)

- (1) 「分析」→「モデルのあてはめ」を選択
- (2) 「役割変数の選択」に『グループ変数(名義尺度)』，「モデル効果の構成」に共変量を選択する。
- (3) 「▼」から「確率の計算式の保存」を選択する(もとのデータセットに予測値が表示される)。

[STEP.2] 事前に入手したアドイン(上記参照)をインストールした場合、
以



[STEP.3] マッチングを行う.

- (1) 「アドイン」→「最近傍マッチングの実行」を選択
- (2) 「Y, 応答変数」に『アウトカム』, 「T, 処理変数」に『グループ変数』, 「マッチングに使うスコア」に『線形[1]』を選択.
- (3) 「ロジットを変換するか」のチェックを外す.
- (4) 「元データにペアの列を追加するか?」にチェックを入れる.
- (5) キャリパーを任意で設定(0~1であり, 推奨は0.2~0.25)

[STEP.4] 治療効果を推定する.

- (1) 「分析」→「二変量の関係」を選択
- (2) 「Y, 目的変数」に『アウトカム』, 「X, 説明変数」に『グループ変数』, 「By」に『マッチングした?』を選択する.

出力は, 「マッチングした」のみを通常の解析方法と同様に解釈する.

新薬(A)を投与した71例と既存薬B(C)を投与した101例の背景因子(性別, 年齢, 喫煙の有無, BMI, 重症度スコア)と治療効果を調査した後ろ向き研究のデータである。

Sex：性別(M：男性, F：女性),

Age：年齢,

Smoke：喫煙歴(1：有, 0：無),

BMI：Body Mass Index,

Score：重症度スコア,

group：薬剤(A:新薬, C：既存薬)

Outcome：アウトカム(1：改善, 0：非改善)

[STEP.1] 傾向スコアを計算する.

応答変数： group(薬剤)

説明変数： Sex(性別), Age(年齢), Smoke(喫煙歴), BMI, Score(重症度スコア)

* 連続変数の場合には, 2値化をしてはいけない.

[STEP.2] 傾向スコアを用いてマッチングする.

[STEP.3] マッチングしたデータを用いて解析を行う.

CSVファイルの読み込み

PSexample.csv

最初の列は変数名

	A	B	C	D	E	F	G
1	Sex	Age	Smoke	BMI	Score	group	Outcome
2	M	50	1	29.5	10	A	1
3	F	67	0	24	8	C	0
4	F	50	0	20.2	8	C	0

新しいデータセットを作成する(直接入力)
既存のデータセットを読み込む

データのインポート

パッケージに含まれるデータを読み込む
データセットを複製する
データセットの名前を変更する
2つのデータセットを結合する
アクティブデータセットを保存する

スクリプトファイルを開く
スクリプトを上書き保存する
スクリプトを名前を付けて保存する

出力を上書き保存する
出力を名前を付けて保存する

マークダウンファイルを開く
マークダウンファイルを上書き保存する
マークダウンファイルを名前を付けて保存する

Rワークスペースを読み込む
Rワークスペースを上書き保存
Rワークスペースを名前を付けて保存

作業フォルダーを変更する
終了

ファイルまたはクリップボード、URL からテキストデータを読み込む

SPSSのデータセットをインポート
Minitabのデータをインポート
Stataのデータをインポート
Excelのデータをインポート

ファイルまたはクリップボード

データセット名を入力：
Dataset

ファイル内に変数名あり：
☒

列数があわない場合に調整する：
☒

文字列の場合にも空欄は欠損値(NA)として読み込む：
☒

空欄以外に欠損値として読み込むべき記号：
NA

データファイルの場所

☒ ローカルファイルシステム
☐ クリップボード
☐ インターネットの URL

フィールドの区切り記号

☐ 空白
☒ カンマ
☐ タブ

その他 ☐ 指定：

小数点の記号

☒ ドリッド[.]
☐ カンマ[,]

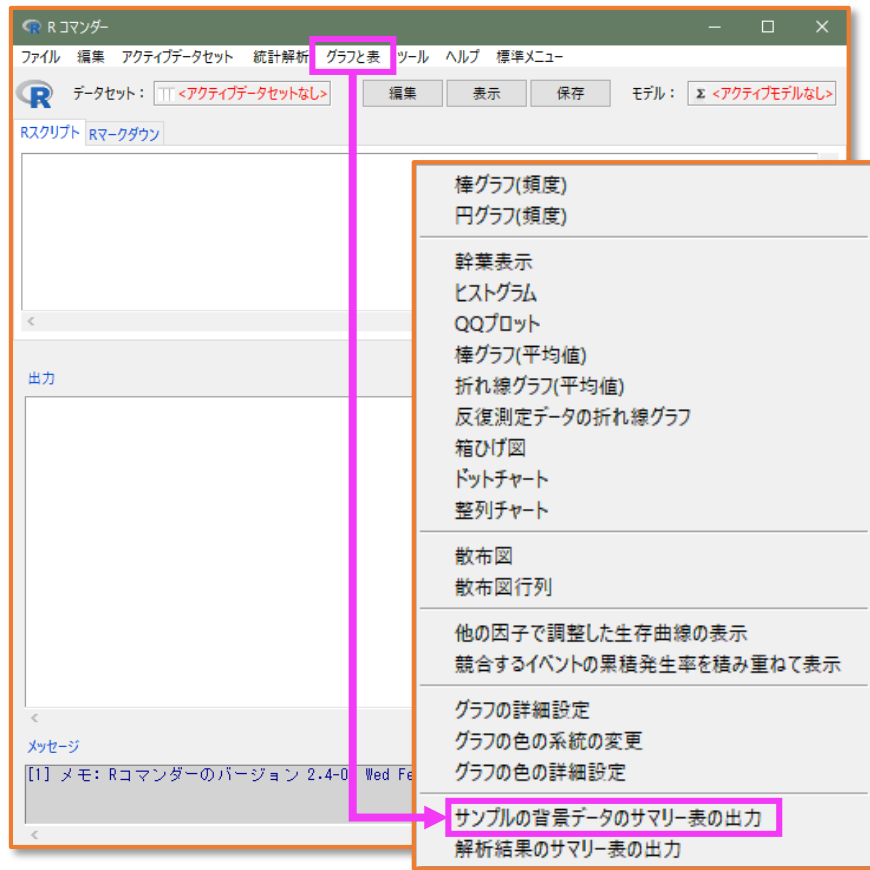
ヘルプ OK キャンセル

最初の列が変数名ならば、チェックを入れる。

NAは欠測

CSVファイルはカンマ区切り、
その他のテキストファイルは
notepadなどで確認する。

あとは、「PSexample.csv」を読み込めばよい。



グループ変数： group(薬剤)
 連続変数：年齢(Age), BMI, Score(重症度スコア)
 カテゴリカル変数： Sex(性別), Smoke(喫煙歴)


```
出力
> FinalTable <- FinalTable[which(rownames(FinalTable)!="n"),]
> FinalTable <- rbind(n=row1, FinalTable)
> FinalTable <- rbind(row0, FinalTable)
> row0 <- rep("", length(colnames(FinalTable)))
> row0[3] <- "group"
> FinalTable <- rbind(row0, FinalTable)
> finalTable_dataframe <- print(FinalTable)

  Factor  Group  group A  C  p.value
-----
n      |      | 71  | 110 |
-----
Sex (%) |      |      |      | 0.150
  F     |      | 19 (26.8) | 41 (37.3) |
  M     |      | 52 (73.2) | 69 (62.7) |
Smoke (%) | 0    | 37 (52.1) | 80 (72.7) | 0.007
        | 1    | 34 (47.9) | 30 (27.3) |
Age      |      | 54.11 (8.62) | 53.45 (8.32) | 0.604
BMI      |      | 25.03 (3.77) | 24.10 (2.92) | 0.063
Score    |      | 8.85 (1.68) | 7.30 (1.94) | <0.001

> write.table(FinalTable, "clipboard", sep=" ", row.names=FALSE, col.names=FALSE)
```

クリップボードのデータからExcelで作成

		group		p.value
		A 71	C 110	
Sex (%)	F	19 (26.8)	41 (37.3)	0.150
	M	52 (73.2)	69 (62.7)	
Smoke (%)	0	37 (52.1)	80 (72.7)	0.007
	1	34 (47.9)	30 (27.3)	
Age		54.11 (8.62)	53.45 (8.32)	0.604
BMI		25.03 (3.77)	24.10 (2.92)	0.063
Score		8.85 (1.68)	7.30 (1.94)	<0.001

喫煙歴、重症度スコアが有意である。したがって、この研究の被験者層において、これらの共変量に対して違いが認められる。

[STEP.1] 傾向スコアの計算

傾向スコアは、通常のロジスティック回帰によって求めることができる。

- 名義変数の解析 ▶
- 連続変数の解析 ▶
- ノンパラメトリック検定 ▶
- 生存期間の解析 ▶
- 検査の正確度の評価 ▶
- マッチドペア解析 ▶
- メタアナリシスとメタ回帰 ▶
- 必要サンプルサイズの計算 ▶

- 頻度分布
- 比率の信頼区間の計算
- 1標本の比率の検定
- 2群の比率の差の信頼区間の計算
- 2群の比率の比の信頼区間の計算
- 分割表の直接入力と解析
- 分割表の作成と群間の比率の比較(Fisherの正確検定)
- 対応のある比率の比較(二分割表の対称性の検定、McNemar検定)
- 対応のある3群以上の比率の比較(Cochran Q検定)
- 比率の傾向の検定(Cochran-Armitage検定)
- 二値変数に対する多変量解析(ロジスティック回帰)

二値変数に対する多変量解析(ロジスティック回帰)

モデル名を入力: GLM.1

変数 (ダブルクリックして式に入れる)

Age
BMI
group [因子]
Outcome
Score
Sex [因子]
Smoke

モデル式: + * : / %in% - ^ ()

目的変数 group ~ 説明変数 Age + BMI + Score + Sex + Smoke

群(group) 傾向スコアを求めるための共変量

☐ 3レベル以上の因子についてその因子全体のP値の計算(Wald検定)

☐ モデル解析用に解析結果をアクティブモデルとして残す

☐ ROC曲線を表示する

☐ 基本的診断プロットを表示する

☒ 傾向スコア変数を自動作成する

チェックを入れると傾向スコアの変数が追加される

☐ AICを用いたステップワイズの変数選択を行う。

☐ BICを用いたステップワイズの変数選択を行う。

☐ P値を用いたステップワイズの変数選択(減少法)を行う。

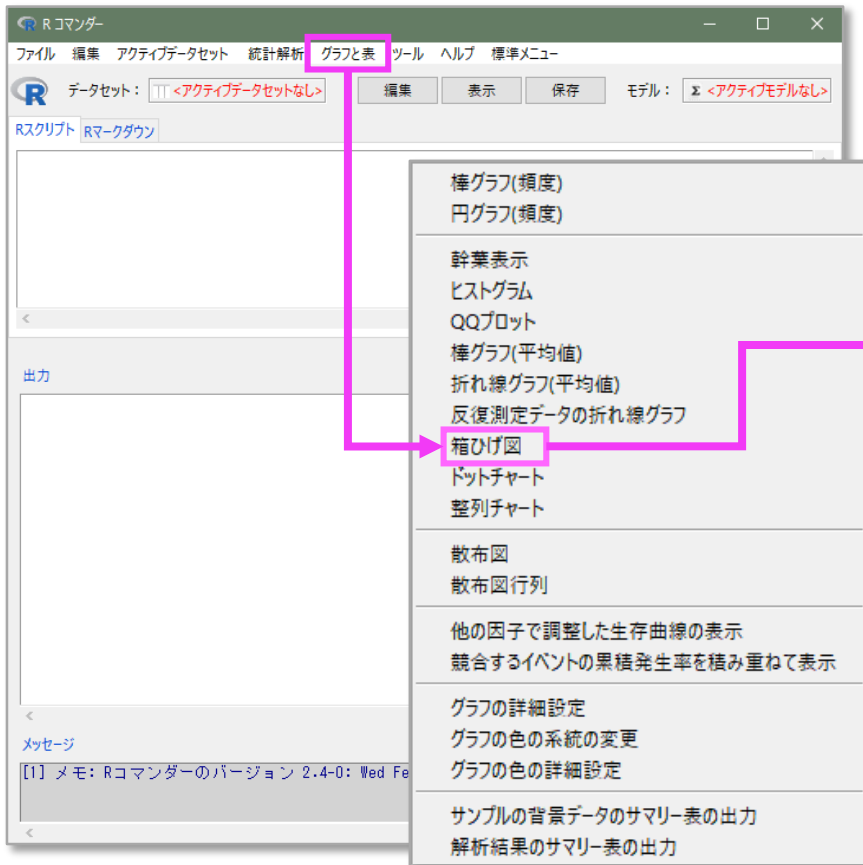
↓一部のサンプルだけを解析対象にする場合の条件式。例: age>50 & Sex==0 や age<50 | Sex==1

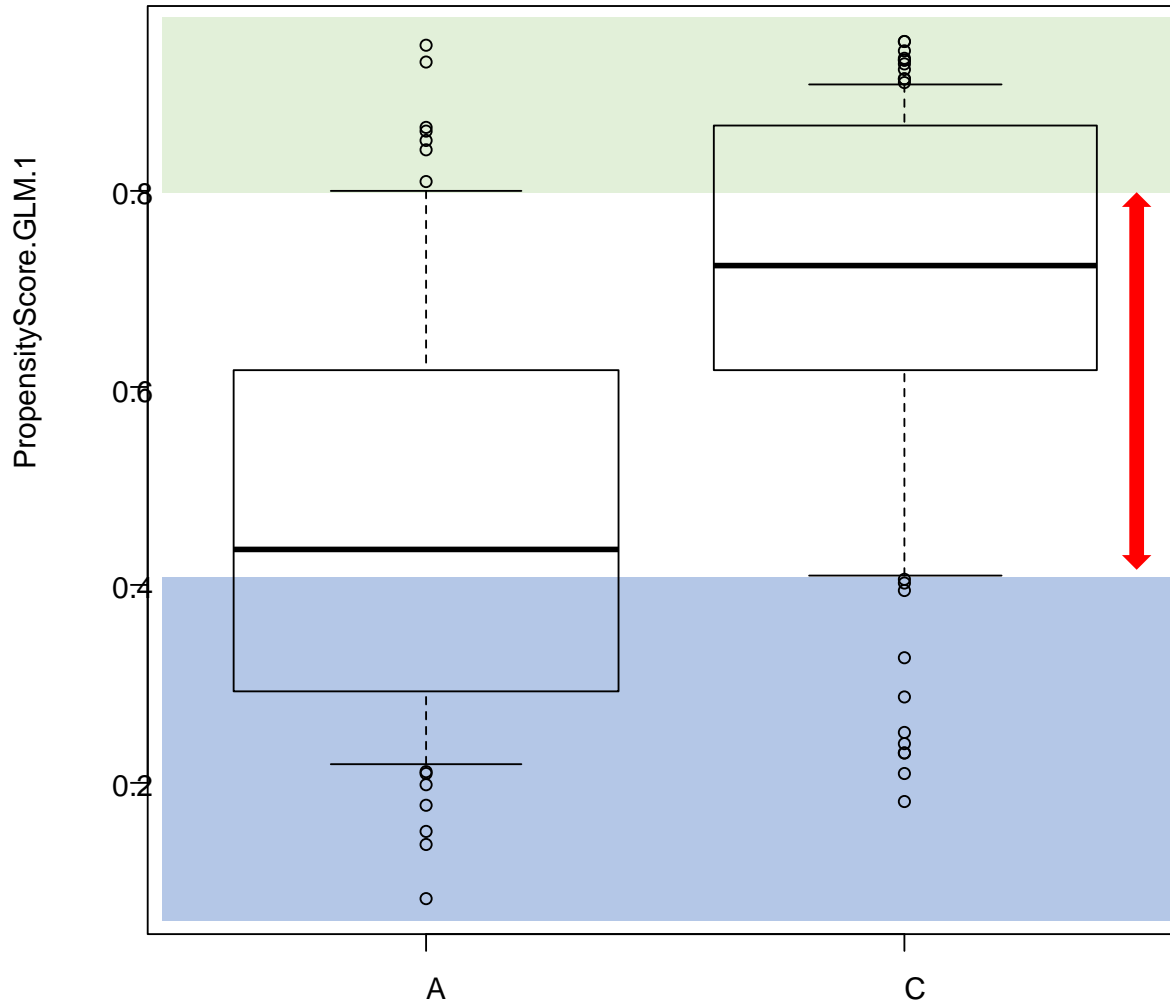
<全ての有効なケース>

ヘルプ リセット OK キャンセル 適用

新たな変数(傾向スコア)が追加される。

(余談) 傾向スコアをグラフ表示する





傾向スコアが小さい被験者は対照薬(C) を投与されているが新薬(A)を投与されている被験者は少ない。

新薬(A)と対照薬(C)のいずれも投与される可能性がある。

傾向スコアが小さい被験者は新薬(A)を投与されているが対照薬(C)を投与されている被験者は少ない。

卑近なマッチング

- 名義変数の解析
- 連続変数の解析
- ノンパラメトリック検定
- 生存期間の解析
- 検査の正確度の評価
- マッチドペア解析**
- メタアナリシスとメタ回帰
- 必要サンプルサイズの計算

マッチさせたコントロールの抽出

- マッチさせたサンプルの比率の比較(Mantel-Haenzel検定)
- マッチさせたサンプルの比率の多変量(条件付ロジスティック回帰)
- マッチさせたサンプルの生存率の多変量(層別化比例ハザード回帰)

キャリアパーを設定できないので推奨されない。

Rを応用する(追加パッケージmatchingを利用する)

matchingパッケージではグループ変数が0(コントロール群), 1(アクティブ群)で定義しなければならない(グループ変数groupはAとCになっている).

「アクティブデータセット」→「変数の操作」→「ダミー変数を作成する」



群(group)

新たに生成される変数

groupDummyGroupA

処理群(A)を1, 対照群(C)を0とした場合

groupDummy.GroupC

処理群(A)を0, 対照群(C)を1とした場合

Rを用いた解析

[STEP.1]

パッケージmatching をインストールする

これは1回やればよい
(毎回実施しなくてもいい)

The screenshot shows the R Commander interface. The command `install.packages("Matching")` is entered in the script area. Below the script area, a flowchart indicates the steps: '入力する' (Enter) leads to '入力したコマンドをドラックする' (Drag the entered command), which leads to '実行ボタンを押す' (Press the execution button). The execution button is labeled '実行' (Execute). Below the execution button, a box indicates '質問はすべてOK' (All questions are OK). To the left of this box, another box indicates 'Secure CRAN mirrorsのメニュー Japan(Tokyo)[https]を選択する' (Select the menu for Secure CRAN mirrors Japan(Tokyo)[https]).

```
graph LR; A[入力する] --> B[入力したコマンドをドラックする]; B --> C[実行ボタンを押す]; C --> D[実行]; D --> E[質問はすべてOK]; E --> F[Secure CRAN mirrorsのメニュー Japan(Tokyo)[https]を選択する];
```

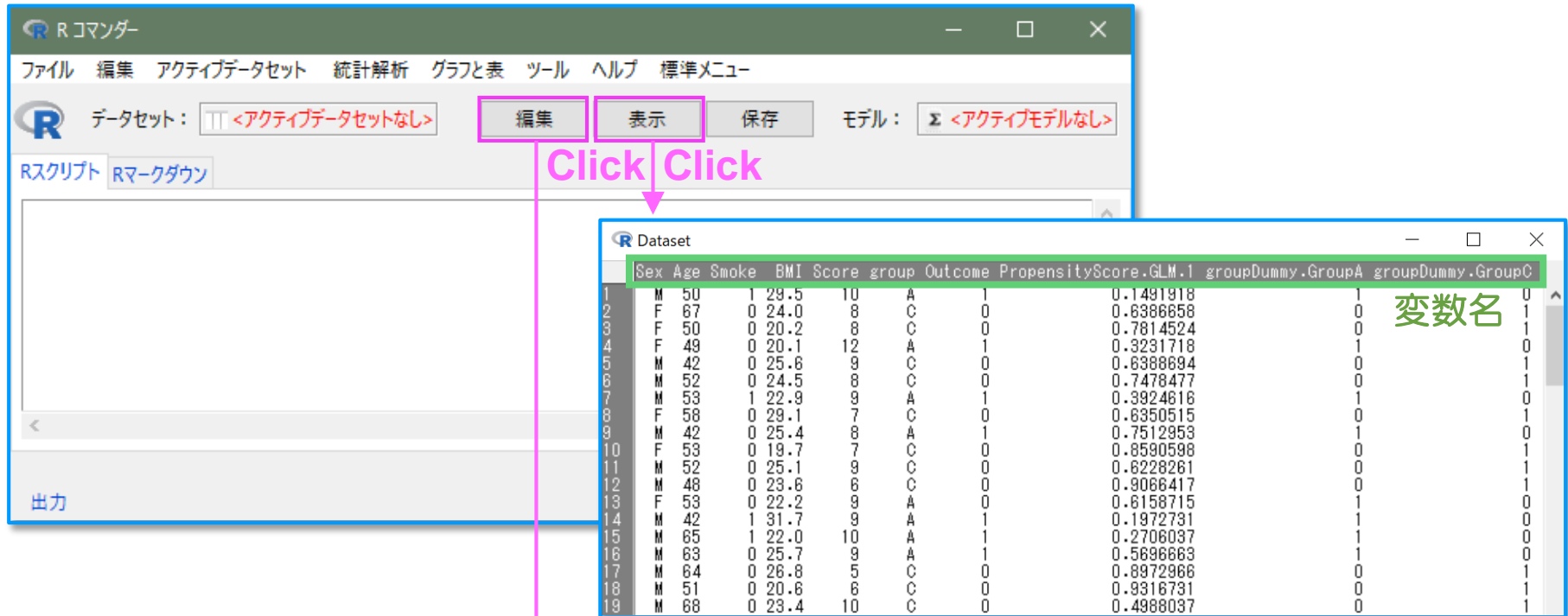
[STEP.2]

パッケージmatching を読み込む

The screenshot shows the R Commander interface. The command `library(Matching)` is entered in the script area. Below the script area, a flowchart indicates the steps: '入力する' (Enter) leads to '入力したコマンドをドラックする' (Drag the entered command), which leads to '実行ボタンを押す' (Press the execution button). The execution button is labeled '実行' (Execute).

```
graph LR; A[入力する] --> B[入力したコマンドをドラックする]; B --> C[実行ボタンを押す]; C --> D[実行];
```

(余談) 変数名を確認するには？



R コマンダー

ファイル 編集 アクティブデータセット 統計解析 グラフと表 ツール ヘルプ 標準メニュー

データセット: 編集 表示 保存 モデル:

Rスクリプト Rマークダウン

出力

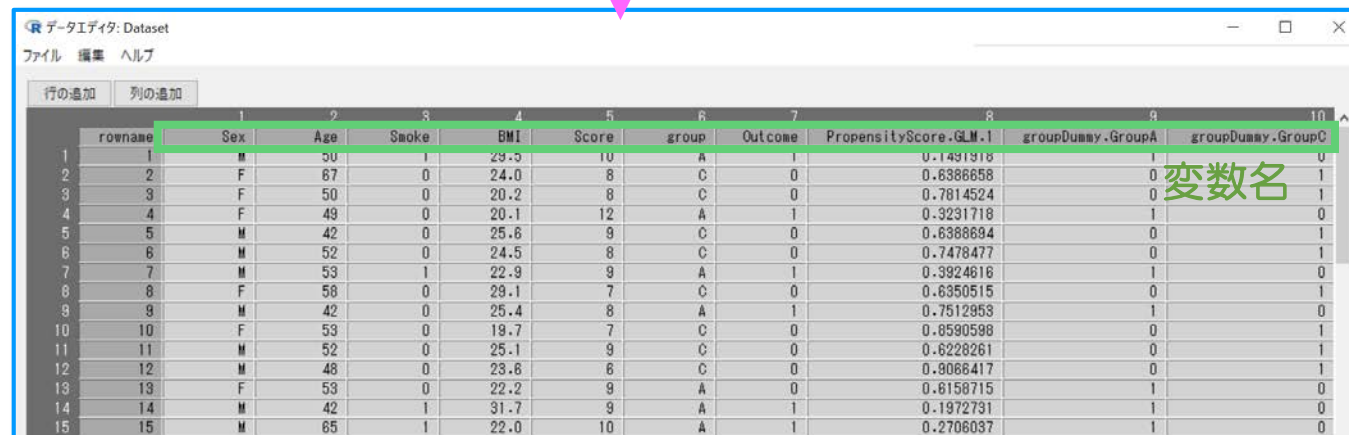
Click Click

R Dataset

	Sex	Age	Smoke	BMI	Score	group	Outcome	PropensityScore.GLM.1	groupDummy.GroupA	groupDummy.GroupC
1	M	50	1	29.5	10	A	1	0.1491918	1	0
2	F	67	0	24.0	8	C	0	0.6386658	0	1
3	F	50	0	20.2	8	C	0	0.7814524	0	1
4	F	49	0	20.1	12	A	1	0.3231718	1	0
5	M	42	0	25.6	9	C	0	0.6388694	0	1
6	M	52	0	24.5	8	C	0	0.7478477	0	1
7	M	53	1	22.9	9	A	1	0.3924616	1	0
8	F	58	0	29.1	7	C	0	0.6350515	0	1
9	M	42	0	25.4	8	A	1	0.7512953	1	0
10	F	53	0	19.7	7	C	0	0.8590598	0	1
11	M	52	0	25.1	9	C	0	0.6228261	0	1
12	M	48	0	23.6	6	C	0	0.9066417	0	1
13	F	53	0	22.2	9	A	0	0.6158715	1	0
14	M	42	1	31.7	9	A	1	0.1972731	1	0
15	M	65	1	22.0	10	A	1	0.2706037	1	0
16	M	63	0	25.7	9	A	1	0.5696663	1	0
17	M	64	0	26.8	5	C	0	0.8972966	0	1
18	M	51	0	20.6	6	C	0	0.9316731	0	1
19	M	68	0	23.4	10	C	0	0.4988037	0	1

変数名

利点：開いたまま解析できる, 欠点：データを編集できない



R データエディタ: Dataset

ファイル 編集 ヘルプ

行の追加 列の追加

	1	2	3	4	5	6	7	8	9	10
rowname	Sex	Age	Smoke	BMI	Score	group	Outcome	PropensityScore.GLM.1	groupDummy.GroupA	groupDummy.GroupC
1	M	50	1	29.5	10	A	1	0.1491918	1	0
2	F	67	0	24.0	8	C	0	0.6386658	0	1
3	F	50	0	20.2	8	C	0	0.7814524	0	1
4	F	49	0	20.1	12	A	1	0.3231718	1	0
5	M	42	0	25.6	9	C	0	0.6388694	0	1
6	M	52	0	24.5	8	C	0	0.7478477	0	1
7	M	53	1	22.9	9	A	1	0.3924616	1	0
8	F	58	0	29.1	7	C	0	0.6350515	0	1
9	M	42	0	25.4	8	A	1	0.7512953	1	0
10	F	53	0	19.7	7	C	0	0.8590598	0	1
11	M	52	0	25.1	9	C	0	0.6228261	0	1
12	M	48	0	23.6	6	C	0	0.9066417	0	1
13	F	53	0	22.2	9	A	0	0.6158715	1	0
14	M	42	1	31.7	9	A	1	0.1972731	1	0
15	M	65	1	22.0	10	A	1	0.2706037	1	0

変数名

利点：データを編集できる, 欠点：開いたまま解析できない

Rによる傾向スコアの計算（続き）

パッケージmatchingにおけるマッチングの関数

```
Match(Y=応答, Tr=群, X=傾向スコア, caliper=キャリパー, ties=F, replace=F)
```

Outcome

groupDummyGroupA

試験薬(A)が1, コントロール
群(C)が0のダミー変数

PropensityScore.GLM.1

0.20~0.25の間で自由に設
定

傾向スコアの値

[STEP.3] マッチングを行う

Rスクリプト Rマークダウン

```
attach(Dataset)
Matching <- Match(Y=Outcome, Tr=groupDummy.GroupA, X=PropensityScore.GLM.1,
caliper=0.25, ties=F, replace=F)
detach(Dataset)
summary(Matching)
```

入力する

入力したコマンドを
ドラックする

実行ボタンを押す

出力



出力

実行

```
> attach(Dataset)
> Matching <- Match(Y=Outcome, Tr=groupDummy.GroupA, X=PropensityScore.GLM.1, caliper=0.25,ties=F, replace=F)
> Detach(Dataset)
> attach(Dataset)
> Matching <- Match(Y=Outcome, Tr=groupDummy.GroupA, X=PropensityScore.GLM.1, caliper=0.25,ties=F, replace=F)
> detach(Dataset)
> summary(Matching)

Estimate... 0.2381
SE..... 0.11071
T-stat..... 2.1507
p.val..... 0.031499

Original number of observations..... 181
Original number of treated obs..... 71
Matched number of observations..... 42
Matched number of observations (unweighted). 42

Caliper (SDs)..... 0.25
Number of obs dropped by 'exact' or 'caliper' 29
```

被験者総数

試験群の例数

マッチング後の方群当たりの例数

Rによる傾向スコアの計算（続き）

[STEP.4] マッチング後のデータを抽出する

Rスクリプト Rマークダウン

```
treat <- Dataset[Matching$index.treated,]  
control <- Dataset[Matching$index.control,]  
Match.data <- rbind(treat,control)
```

入力する → 入力したコマンドをドラックする → 実行ボタンを押す

出力

実行

マッチングされたデータが新たなデータ集合Match.dataに保存される。

R コマンドー

ファイル 編集 アクティブデータセット 統計解析 ク

データセット: Dataset 編集

データセットの選択

データセット(1つ選択)

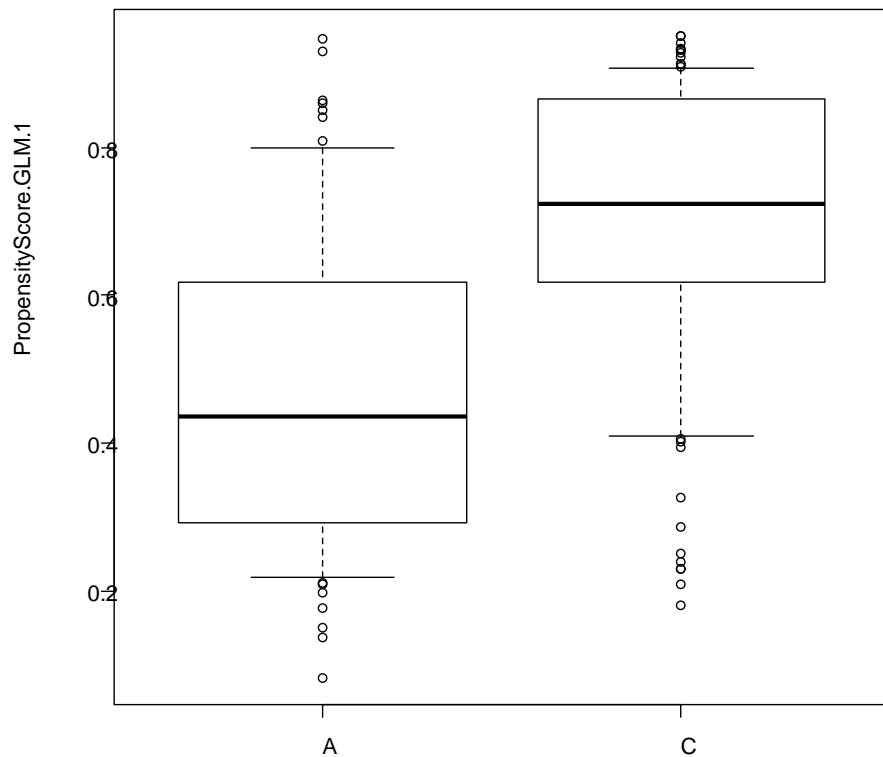
- control
- Dataset
- Match.data
- odds
- TempDF
- treat

OK キャンセル

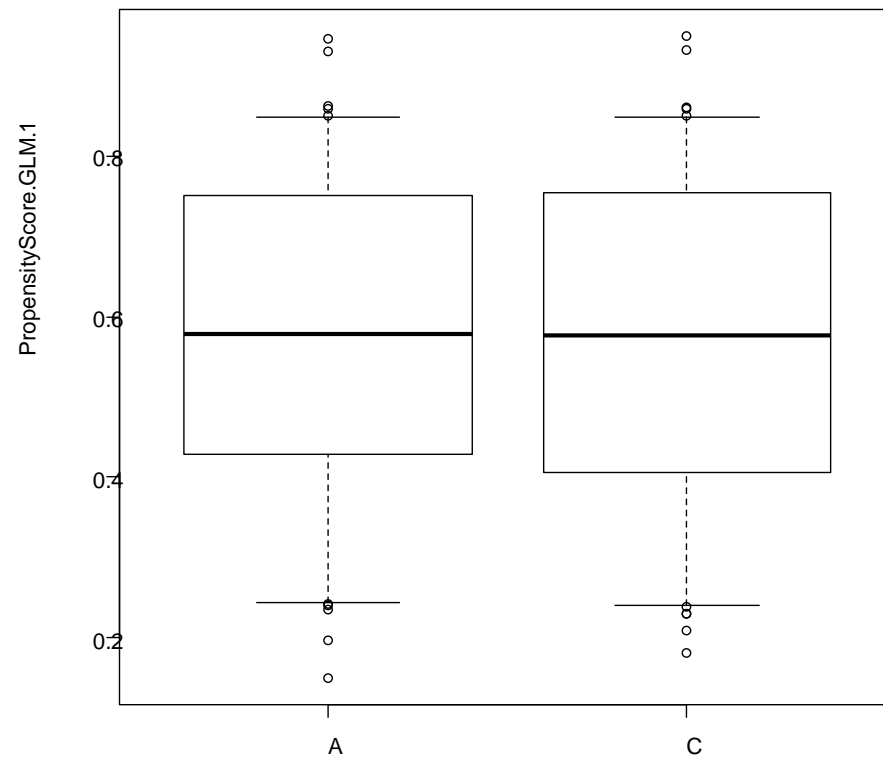
Match.dataを選択してOKボタンを押す

マッチング後の傾向スコアの分布を箱ひげ図で確認する

マッチング前



マッチング後



明らかにマッチング後では分布の違いが認められなくなった。

共変量毎に確認する.

		group		p.value
		A 71	C 110	
Sex (%)	F	19 (26.8)	41 (37.3)	0.150
	M	52 (73.2)	69 (62.7)	
Smoke (%)	0	37 (52.1)	80 (72.7)	0.007
	1	34 (47.9)	30 (27.3)	
Age		54.11 (8.62)	53.45 (8.32)	0.604
BMI		25.03 (3.77)	24.10 (2.92)	0.063
Score		8.85 (1.68)	7.30 (1.94)	<0.001

		group		p.value
		A 71	C 110	
Sex (%)	F	14 (33.3)	16 (38.1)	0.820
	M	28 (66.7)	26 (61.9)	
Smoke (%)	0	27 (64.3)	26 (61.9)	1.000
	1	15 (35.7)	16 (38.1)	
Age		54.83 (9.38)	53.52 (7.84)	0.490
BMI		24.61 (2.96)	24.14 (3.16)	0.482
Score		8.26 (1.65)	8.31 (1.94)	0.904

分布間の差が認められなくなった

名義変数の解析	▶
連続変数の解析	▶
ノンパラトリック検定	▶
生存期間の解析	▶
検査の正確度の評価	▶
マッチドペア解析	▶
メタアナリシスとメタ回帰	▶
必要サンプルサイズの計算	▶

分割表の直接入力と解析
分割表の作成と群間の比率の比較(Fisherの正確検定)
対応のある比率の比較(二分割表の対称性の検定、McNemar検定)
対応のある3群以上の比率の比較(Cochran Q検定)
比率の傾向の検定(Cochran-Armitage検定)
二値変数に対する多変量解析(ロジスティック回帰)

R 分割表の作成と群間の比率の比較(Fisherの正確検定)

↓ 複数の選択はCtrlキーを押しながらクリック。

行の選択 (1つ以上選択)

列の変数 (1つ選択)

Age
BMI
group
groupDummy.GroupA
groupDummy.GroupC
Outcome
PropensityScore.GLM.1
Score
Sex
Smoke

Age
BMI
group
groupDummy.GroupA
groupDummy.GroupC
Outcome
PropensityScore.GLM.1
Score
Sex
Smoke

パーセントの計算

☒ 行のパーセント

☐ 列のパーセント

☐ 総計のパーセント

☐ パーセント表示無し

仮説検定

☐ カイ2乗検定

☐ カイ2乗統計量の要素

☐ 期待度数の表示

☒ フィッシャーの正確検定

カイ2乗検定の連続性補正

☒ Yes

☐ No

↓ 2群ずつの比較(post-hoc検定)は比

☐ 2群ずつの比較(Bonferroniの多重比較)

☐ 2群ずつの比較(Holmの多重比較)

↓ 一部のサンプルだけを解析対象にする場合の条件式。例: age>50 & Sex==0 や age<50 | Sex==1

<全ての有効なケース>

ヘルプ リセット OK キャンセル 適用

アウトプット

```
> rowPercents(.Table) # 行のパーセント表示
```

```
      Outcome
group    0    1 Total Count
  A 50.0 50.0   100     42
  C 73.8 26.2   100     42
```

有効割合

```
> fisher.test(.Table)
```

Fisher's Exact Test for Count Data

```
data: .Table
p-value = 0.04238
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.1272557 0.9693956
sample estimates:
odds ratio
0.3594044
```

C/Aのオッズ比

(A/Cのオッズ比は逆数を取ればよい)

```
> res <- NULL
> res <- fisher.test(.Table)
> Fisher.summary.table <- rbind(Fisher.summary.table, summary.table.twoway(table=.Table, res=res))
> colnames(Fisher.summary.table)[length(Fisher.summary.table)] <- gettextRcmdr(
+   colnames(Fisher.summary.table)[length(Fisher.summary.table)])
```

```
> Fisher.summary.table
```

```
      Outcome=0 Outcome=1 Fisher検定のP値
group=A         21       21          0.0424
group=C         31       11
```

新薬と既存薬で有効割合に有意な違いが認められた。