

人を対象とする医学系研究における 研究計画の立て方

下川敏雄

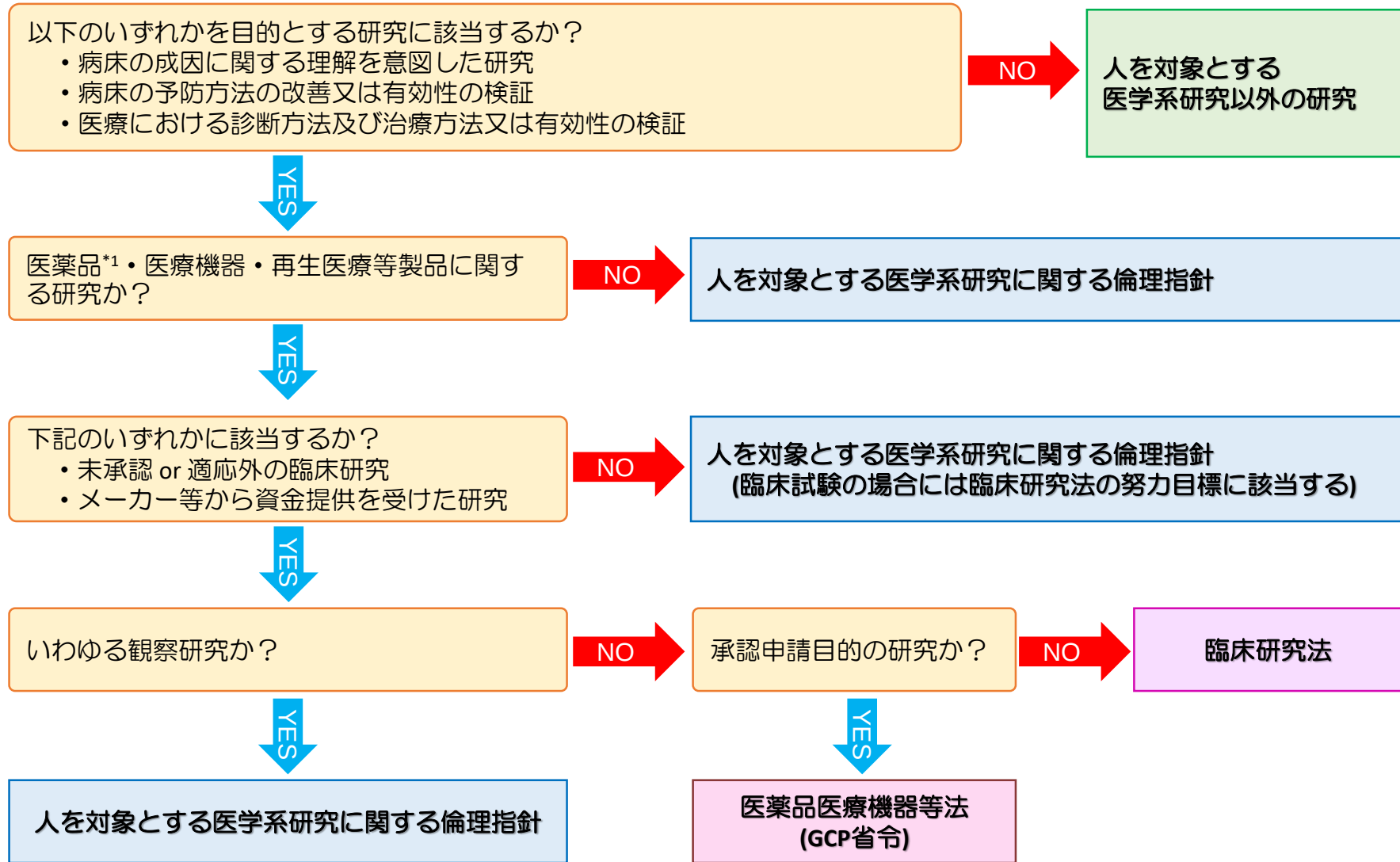
和歌山県立医科大学 医学部
和歌山県立医科大学附属病院 臨床研究センター

本日のお話の流れ

- 人を対象とする医学系研究と研究の区分
- 人を対象とする医学系研究の立案に関する個人的意見
- 臨床試験の区分と留意点
- 観察研究の区分と留意点
- (時間があれば...)臨床試験よもやばなし

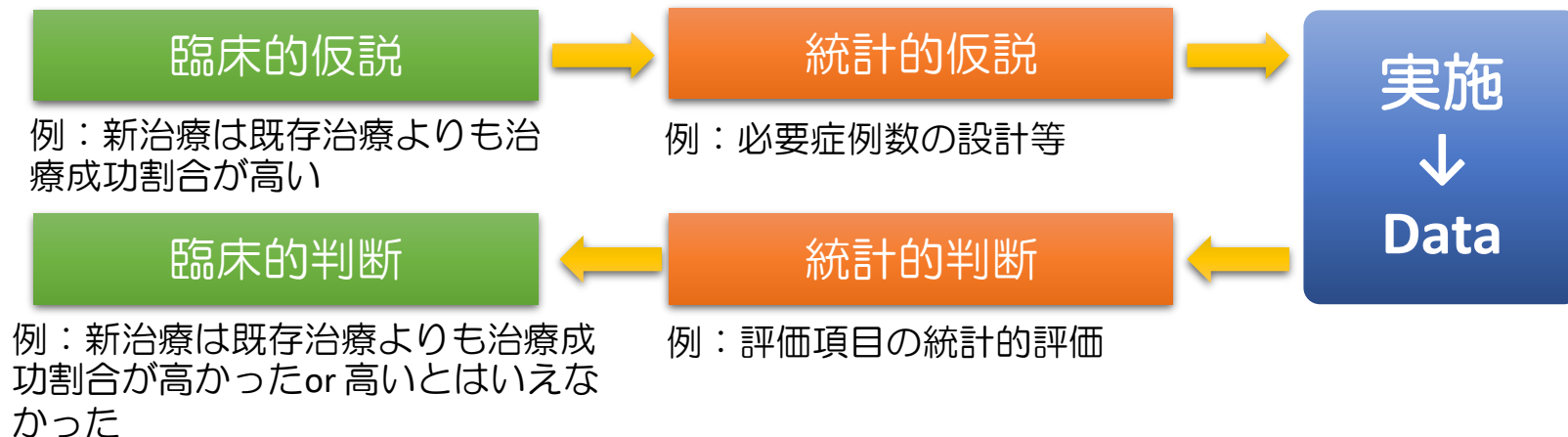
■ 人を対象とする医学系研究と研究の区分

人を対象とする医学系研究と諸種の規則



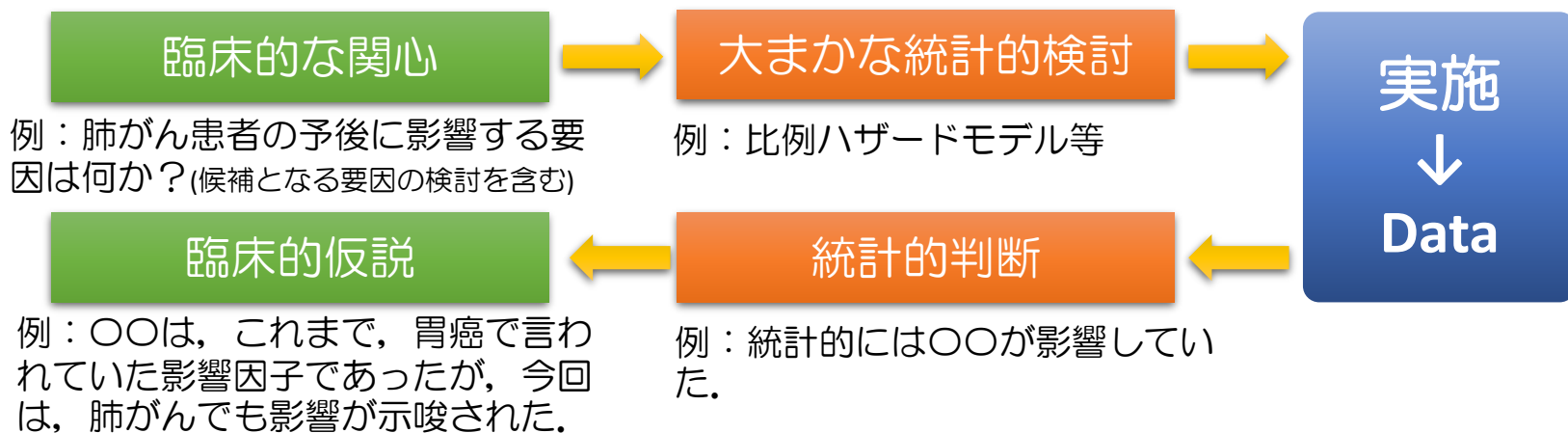
実験研究(experimental study)と観察研究(observational study)

■ 実験研究(experimental study)：臨床試験



つまり、実験研究(観察研究)とは、臨床的仮説が正しいか否かを判断するために実施する。

■ 観察研究(observational study)：臨床試験



つまり、実験研究(観察研究)とは、臨床的仮説を新たに見出すために実施する。

■ 人を対象とする医学系研究の立案に関する個人的意見

臨床研究の御法度(福原, 2013)

1. データを取ってから研究デザインを考える.
2. リサーチクエスチョンが明確・具体的でない.
3. 対象が不明確. 抽出方法, 参入・除外基準を設定しない.
4. 主要なアウトカム変数を設定しない. 変数の吟味はしない.
5. 変数の測定方法の信頼性と妥当性を検討しない.
6. 研究計画を事前に作成しない. サンプルサイズ, 検出力, エフェクトサイズを事前に設定しない.
7. 結果の解釈, 統計的有意差のみで臨床的・社会的に意味のある差かどうかを検討しない.

臨床研究における誤差

偶然誤差
(random error)

偶然によって生じる誤差
ランダムサンプリングによる影響, 個体内変動



標本サイズが影響する
(標本サイズの増加は**定度(精度)**が増す)

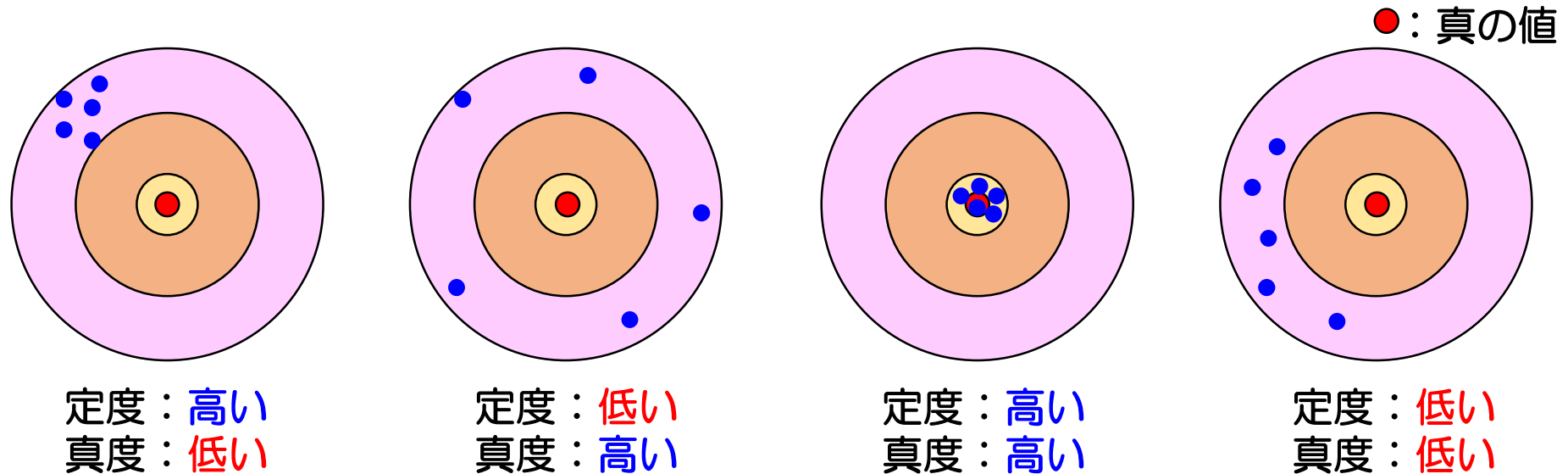
系統誤差
(systematic error)

何らかのバイアスによって生じる誤差



研究のデザインが影響する
(適切なデザインは**真度(正確性)**が増す)

測定の定度(研究のバラツキ)と真度(研究のバイアス)



- 症例数を増やすことは、定度を高めるのに有効(臨床試験においては臨床的有効性との関係で計算)である。なお、反復した測定は、その有効性を更に高める。
- 適切な研究デザインは、バイアスを減らし、真度を高めるのに有効である。
- 研究計画書あるいはマニュアルは、測定者・対象者による誤差を減らし、定度・真度のいずれも増加することに貢献する。

リサーチクエスションの構造化：研究の下ごしらえ

■ リサーチクエスションとは？

- － 研究で明らかにしたいことを宣言した最も短い文
- － 研究計画に最低限必要な要素で構成される

■ なぜ必要か？

- － 漠然としたクリニカルクエスションを研究によって回答可能な形に構造化する
- － 実行までに必要な作業を洗い出す

■ 利益

- － 他の研究者とのコミュニケーションが容易になる
- － 明らかにしたいことが明確になり、自分の頭が整理される

良いリサーチクエスション：FINER

Feasible：実現可能性 **I**nteresting：真に興味深く **N**ovel：独自性があり

Ethical：倫理的で **R**elevant：切実な問題

臨床試験を計画するうえでの論点整理の重要性

■ 臨床試験の場合

既存の治療法(標準治療)の内容

➡ 標準治療における課題及び不明点

↓ 標準治療での治療方法, 治療成績, 課題及び不明点を明確に示す

当該研究の概要

当該研究の新規治療の概要, 標準治療に対して優れていると期待すること (efficacy and/or safetyなど), 新規性に優れていることを明記する.

新規治療が優れている点 or 当該研究領域においてリーズナブルな評価項目を主要アウトカムに設定する

例: 肺がん患者を対象に, 新薬と標準薬で再発までの時間を調べた.

P: Patients (どのような研究対象に対して)

肺がん患者に対して

I: Intervention (どのような介入をすると)

新薬を投与すると

C: Comparison (何と比べて)

標準薬に比べて

O: Outcome (どのような結果になるか・アウトカムは何か)

再発までの期間が伸びた

臨床試験を計画するうえでの論点整理の重要性

■ 観察研究の場合

対象疾患に対する現況の整理

→ 対象疾患時ついでの影響要因等での既存研究と不明点を明らかにする
↓ 既存の論文等での状況、既報において明らかになっていないことの論点整理を行う。

当該研究の概要

当該研究で明らかにしたいこと、本研究に新規性があることを明記する。

観察研究において、「影響要因を探索する..」等の記述が多いが、その影響要因の候補となる者を吟味しないと何もできない。また、影響要因は結果(当該疾患の罹患の有無等)よりも時間軸で前にないといけない(記述的な場合は同じ時点)。決して、アウトカムよりも遅い場合はダメ

例：喫煙による肺癌への発症のリスクを調べた。

P :Patients (どのような研究対象に対して)

健常者に対して

E :Exposure(暴露すると)

喫煙習慣があると

C :Comparison (何と比べて)

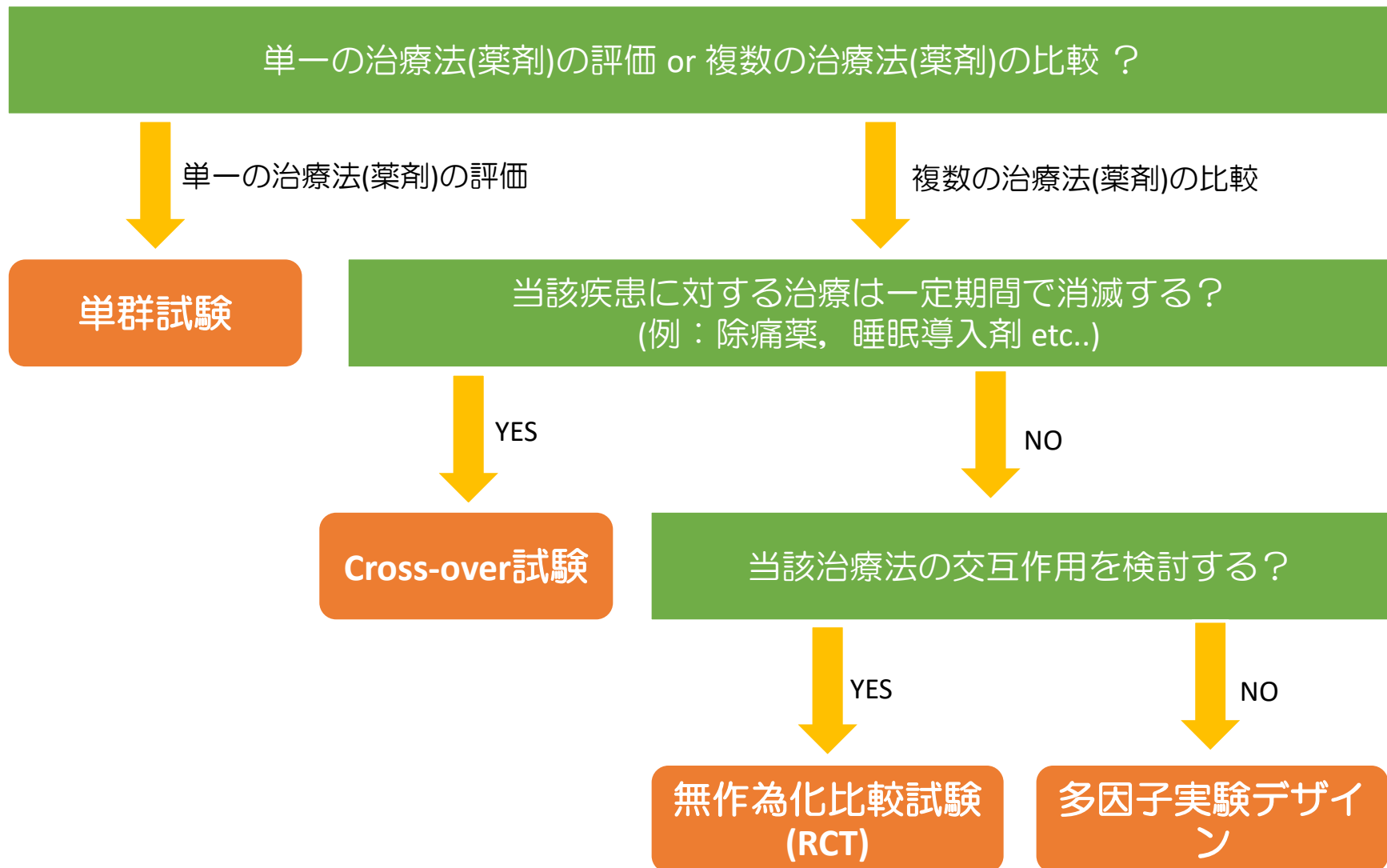
喫煙習慣がない人に比べて

O: Outcome (どのような結果になるか・アウトカムは何か)

肺がんの発症リスクが高まる

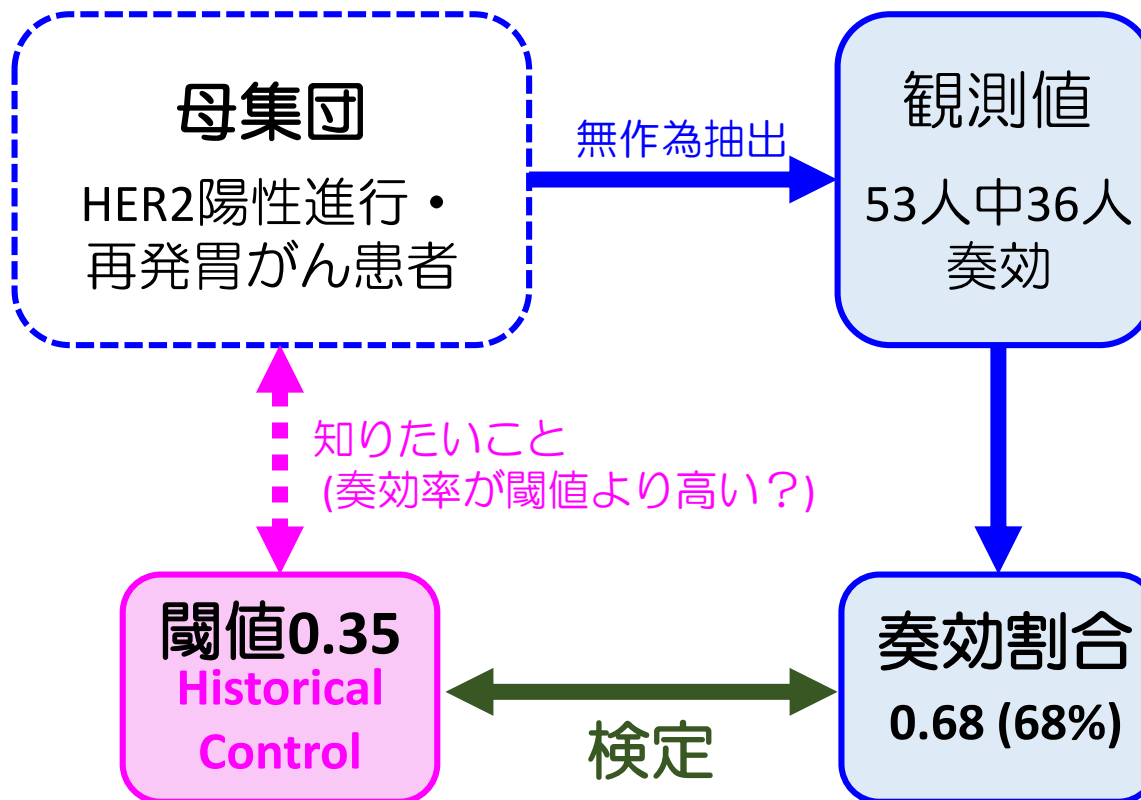
■ 臨床試験の区分と留意点

臨床試験の幾つかのタイプ



臨床研究における単群試験の例示

HER2陽性の進行・再発胃がん患者に対するフッ化ピリミジン系抗がん剤とCDDPを併用したレジメンでの奏効率が35%であると報告されている。新しい臨床試験では、S1+CDDP+Tmabの併用療法による新たなレジメンの有効性を単アーム試験で実施した。その結果、53例中36例の奏効例が認められた(奏効割合=68%) (Kurokawa et al., *BJC*, 2014)



p値<0.001であることから、高度に有意だった。すなわち、本治療レジメンは、HER2陽性の進行・再発胃がん患者に対して、非常に**有望**であることが証明された。

単群試験での留意点

■ 閾値(帰無仮説)には根拠が必要

単群試験では、閾値に対して臨床的な根拠が必要であり、統計解析は閾値よりもポジティブな結果か否かが評価される。ランダム化比較試験では、帰無仮説は「差がない(or $HR=1$)」になる。

HERBIS-1では、ToGA試験(Y-J Bang, et al. *The Lancet*, 28, 687 - 697, 2010.)の結果を参考に行っている。ToGA試験では、化学療法群(5-FU /CDDP あるいはCapecitabine /CDDP)での**奏功率が35%**であった。

■ 期待値(対立仮説)には実現可能性及び臨床的有効性に基づいて設定

期待値は、臨床的に合意されるような有効性(エフェクトサイズ)として設定される。したがって、期待値が本臨床試験のエンドポイントの目標値となる。

ToGA試験では、(1) Trastuzumab併用群の**奏功率が47%**であること、(2) HER2による選択基準(IHC 3+ or FISH positive + IHC 2+)、(3) 本邦の状況を鑑みて、閾値奏功率を50%と設定している。

Cross-over試験



- 異なる治療を同一被験者に施行することで、群間の被験者の違いがなくなることから、標本サイズが大幅に削減できる.
- 持ち越し効果(carryover)が存在する場合には統計モデルが破綻するため注意が必要.
- プラセボ対照試験の場合、心的持ち越し効果(盲検化の破綻)によるプラセボ効果の影響が分離できない可能性がある.

クロスオーバー試験は症例を大幅に削減できるものの用途は限定的であり、注意が必要

無作為化比較試験(並行群間試験) : randomized clinical trials



- 群間の患者像が同じようにランダムに割り付ける試験
→ 無作為化比較試験
- 群間の患者像が同じようにランダムに割り付けない試験
→ 非無作為化比較試験

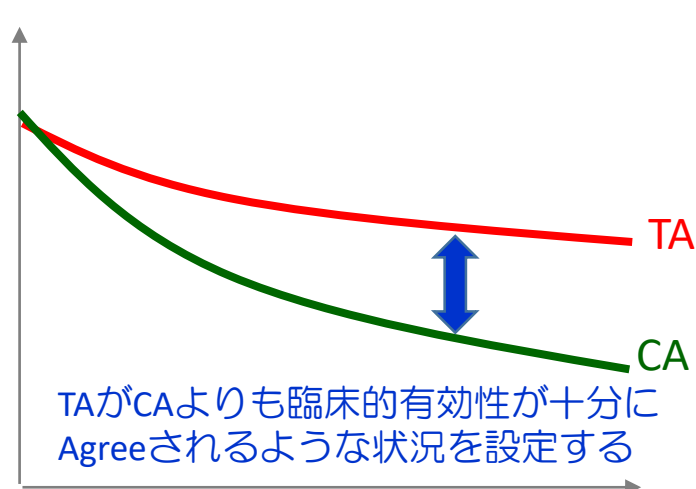
無作為化比較試験では、主要評価項目に影響を与えると考えられる要因に関して均等化を行うことで、群間の患者像をそろえる。これを層別化といい、その要因を層別因子という。

無作為化比較試験の幾つかのタイプ：優越性試験

優越性試験の例：新規治療の有効性は期待できるが、toxicityが悪そうな場合 (Toxic New)



- 上記の状況を証明するには、**優越性試験**のみが許容される。
- つまり、試験治療が既存治療に対して臨床的に意味がある差異として認識されなければ、安全性を担保できない



優越勢デザインでは、TAのハザード H_T がCAのハザード H_C に比べて臨床的に有効であることを統計的に証明しなければならない、次の仮説

帰無仮説： $H_T \leq H_C$

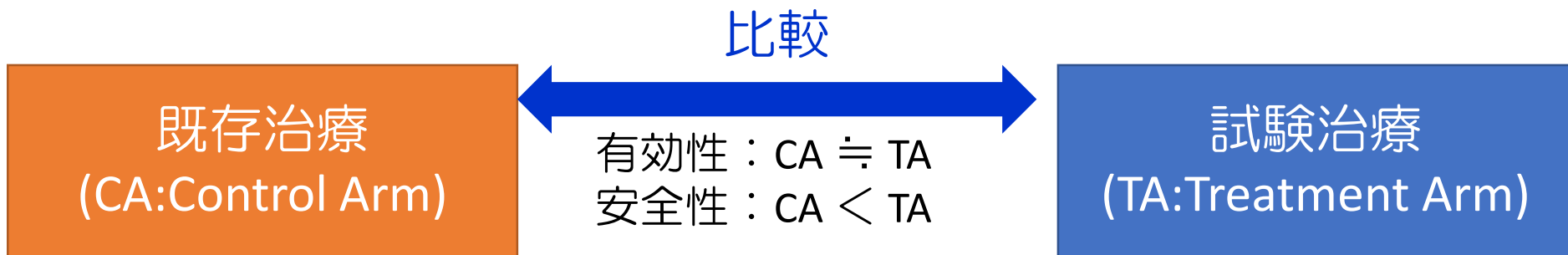
対立仮説： $H_T < H_C$ (片側 or 両側対立仮説)

を検定する(両側の場合には" $\neq$ "を証明)。

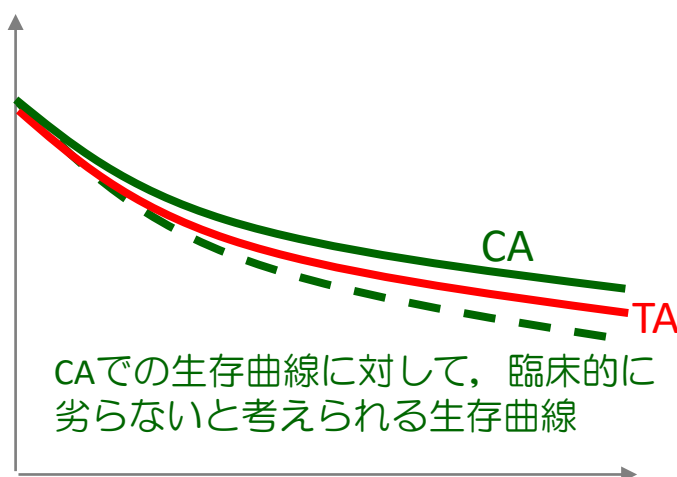
このときの臨床的に有効な差異は、標本サイズの選定のときに考慮される。

無作為化比較試験の幾つかのタイプ：非劣勢試験

第III相試験の目標：新規治療の安全性(など)のBenefitがある場合



- 上記の状況を証明するには、**非劣勢試験**として行えることがある。
- つまり、試験治療が既存治療に対して有意に劣らなければ、ひとつの治療アームとしての価値があることを示すことができる。
- 安全性だけでなく、たとえば、既存治療が入院だったのに対して、新規治療が外来である、あるいは治療コストが安価になるなどのBenefitも考えられる。



非劣勢デザインでは、TAのハザード H_T がCAのハザード H_C に比べて臨床的に劣らない範囲 Δ を決定し、次の仮説

帰無仮説： $H_T = H_C + \Delta$

対立仮説： $H_T < H_C + \Delta$ (片側対立仮説)

を検定する。すなわち、TAの治療は、CAに比べて、 Δ (臨床的に許容される下限値)まで劣ることはない治療であることを証明している。

無作為化比較試験での留意点：CONSORT声明

CONSORT(Consolidated Standards of Reporting Trials：臨床試験報告に関する統合基準)とは，無作為化比較試験の報告の質の向上について記載された基準であり，論文とともにチェックリストを提出することが義務付けられるようになってきている．最新版は「CONSORT2010」

□ CONSORT2010 Check List

無作為化比較試験を報告する際に記載すべき情報のチェックリストになっている．本チェックリストを埋めるには，適切なプロトコルを作成することが必須．

□ CONSORT Flow Diagram

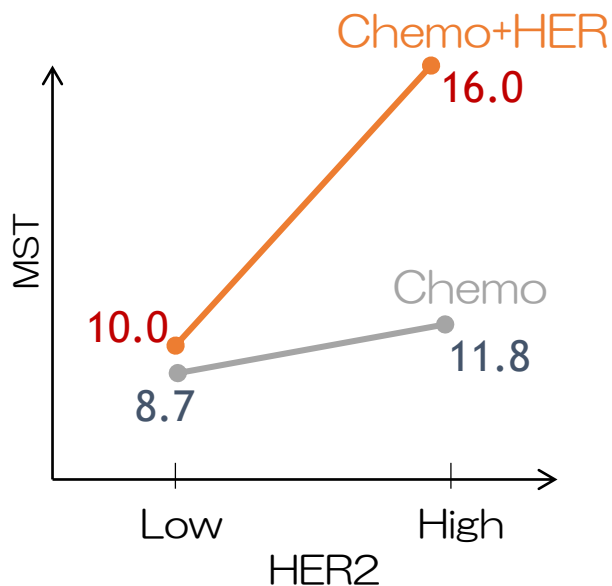
被験者の登録・割付・追跡・解析に至るすべての状況において除外された症例とその理由等を記載するためのダイアグラムである．臨床試験の適切な運用が求められる．

詳細は付録を参照

交互作用とは

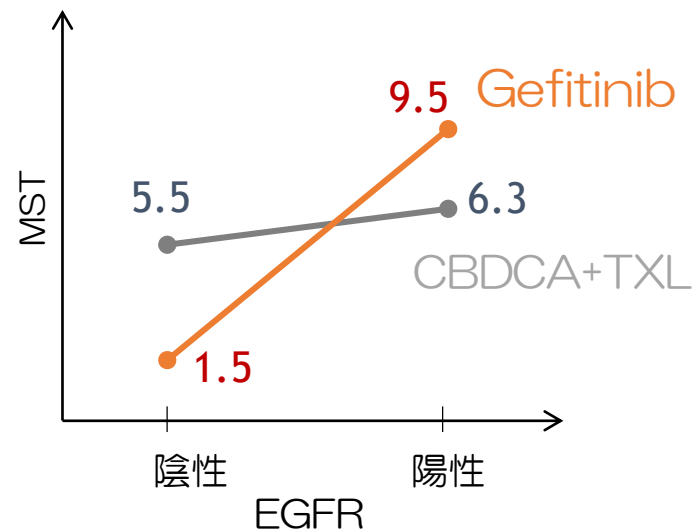
ToGA Study

(a) MSTが上昇



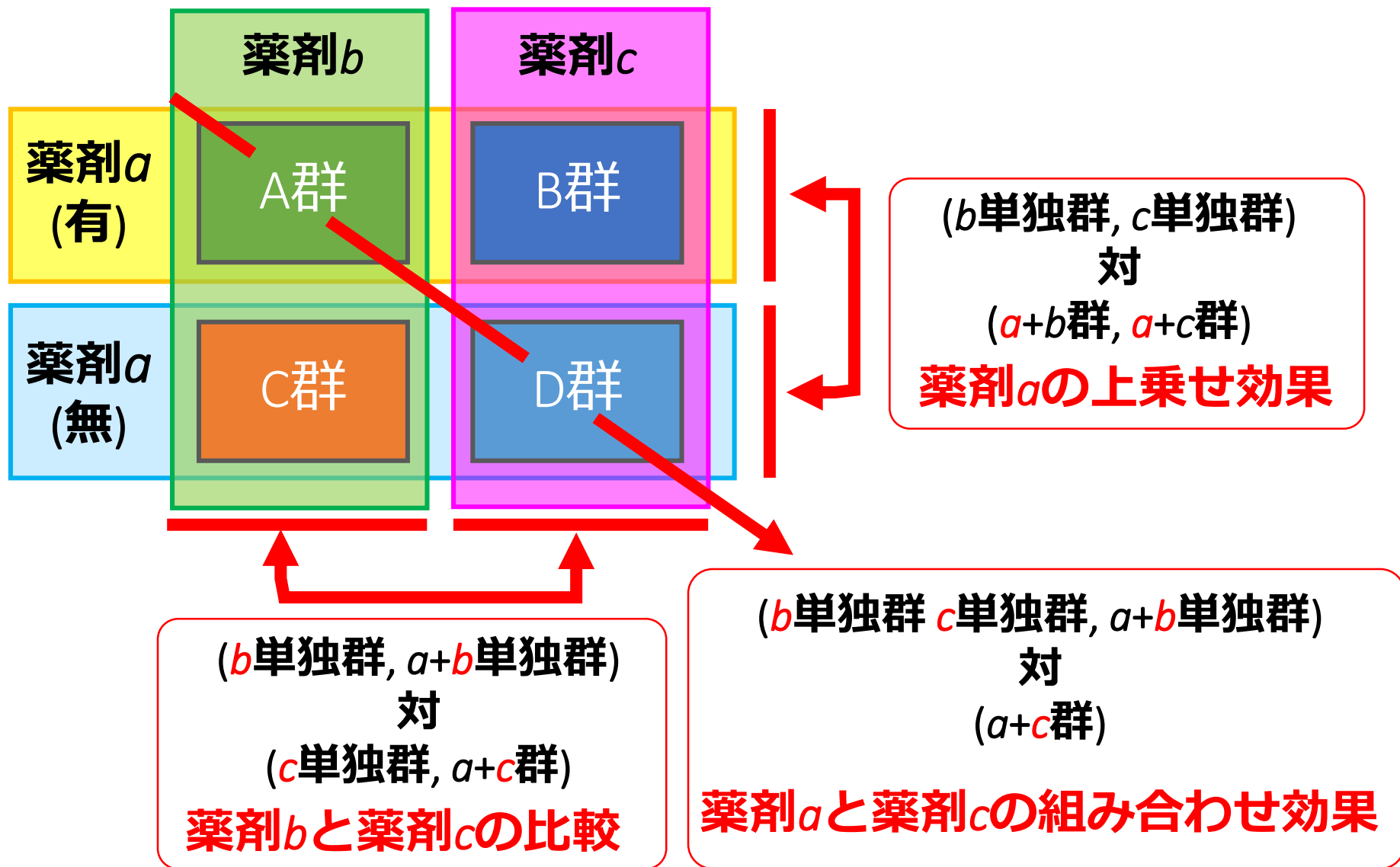
IPASS Study

b) MSTが交差



何らかの要因によって、抗がん剤の効果が逆転したり、相乗効果によって、より効果が上昇するような現象を交互作用(interaction)という。
交互作用は、治療方法の一般可能性に繋がる。

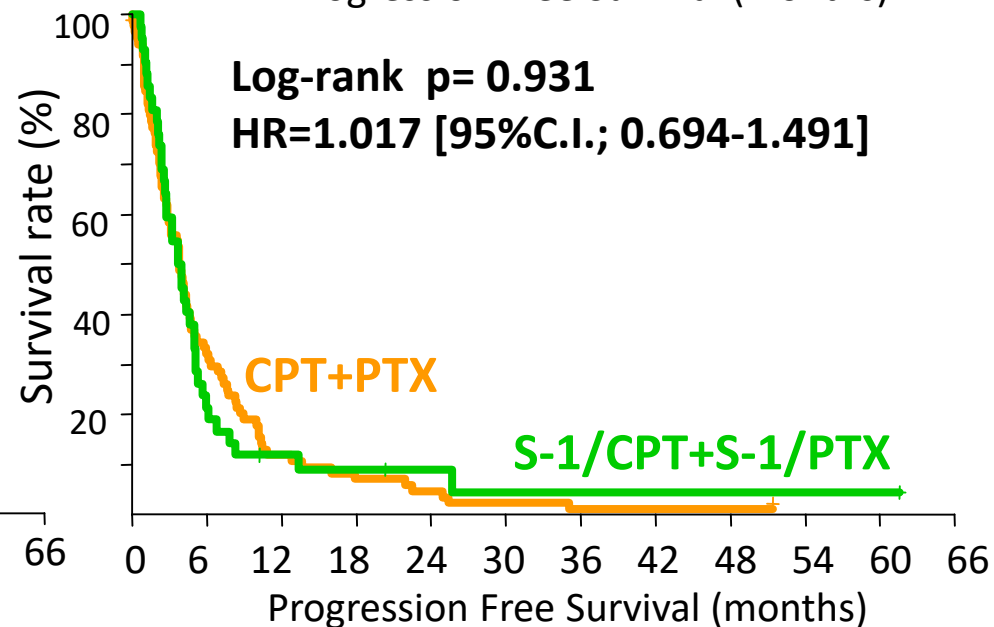
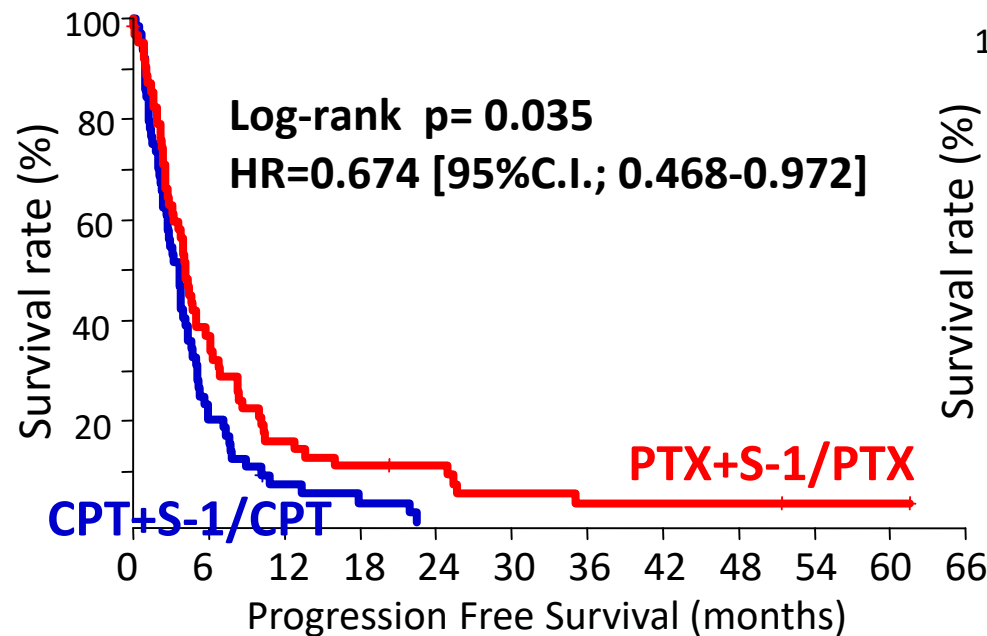
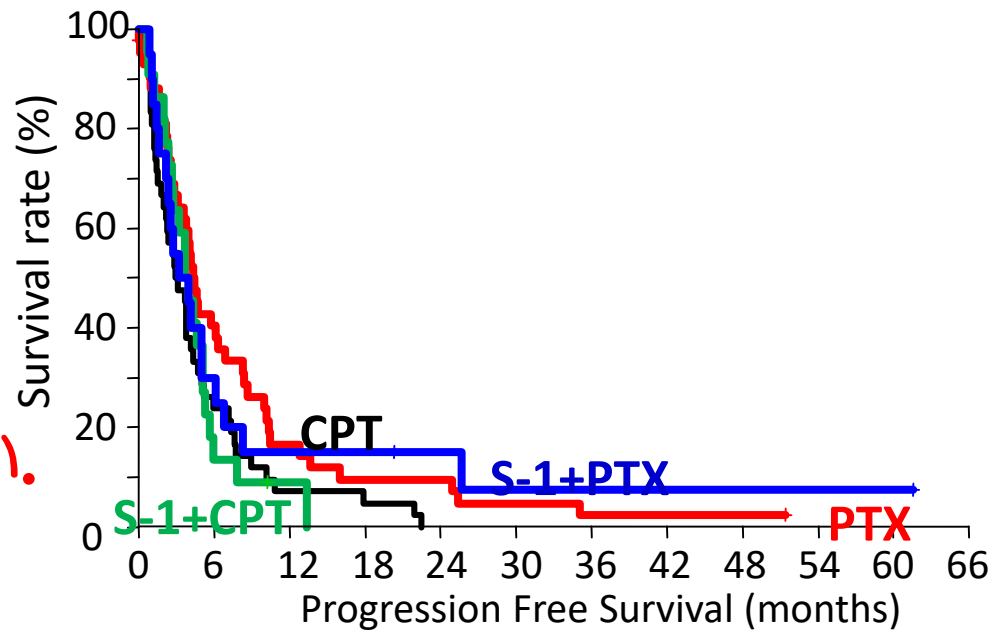
多因子実験デザイン(factorial design)



多因子実験デザインの例示

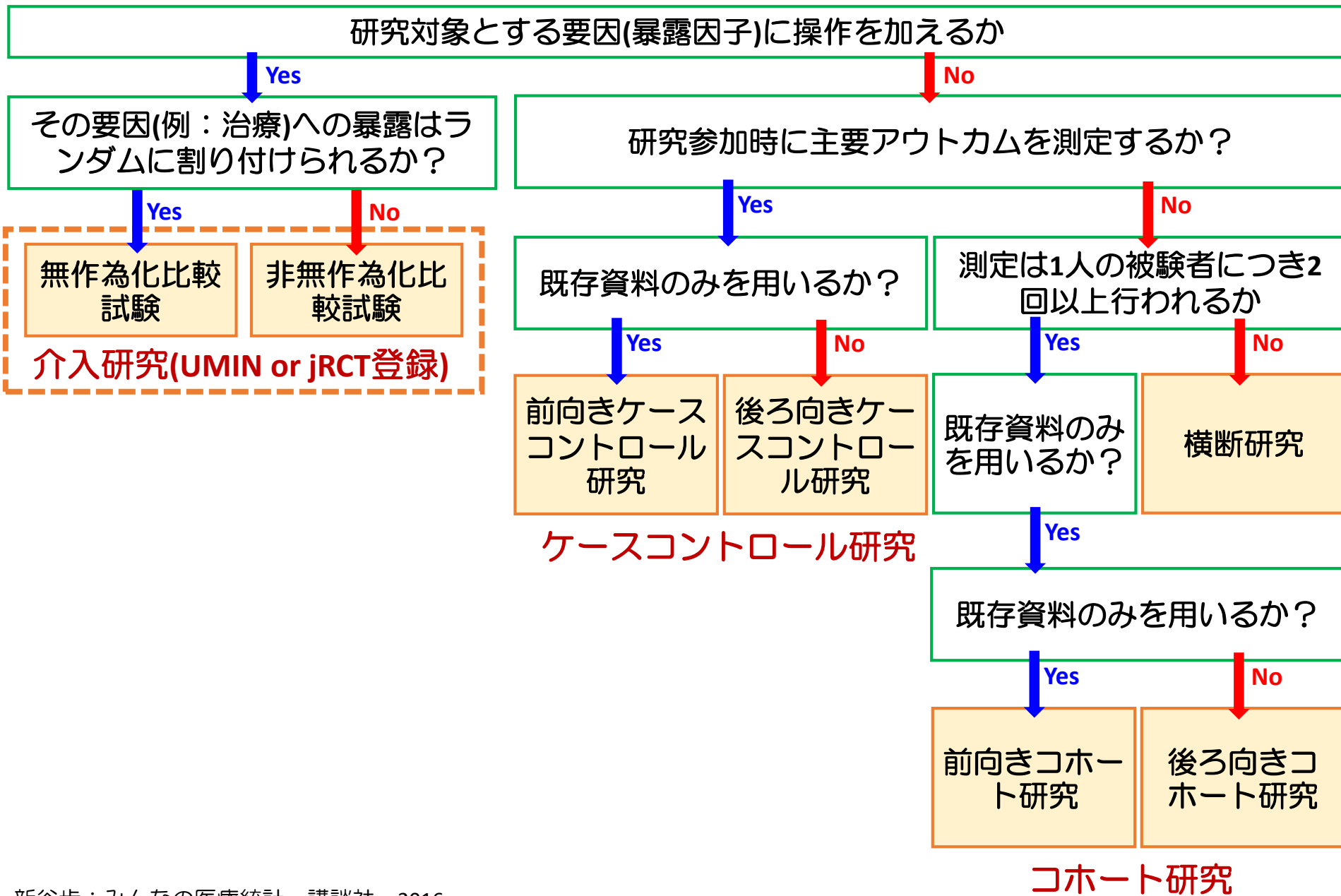
TS-1単独またはTS-1/CDDP併用治療に抵抗性となった進行・再発胃癌に対する二次治療の無作為化第II相試験 (Nishikawa et al; ASCO-GI2015)

- CPT(+S-1)よりもPTX(+S-1)のほうがPFSを伸ばす。
- S-1の上乗せ効果は認められない。
- 交互作用は示唆されない。



■ 観察研究の区分と留意点

研究のデザイン(観察研究を中心に...)



縦断研究(コホート研究とケースコントロール研究)

■ コホート研究(Cohort study)

研究対象となるアウトカムの発症していない集団をある要因に曝露した集団と曝露していない集団に分けて一定期間追跡し、アウトカムの発生率を比較することで、要因とアウトカムの因果関係を調べる研究。

■ ケースコントロール・症例対照研究(Case-control study)

アウトカムに罹患した集団とそうでない集団に分けて、研究対象要員に曝露しているかどうかを調べることにより、要因とアウトカムの因果関係を調べる研究

これらの研究は、観察的疫学研究のなかでも、とくに分析疫学と呼ばれる。観察疫学研究には、**記述疫学**(因果関係や仮説の検証は含まないが状況把握や仮説構築を行う)研究がある。

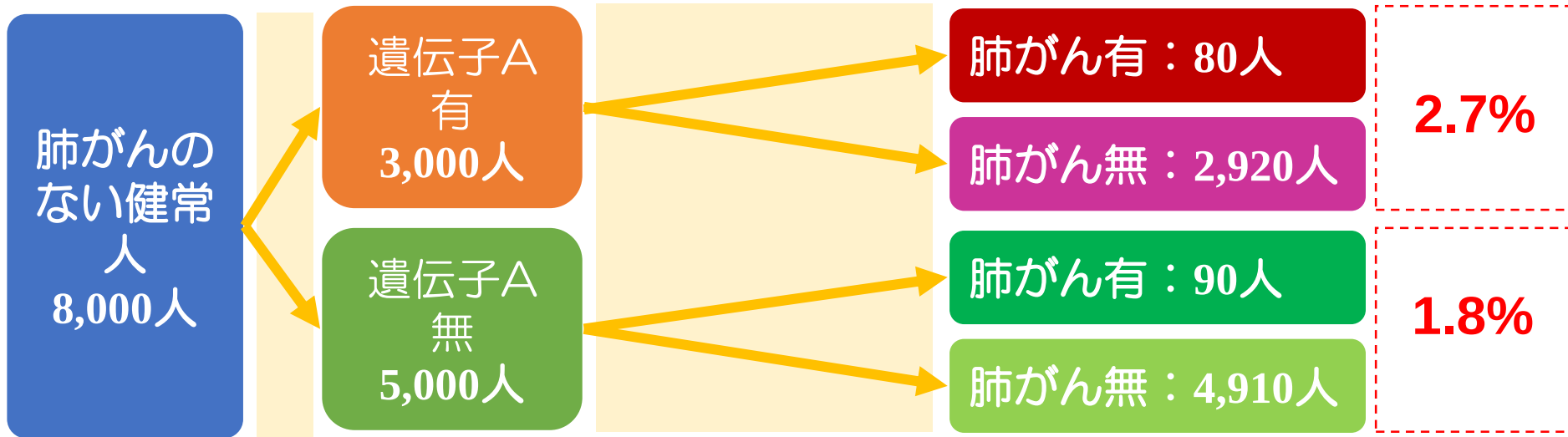
コホート研究の例示

ある遺伝子Aが肺がんに関係するかを見るための研究

遺伝子検査

フォローアップ30年

発症率

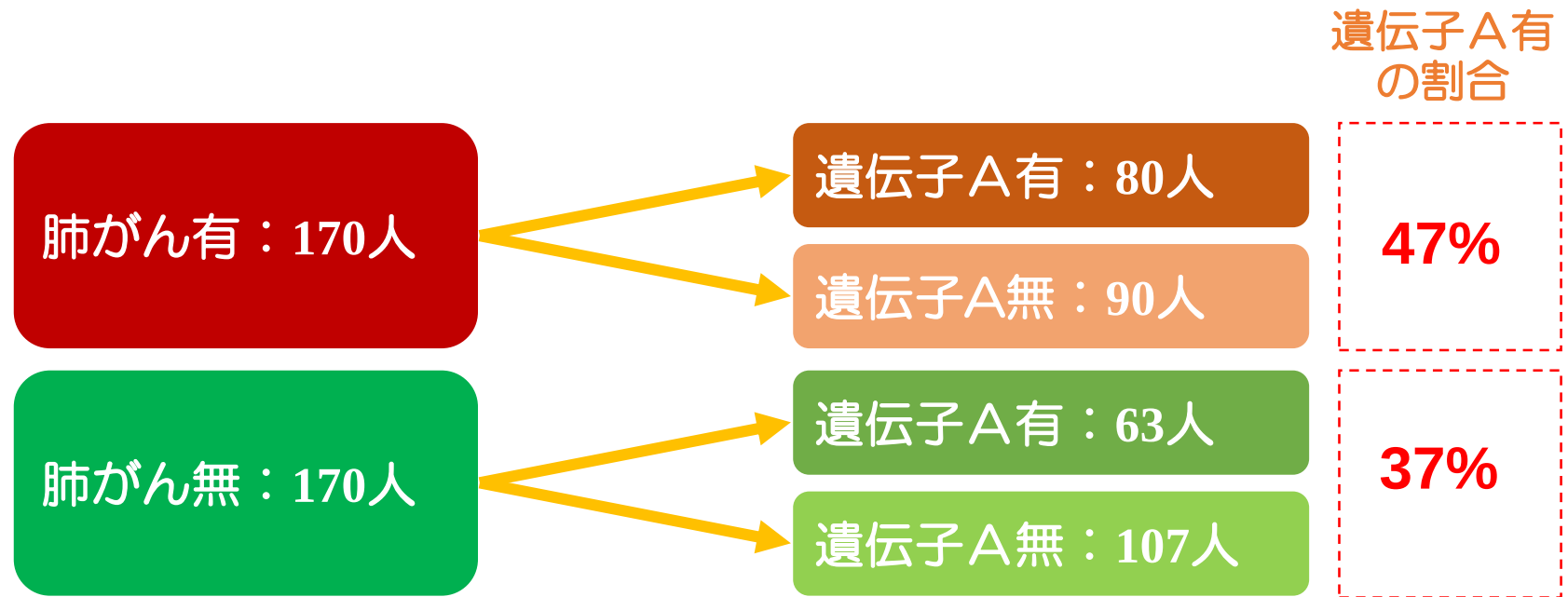


遺伝子Aが有る人はない人に比べて、肺がんの発症率を0.9%上昇させる。

- リスク比(相対危険度)を求めることができ、リスク因子の複数の疾病発生への影響を調べることができる。
- 研究に多くの時間と費用がかかる。
- 対照群の設定が難しい。例えば、上記研究においてもそれぞれの集団がリスク因子(遺伝子A)の曝露だけが異なるとは限らない。

ケースコントロール研究の例示

ある遺伝子Aが肺がんに関係するかを見るための研究

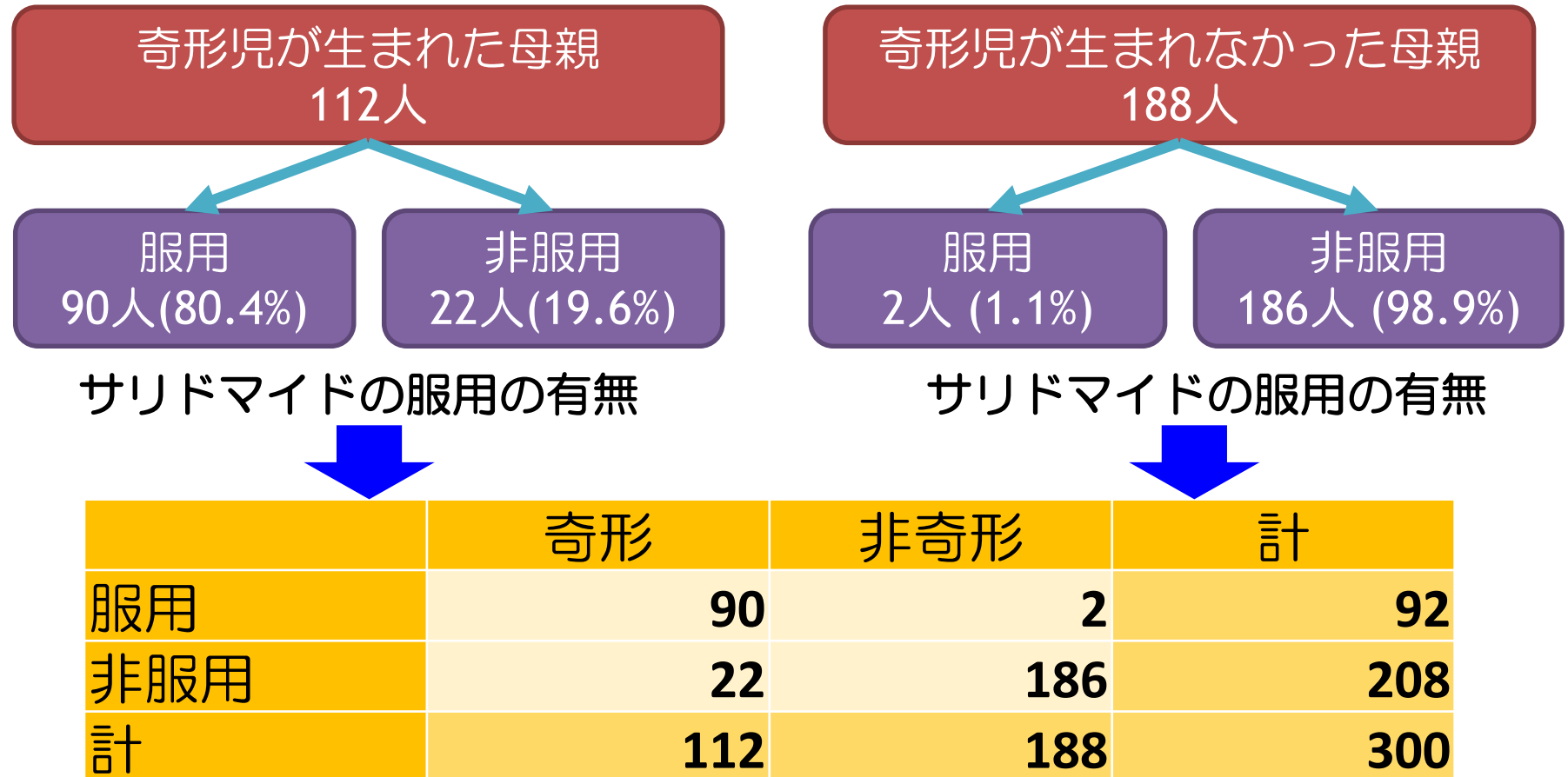


肺がん患者は非肺がん患者よりも遺伝子A有の割合が10%高い。

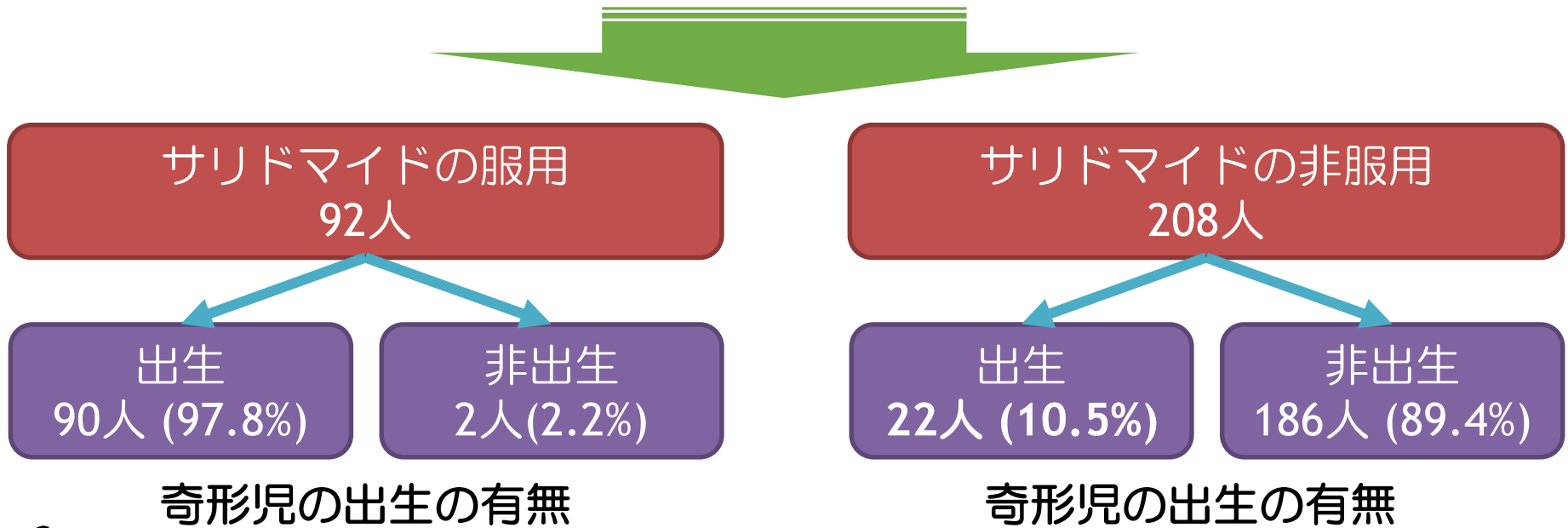
- ケースコントロールに比べて時間と費用が大幅に削減できる。
- 対照群の理想は、関心のある疾患を持っていない点だけが患者群と異なることであるが、そのような設定が難しい。

ケースコントロール研究の落とし穴

このデータは、サリドマイドによる薬害を調査したLentz博士が実施したケースコントロール研究の結果である。



ケースコントロール研究の落とし穴



サリドマイドを服用していなくても、10.5%の母親が奇形児を出生しているというのは多すぎないか？

奇形児の出生の有無に基づいて、サリドマイドの服用を調査している。したがって、上記のような割合を出すことはできない。

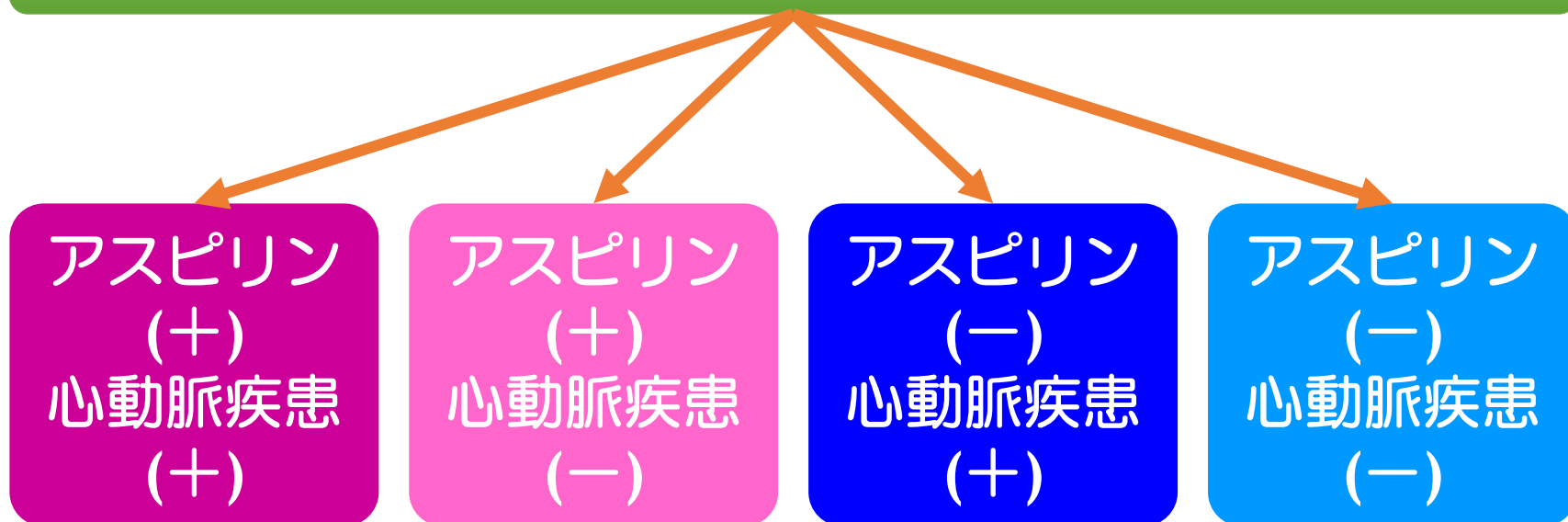
臨床研究において、「時間的な概念」は非常に重要である。

横断研究：Cross sectional study

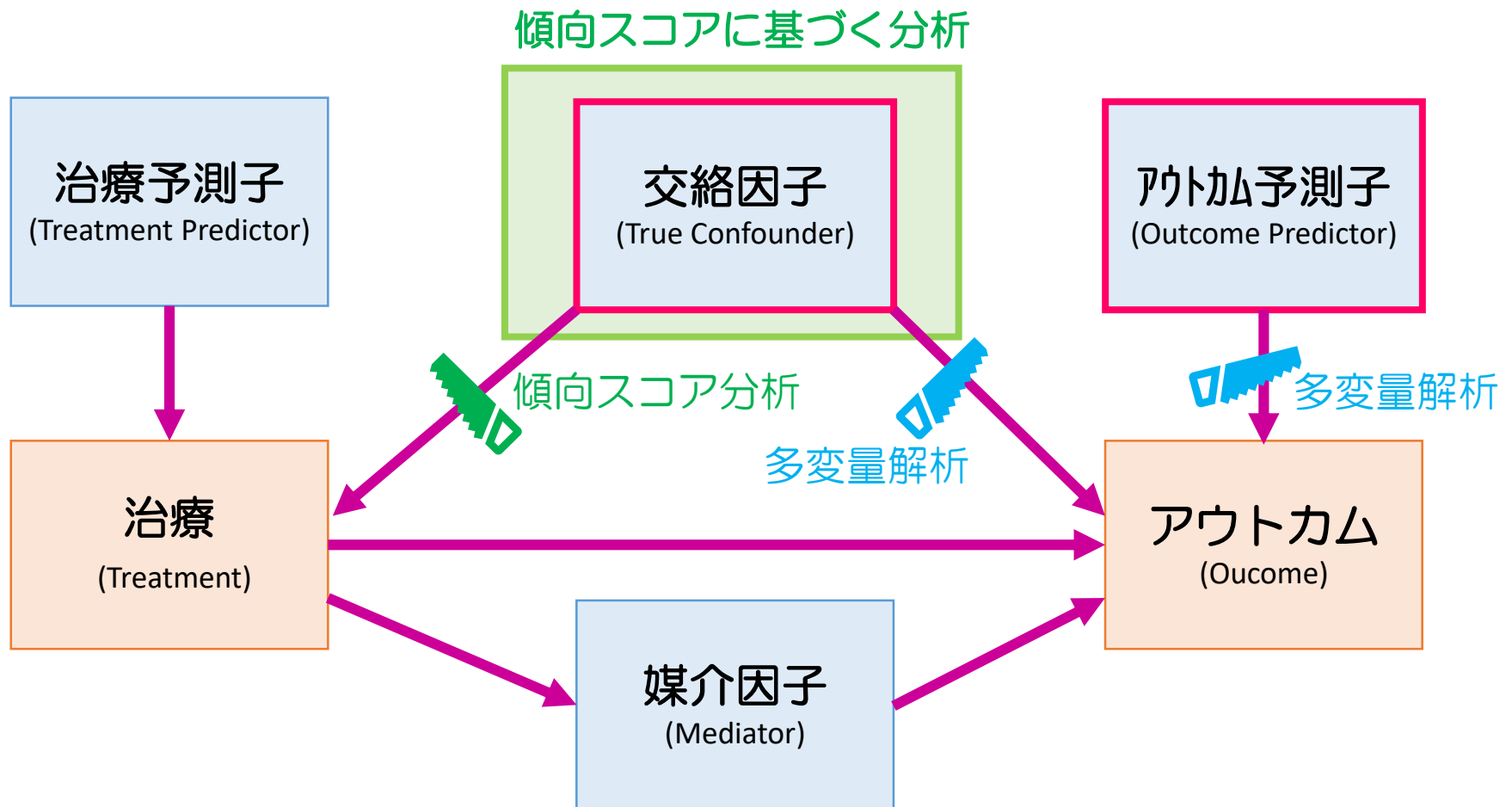
- 追跡を伴わない研究(例：アンケート調査).
- 一方で，前向き研究や後ろ向き研究では，時間軸が存在するため，縦断研究(longitudinal study)という

例：通院患者にアンケートをとり，過去に心動脈疾患になったか，アスピリンを常用していたかどうか調査する。

曝露と疾患に関するデータ収集



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(偏りは関係ない)。

傾向スコアとその留意点

傾向スコアとは、研究対象因子に暴露する確率で表される(つまり、傾向スコアは0～1の範囲をとる)。

暴露群

非暴露群

傾向スコア(propensity score)

暴露・非暴露が完全に分離しているため、比較可能性が担保できない(いいかえれば、患者の違い明確であり、アウトカムを比較する状況にないことを示している)

暴露群

非暴露群

傾向スコア(propensity score)

傾向スコアが重複している部分については、暴露・非暴露のいずれも候補になり得る。このような状況においては、比較可能性が担保されている。

傾向スコアの例示：早期胃がん患者に対する研究

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†*

早期胃がん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

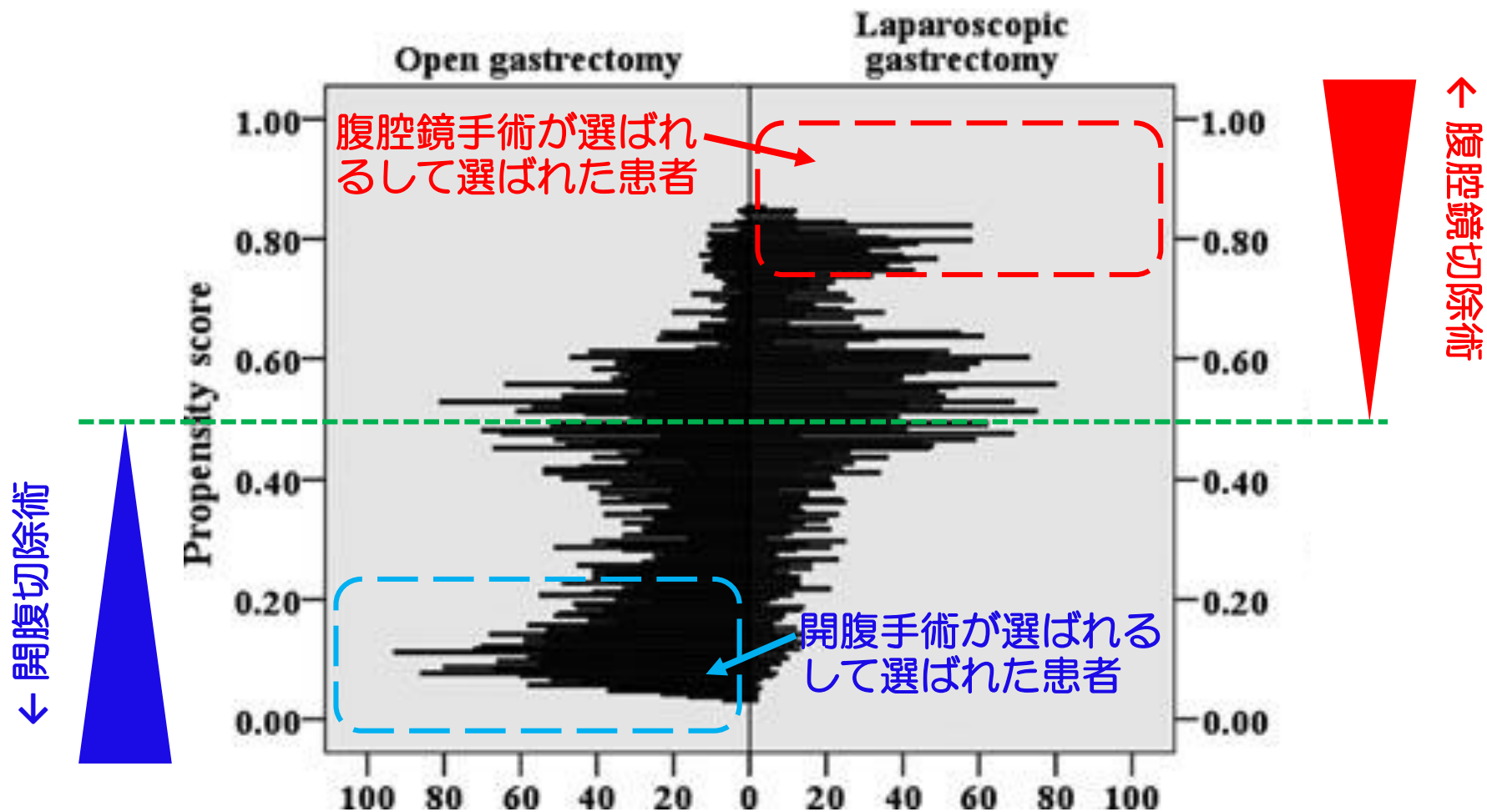
To estimate the propensity score, we fitted a **logistic regression model** for the receipt of laparoscopic gastrectomy as a function of patient demographic and hospital factors including **age, sex, Charlson comorbidity index, body mass index, smoking index, cancer stage (I or II), hospital volume category, 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



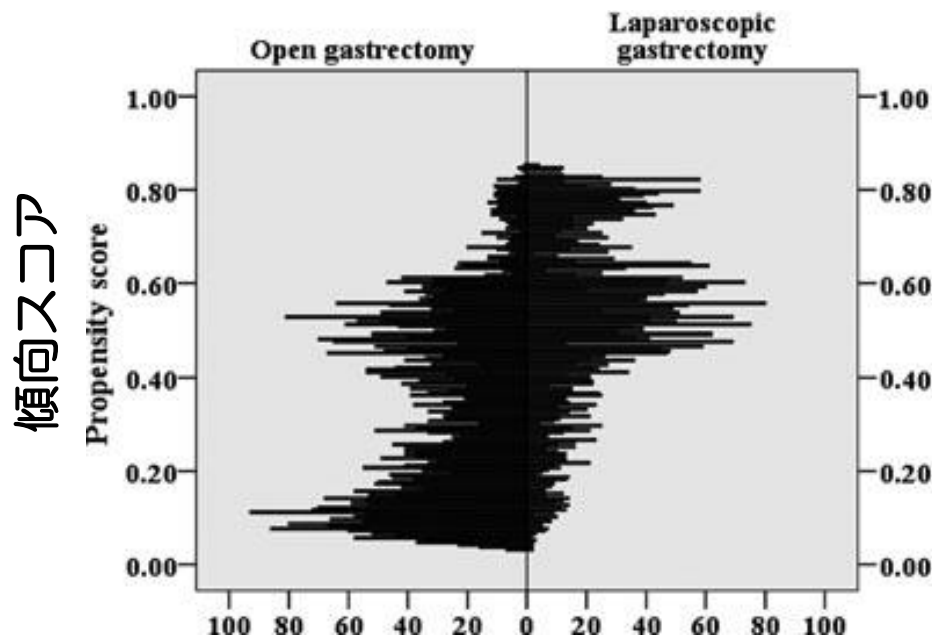
先ほどの事例におけるマッチングの結果

論文の記載

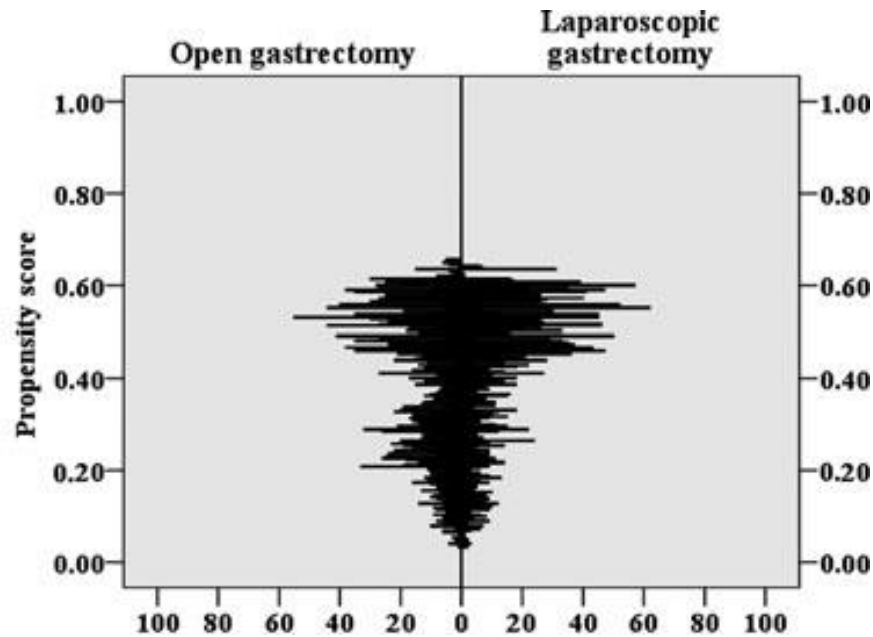
We performed a one-to-one matching analysis between laparoscopic and open distal gastrectomy groups on the basis of estimated propensity scores of each patient.

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

各共変量の比較

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。

マッチング後の主要評価項目の解析結果

	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		

当該研究において、以下のことが分かった

- 麻酔時間 腹腔鏡 > 開腹
- 術後在院日数 腹腔鏡 < 開腹
- 費用 腹腔鏡 > 開腹

■ 臨床試験よもや話

症例数設計について：計画段階



- 臨床試験を行いたいが集積可能な被験者数は決まっている
- 研究したい仮説も決まっているので、この被験者数が妥当であるという理由を統計学的な観点から作成して欲しい

■ 医学統計家は仮説に基づいて必要症例数を計算する。

- 仮説に基づいて統計学的に症例数を計算したとき、必要症例数が集積可能な被験者数を上回った場合どうするか？

検討事項1：統計的な例数設計を捨てて、パイロット試験として実現可能性の観点から例数設計を行う(この場合、検定結果は参考資料扱い)。

検討事項2：単施設での研究であれば、多施設にできないかどうか、多施設共同研究の場合には、さらに施設を増やせないかを相談する(施設数を増やさずに、もっと頑張るから...となった場合、十中八九、症例が集積でない)。

検討事項3：無作為化比較試験で検討していた場合には、単群試験、無作為化第II相試験にするなど、フェーズを1段階下げること検討する。

症例数設計について：論文投稿段階



- 非盲検試験のデータを眺めていると、どうやら主要評価項目では差がつきそうにない
- この試験の主要評価項目で差がつかないと研究がネガティブであると評価される
- 副次的評価項目の中のひとつが差が出そうな雰囲気である
- この段階でプロトコルを修正し、差が出そうな副次的評価項目を主要評価項目としたい

主要評価項目の選定は、研究の計画段階で決めることで、それを変えるということは基本的にはあり得ない話です。

主要評価項目に基づいて例数設計が行われているので、その仮定も変わります。

「有意差が出そうだから」ではなく、「臨床的に最も有益である」の観点で主要評価項目は決まったはずです。

副次的評価項目は検証的に評価されていませんので、主要評価項目でnegativeになった場合には、negativeということで書きましょう。

そのうえで、副次的評価項目において、有望な結果が得られたと正直に書きましょう。

パブリケーションバイアスの観点から、negative studyだからといって、良いjournalに掲載されるというわけではありません(top journalがすべてpositive studyではありません)。

これらは無作為??



- ・ランダム化を行うためのシステムを利用することが面倒である。
- ・被験者が来た順番に従って、2つの治療を順番に割り付けたい。
- ・被験者の来院はランダムに生じるのだから、ランダム化試験であると言えるはずだ。

ランダム化ではありません!!



全被験者

2群間の共
変量の影響
を均一化

R

新治療群

既存治療

新治療
での結果

明らかになる
新治療群の結果
とほぼ同じ
明らかにならない

既存治療
での結果

明らかにならない
既存治療群の結
果とほぼ同じ
明らかになる

共変量

ランダム化を実施することで、新治療群と既存治療群の群間で同じ集団になることを担保しなければならない。順番に割り付け場合、同じ集団になることは保証できない。

ご清聴ありがとうございます



shimokaw@wakayama-med.ac.jp

■ バックアップ

- 人対象指針及び臨床研究法においてプロトコルに求められている内容一覧
- 臨床試験における必要症例数設計の意味
- 交互作用を伴う臨床試験結果の例示(ToGA study, IPASS study)
- 諸種のガイドライン(CONSORT, STROBE)
- EZRによる傾向スコア分析

人を対象とする医学系研究に関する倫理指針で求められている研究計画書の項目

- ① 研究の名称
- ② 研究の実施体制（研究機関の名称及び研究者等の氏名を含む。）
- ③ 研究の目的及び意義
- ④ 研究の方法及び期間
- ⑤ 研究対象者の選定方針
- ⑥ 研究の科学的合理性の根拠
- ⑦ インフォームド・コンセントを受ける手続等（インフォームド・コンセントを受ける場合には、同規定による説明及び同意に関する事項を含む。）
- ⑧ 個人情報等の取扱い（匿名化する場合にはその方法を含む。）
- ⑨ 研究対象者に生じる負担並びに予測されるリスク及び利益、これらの総合的評価並びに当該負担及びリスクを最小化する対策
- ⑩ 試料・情報（研究に用いられる情報に係る資料を含む。）の保管及び廃棄の方法
- ⑪ 研究機関の長への報告内容及び方法
- ⑫ 研究の資金源等、研究機関の研究に係る利益相反及び個人の収益等、研究者等の研究に係る利益相反に関する状況
- ⑬ 研究に関する情報公開の方法
- ⑭ 研究対象者等及びその関係者からの相談等への対応
- ⑮ 代諾者等からインフォームド・コンセントを受ける場合には、その手続
- ⑯ インフォームド・アセントを得る場合の手続
- ⑰ 生命の危険に瀕している場合には、同規定に掲げる要件の全てを満たしていることについて判断する方法
- ⑱ 研究対象者等に経済的負担又は謝礼がある場合には、その旨及びその内容
- ⑲ 侵襲（軽微な侵襲を除く。）を伴う研究の場合には、重篤な有害事象が発生した際の対応
- ⑳ 侵襲を伴う研究の場合には、当該研究によって生じた健康被害に対する補償の有無及びその内容
- ㉑ 通常の診療を超える医療行為を伴う研究の場合には、研究対象者への研究実施後における医療の提供に関する対応
- ㉒ 研究の実施に伴い、研究対象者の健康、子孫に受け継がれ得る遺伝的特徴等に関する重要な知見が得られる可能性がある場合には、研究対象者に係る研究結果（偶発的所見を含む。）の取扱い
- ㉓ 研究に関する業務の一部を委託する場合には、当該業務内容及び委託先の監督方法
- ㉔ 研究対象者から取得された試料・情報について、研究対象者等から同意を受ける時点では特定されない将来の研究のために用いられる可能性又は他の研究機関に提供する可能性がある場合には、その旨と同意を受ける時点において想定される内容
- ㉕ モニタリング及び監査を実施する場合には、その実施体制及び実施手順

臨床研究法で求められている研究計画書の項目 (施行規則14条)

1. 臨床研究の実施体制に関する事項
2. 臨床研究の背景に関する事項（当該臨床研究に用いる医薬品等の概要に関する事項を含む）
3. 臨床研究の目的に関する事項
4. 臨床研究の内容に関する事項
5. 臨床研究の対象者の選択及び除外並びに臨床研究の中止に関する基準
6. 臨床研究の対象者に対する治療に関する事項
7. 有効性の評価に関する事項
8. 安全性の評価に関する事項
9. 統計的な解析に関する事項
10. 原資料等の閲覧に関する事項
11. 品質管理及び品質保証に関する事項
12. 倫理的な配慮に関する事項
13. 記録（データを含む。）の取扱い及び保存に関する事項
14. 臨床研究の実施に係る金銭の支払及び補償に関する事項
15. 臨床研究に関する情報の公表に関する事項
16. 臨床研究の実施期間
17. 臨床研究の対象者に対する説明及びその同意（これらに用いる様式を含む）に関する事項
18. その他、臨床研究の適正な実施のために必要な事項

「臨床研究法の施行に伴う新政令の制定について(厚生労働省医政局長通知 [医政発0228第10号])」にはそれぞれの項目に関する留意点が述べられている。

臨床試験(介入研究)において最も留意しなければならないこと

臨床試験における統計的評価とは、臨床的に設定された有効性・安全性(臨床的効果)に関する仮説が正しいか否かを統計的に判断することである。

臨床的効果^{*1}
の設定

必要症例数
を設定

臨床試験
の実施

統計解析

有意である

臨床的に有効
であると判断

有意でない

臨床的に有効
であるとはいえ
ないと判断

*1：文献・これまでの臨床的知見から設定(第3者が見ても合意される臨床的効果)

症例数設定の一例

論文での記述例

Phase II study of trastuzumab in combination with S-1 plus cisplatin in HER2-positive gastric cancer (HERBIS-1)

The required sample size was estimated based on a threshold RR of 35% and an expected RR of 50%, 80% power, and an alpha value of 0.1 (one-sided) using the binomial test. Given 2% of ineligible patients, the target sample size was determined to be at least 50 patients.

(Kurokawa et al., British Journal of Cancer, 110, 1163-68, 2014)

■ 上記記載の要約

試験デザイン：Phase II Study (単アーム)

主要評価項目：奏功割合 (RR: Response Rate)

閾値奏効率：35%， 期待奏効率：50%

臨床的判断

第1種の過誤(type I error, 有意水準)： $\alpha = 0.1$

1－第2種の過誤(1-type II error, 検出力)： $1 - \beta = 0.8$ (80%)

統計的判断
必要症例数 = 50

試験デザインでの留意点

■ 閾値(帰無仮説)には根拠が必要

単群試験では、閾値に対して臨床的な根拠が必要であり、統計解析は閾値よりもポジティブな結果か否かが評価される。ランダム化比較試験では、帰無仮説は「差がない(or $HR=1$)」になる。

HERBIS-1では、ToGA試験(Y-J Bang, et al. *The Lancet*, 28, 687 - 697, 2010.)の結果を参考に行っている。ToGA試験では、化学療法群(5-FU /CDDP あるいはCapecitabine /CDDP)での**奏功率が35%**であった。

■ 期待値(対立仮説)には実現可能性及び臨床的有効性に基づいて設定

期待値は、臨床的に合意されるような有効性(エフェクトサイズ)として設定される。したがって、期待値が本臨床試験のエンドポイントの目標値となる。

ToGA試験では、(1) Trastuzumab併用群の**奏功率が47%**であること、(2) HER2による選択基準(IHC 3+ or FISH positive + IHC 2+)、(3) 本邦の状況を鑑みて、閾値奏功率を50%と設定している。

第1種の過誤(α エラー)・第2種の過誤(β エラー)とは

	検定により評価	標本サイズで規定
	H_0 が正しい [H_1 : 偽]	H_1 が正しい [H_0 : 偽]
H_0 を棄却 [H_1 を受容]	第1種の過誤 α (H_0 が正しいのに棄却)	検出力 $1-\beta$ (H_0 が正しいので受容)
H_0 を受容 [H_1 を受容]	正しい判断 $1-\alpha$ (H_1 が正しいので棄却)	第2種の過誤 β (H_1 が正しいのに受容)

- 第2種の過誤が一定水準未満(β 未満)になるように、標本サイズを規定する (本当はpositiveなのに本試験を通じてnegativeと判断する可能性を低くする)。
- 統計解析の結果、第1種の過誤が一定水準未満(α 未満)であるか否かを確認する(本当はnegativeなのに本試験結果を通じてpositiveと結論付けるエラーがかなり低いことを確認する)。

必要症例数設定のプロセス

STEP.1：主要評価項目(主たるアウトカム)を考える

- 本研究におけるクリニカルクエスチョンに答えられるものか(妥当性)？
- 当該分野においてAgreeされるものか(信頼性)？
- エンドポイントを取得できるか(実施可能性)？
一例えば、抗がん剤治療において標的病変がなければ奏功率が測定できない等

STEP.2：仮説を決定する

- 無作為化比較試験の場合には非劣勢・優越性・同等性を検討する.
- 単アーム試験の場合には閾値を文献等から決定する.
- 本プロトコル治療における主要評価項目での有用性を検討する.
量的変数の場合：共通の分散(標準偏差)を決定する(個体差).
生存時間の場合：受け入れ期間, フォローアップ期間を決定(censoring割合)

STEP.3：有意水準・検出力を決定する

単アーム： $\alpha=0.05$ or 0.10 , $1-\beta = 0.8$

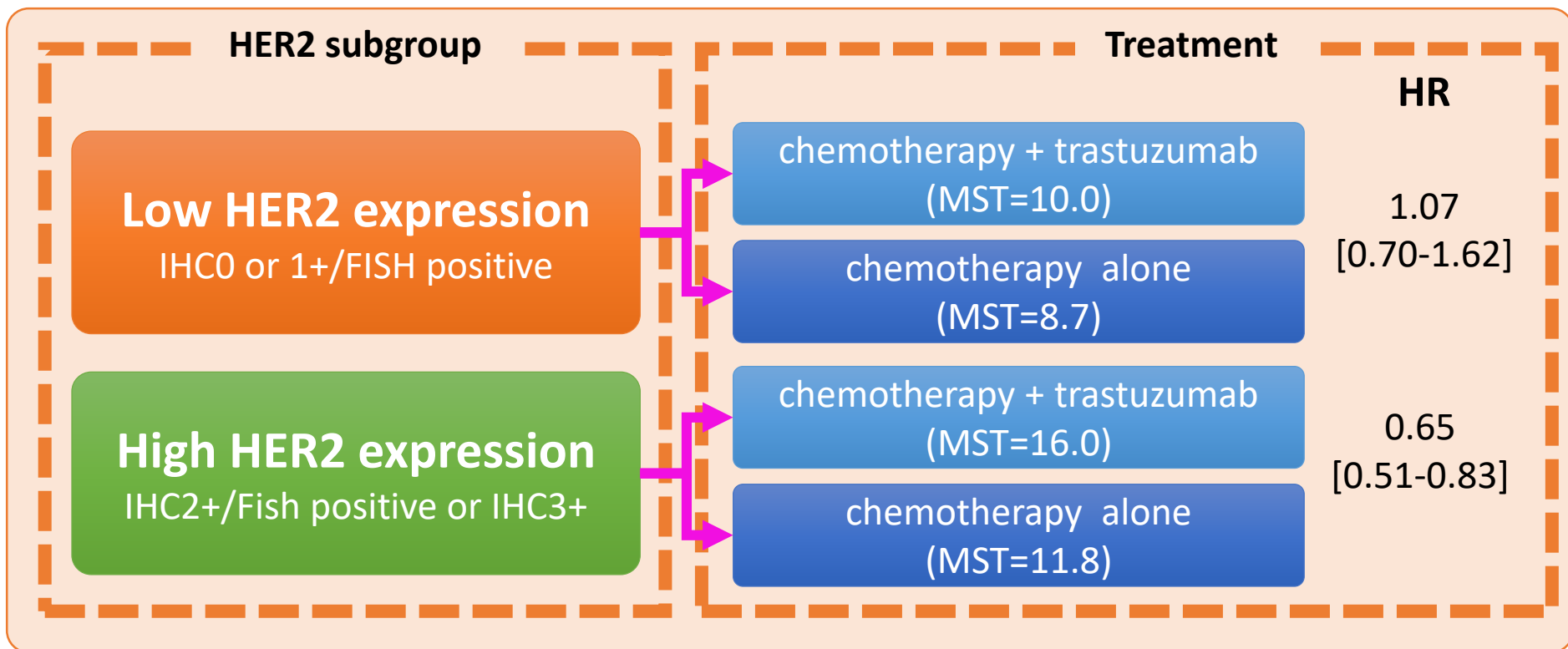
多アーム： $\alpha=0.05$, $1-\beta = 0.80$ or 0.90

交互作用とは(1/2) : ToGA study

ToGA study(Bang et al.: *The Lancet*, 376, 687-697, 2010)の一文

There was evidence of a **significant interaction test** ($p=0.036$) between treatment and the two HER2 subgroups (high HER2 expression vs low HER2 expression).

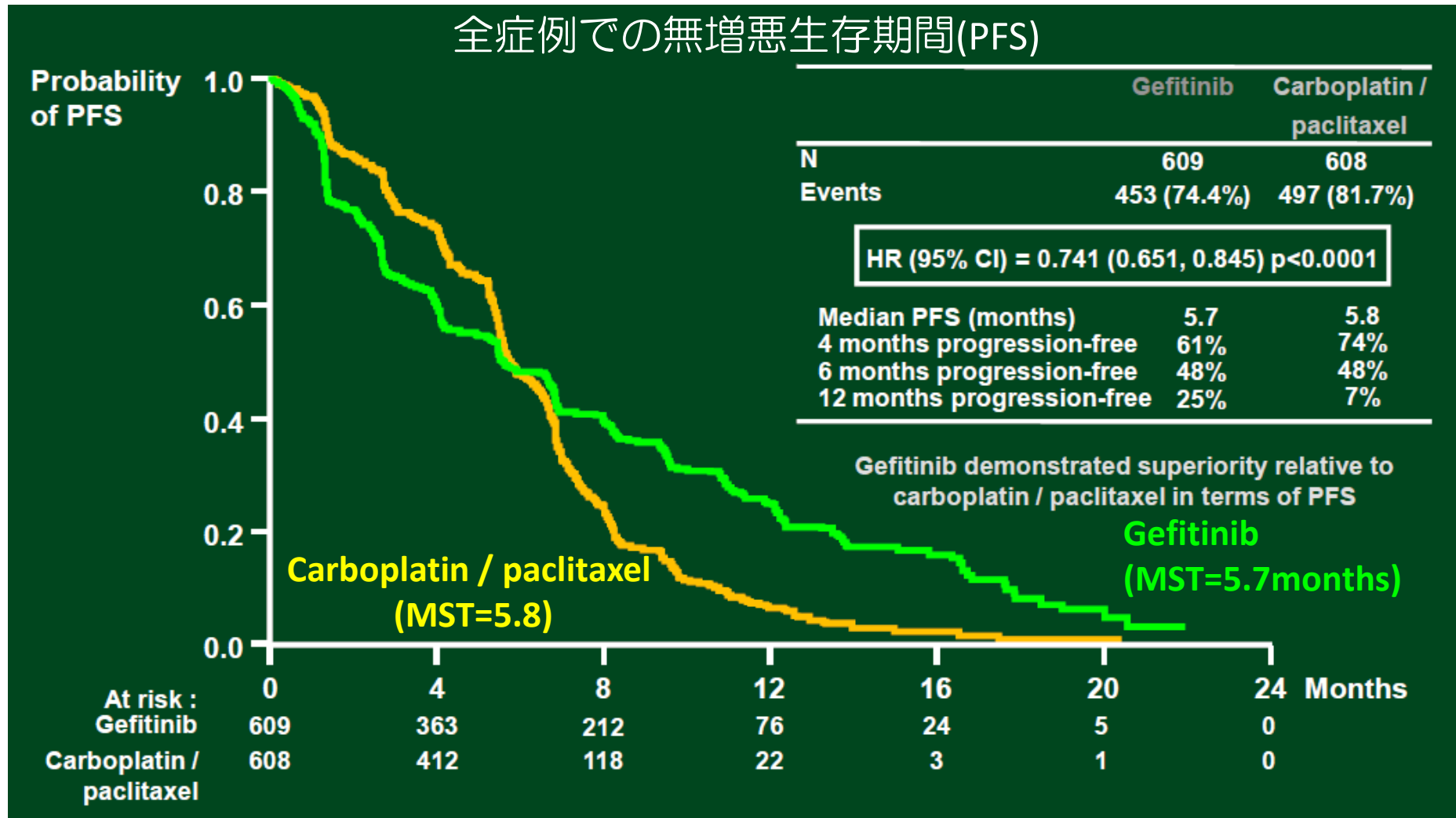
切除不能な局所進行性、再発 and/or 転移性のHER2陽性の患者に対する化学療法単独群（5-FU+CDDP (FP療法)あるいはCapecitabine +CDDP(XP療法)）と化学療法+Trastuzumab併用群を比較する国際共同無作為化非盲検試験



Low HER2ではハザード比(HR)が有意でないが, High HER2では有意に chemotherapy + trastuzumabのハザード比が低い.

交互作用とは (2/2) : IPASS Study

IRESSA Pan-Asia Study(IPASS)試験は、アジア地域における非盲検ランダム化並行群間比較試験である。そこでは、進行非小細胞肺癌患者1,217例を対象に1次治療としてのGefitinib(イレッサ)の有効性・安全性および忍容性をCarboplatin+Paclitaxel併用化学療法と比較した試験である。

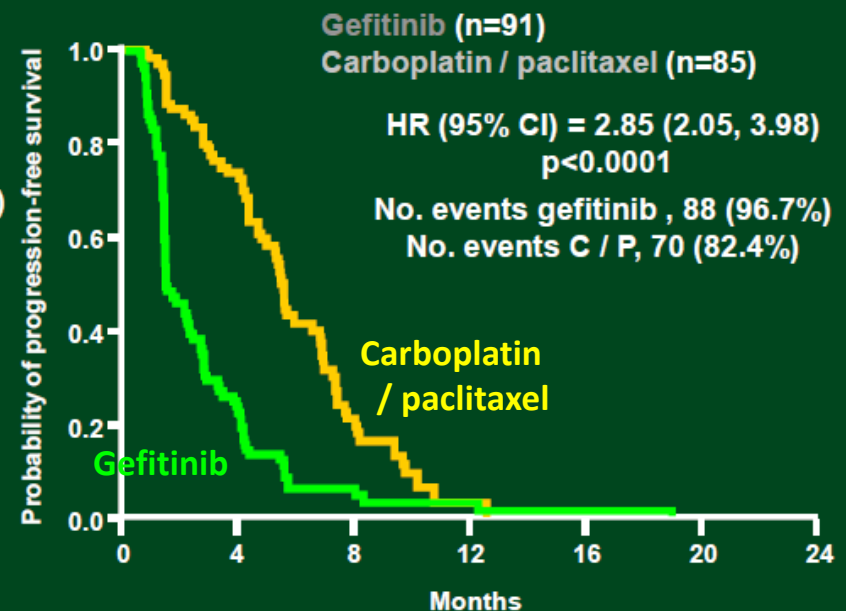
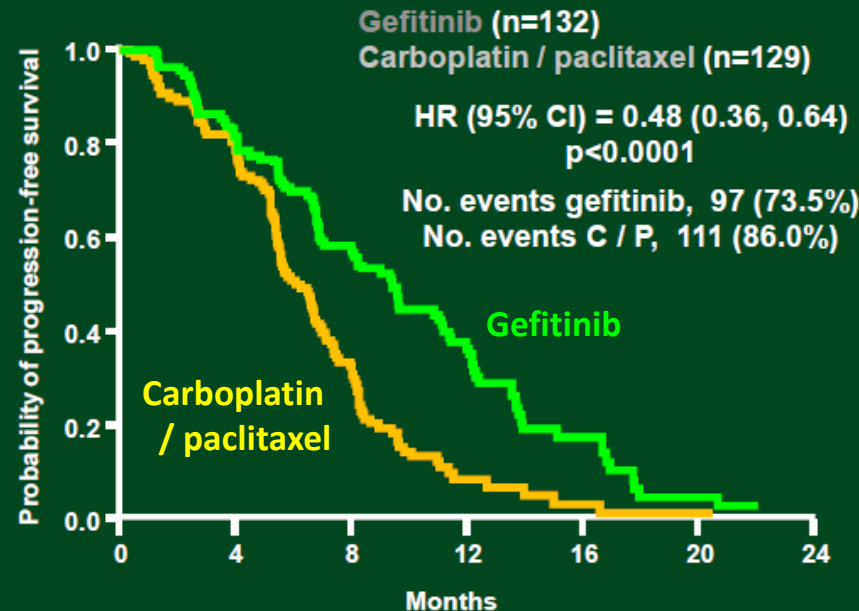


Mok, T.S. et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma, N.Engl.J Med., 361(10), 947-57, 2009.

EGFRで分けると...

EGFR mutation+

EGFR mutation-



Treatment by subgroup interaction test, p<0.0001

無作為化比較試験を行う上でのガイドライン：CONSORT声明

CONSORT(Consolidated Standards of Reporting Trials：臨床試験報告に関する統合基準)とは，無作為化比較試験の報告の質の向上について記載された基準であり，論文とともにチェックリストを提出することが義務付けられるようになってきている．最新版は「CONSORT2010」

□ **CONSORT2010 Check List**

無作為化比較試験を報告する際に記載すべき情報のチェックリストになっている．本チェックリストを埋めるには，適切なプロトコルを作成することが必須．

□ **CONSORT Flow Diagram**

被験者の登録・割付・追跡・解析に至るすべての状況において除外された症例とその理由等を記載するためのダイアグラムである．臨床試験の適切な運用が求められる．

CONSORT Checklist

タイトル・抄録 (Title and Abstract)

- | | | |
|--|----|--|
| | 1a | タイトルにランダム化比較試験であることを記載 |
| | 1b | 試験デザイン(trial design), 方法(method), 結果(result), 結論(conclusion)の構造化抄録 |

はじめに (Introduction)

- | | | |
|-------------------------------------|----|-------------------------|
| 背景・目的
(Background and Objective) | 2a | 科学的背景と論拠 (rationale)の説明 |
| | 2b | 特定の目的または仮説 (hypothesis) |

方法 (Method)

- | | | |
|--------------------------|----|---|
| 試験デザイン
(Trial Design) | 3a | 試験デザインの記述 (並行群間, 要因分析など), 割付け比を含む |
| | 3b | 試験開始後の方法上の重要な変更(適格基準 eligibility criteriaなど)とその理由 |
| 参加者
(Participant) | 4a | 参加者の適格基準 (eligibility criteria) |
| | 4b | データが収集されたセッティング(setting)と場所 |
| 介入(Intervention) | 5 | 再現可能となるような詳細な各群の介入. 実際にいつどのように実施されたかを含む |

アウトカム (Outcome)	6a	事前に特定され明確に定義された主要・副次的アウトカム評価項目、 いつどのように評価されたかを含む。
	6b	試験開始後のアウトカムの変更とその理由。
症例数 (Sample size)	7a	どのように目標症例数が決められたか。
	7b	あてはまる場合には、中間解析と中止基準の説明。

ランダム化 (Randomization)

順番の作成 (Sequence generation)	8a	割振り(allocation)順番を作成(generate)した方法
	8b	割振りのタイプ: 制限の詳細(ブロック化, ブロックサイズなど)
割振りの隠蔽機構 (Allocation concealment mechanism)	9	ランダム割振り順番の実施に用いられた機構(番号付き容器など), 各群の割付けが終了するまで割振り順番が隠蔽されていたかどうかの記述
実施 (Implementation)	10	誰が割振り順番を作成したか, 誰が参加者を組入れ(enrollment)たか, 誰が参加者を各群に割付けた(assign)か
ブラインディング (Blinding)	11a	ブラインド化されていた場合, 介入に割付け後, 誰がどのようにブラインドかされていたか(参加者, 介入実施者, アウトカムの評価者など)
	11b	関連する場合, 介入の類似性の記述
統計学的手法 (Statistical method)	12a	主要・副次的アウトカムの群間比較に用いられた統計学的手法
	12b	サブグループ解析や調整解析のような追加的解析の手法

結果 (Results)

参加者の流れ (Participant flow) (フローチャートを強く推奨)	13a	各群について、ランダム割付けされた人数、意図された治療を受けた人数、主要アウトカムの解析に用いられた人数の記述
	13b	各群について、追跡不能例とランダム化後の除外例を理由とともに記述
募集(Recruitment)	14a	参加者の募集期間と追跡期間を特定する日付
	14b	試験が終了または中止した理由
ベースライン・データ (Baseline data)	15	各群のベースラインにおける人口統計学的(demographic)、臨床的な特性を示す表
解析された人数 (Number analyzed)	16	各群について、各解析における参加者数(分母)、解析が元の割付け群によるものであるか
アウトカムと推定 (Outcome and estimation)	17a	主要・副次的アウトカムのそれぞれについて、各群の結果、介入のエフェクト・サイズの推定とその精度(95%信頼区間など)
	17b	2項アウトカムについては、絶対エフェクト・サイズと相対エフェクト・サイズの両方を記載することが推奨される
補助的解析 (Ancillary analysis)	18	サブグループ解析や調整解析を含む、実施した他の解析の結果。事前に特定された解析と探索的解析を区別する
害 (Harm)	19	各群のすべての重要な害(harm)または意図しない効果

考察 (Discussion)

限界(Limitation)	20	試験の限界, 可能性のあるバイアスや精度低下の原因, 関連する場合は解析の多重性の原因を記載
一般化可能性 (Generalisability)	21	試験結果の一般化可能性 (外的妥当性, 適用性)
解釈(Interpretation)	22	結果の解釈, 有益性と有害性のバランス, 他の関連するエビデンス

その他の情報 (Other information)

登録(Registration)	23	登録番号と試験登録名
プロトコル (Protocol)	24	可能であれば, 完全なプロトコルの入手方法
資金提供者 (Funding)	25	資金提供者と他の支援者 (薬剤の供給者など), 資金提供者の役割

疫学研究を行うためのガイドライン：STROBE

観察的疫学研究報告の質改善のための声明(STROBE: Strengthening the Reporting of Observational Studies in Epidemiology)とは、観察的疫学研究の研究報告の質の向上と結果の透明性を目指して作られた。

□ STROBE Checklist

コホート研究，ケースコントロール研究，及び横断研究の各デザインについて，論文の中に記載すべき内容に関してのチェックリストである．全体で22項目が存在し，そのうちの18項目はすべてのデザインの共通項目になっている．そして，4項目については，上記の3種類のデザインによって異なる．

STROBEの解説と詳細のなかには，具体的な観察的疫学研究の事例とともに，記載内容を明記していることから，研究を整理するうえでも非常に参考になる．

EZRとは



自治医科大学附属さいたま医療センター 血液科の神田善伸教授が統計解析環境RのGUI環境であるRコマンダーをカスタマイズしたものがEZRである。

EZRでは、医学統計学における諸種の統計的方法をマウスのみで実行できる。

ダウンロードは、ブラウザで「EZR」と入力すれば、左図のホームページがトップにくる。

<http://www.jichi.ac.jp/saitama-sct/SaitamaHP.files/statmed.html>

Warning!!

MacOSでは、EZRのもととなるRcommanderを動かすために、x11というプラグインが必要になる。そのため、EZRをインストールする前に、

<https://www.xquartz.org/>

から、x11をダウンロードおよびインストールする必要がある。

傾向スコアマッチングの身近な適用：EZRを用いた場合

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

最初の列は変数名

PSexample.csv

	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

☒ ピリオド[.]

☐ カンマ[,]

ヘルプ

OK

キャンセル

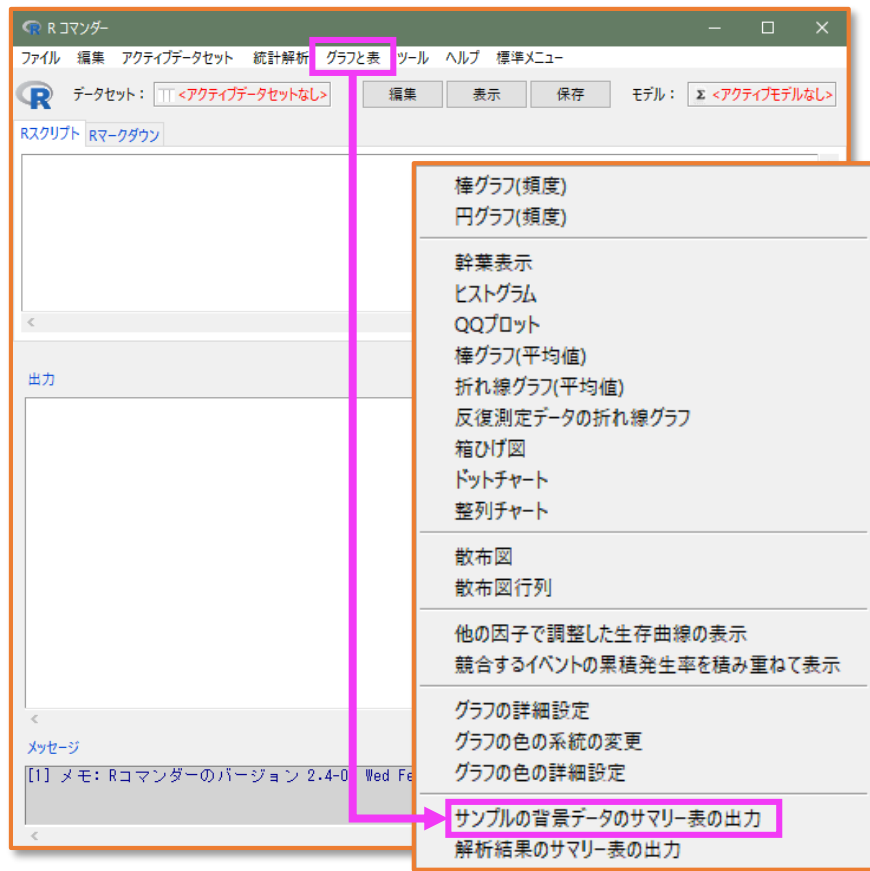
最初の列が変数名ならば、チェックを入れる。

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。

データの概要



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)

結果

```
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 (%) |      |      |      |      |
  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

> 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	C	
		71	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。

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)

R 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

層別化変数 (変数名を##と入力)

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。

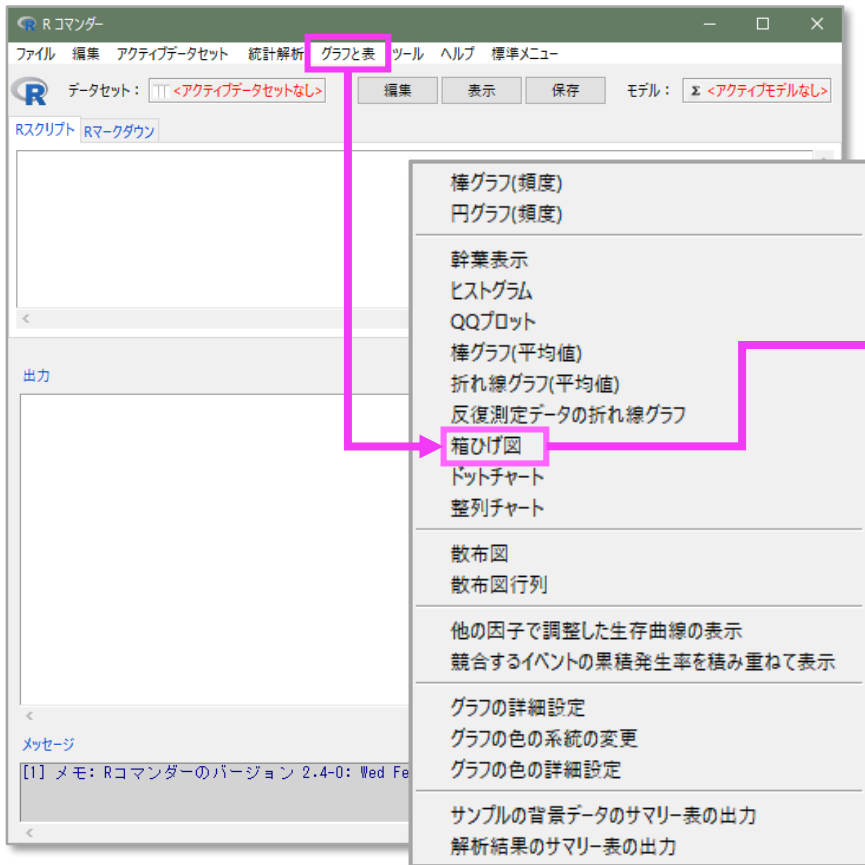
↓ 一部のサンプルだけを解析対象にする場合の条件式。例: age>50 & Sex==0 や age<50 | Sex==1

<全ての有効なケース>

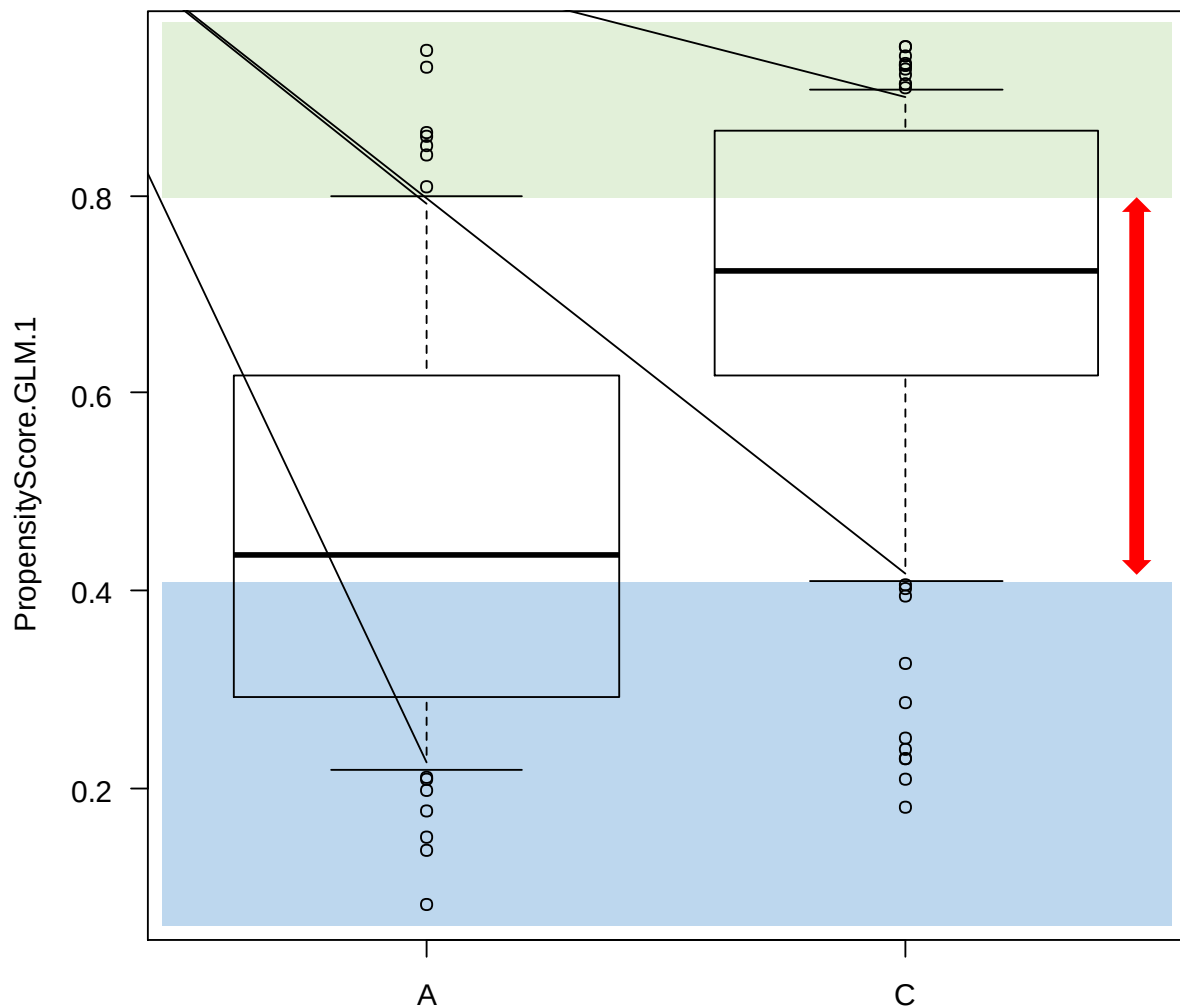
ヘルプ リセット OK キャンセル 適用

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



箱ひげ図



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。

マッチングを用いた解析

[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 with the command `install.packages("Matching")` entered in the script editor. A flowchart with pink boxes and arrows guides the user through the steps: 1. '入力する' (Enter) points to the command input. 2. '入力したコマンドをドラックする' (Drag the entered command) points to the command text. 3. '実行ボタンを押す' (Press the execution button) points to the '実行' (Run) button. 4. '質問はすべてOK' (All questions are OK) points to the '実行' button. 5. '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]) points to the '実行' button.

```
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[質問はすべてOK]; D --> E[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 with the command `library(Matching)` entered in the script editor. A flowchart with pink boxes and arrows guides the user through the steps: 1. '入力する' (Enter) points to the command input. 2. '入力したコマンドをドラックする' (Drag the entered command) points to the command text. 3. '実行ボタンを押す' (Press the execution button) points to the '実行' (Run) button.

```
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];
```

EZR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

傾向スコアの値

0.20～0.25の間で自由に設定

[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



マッチング結果

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

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)

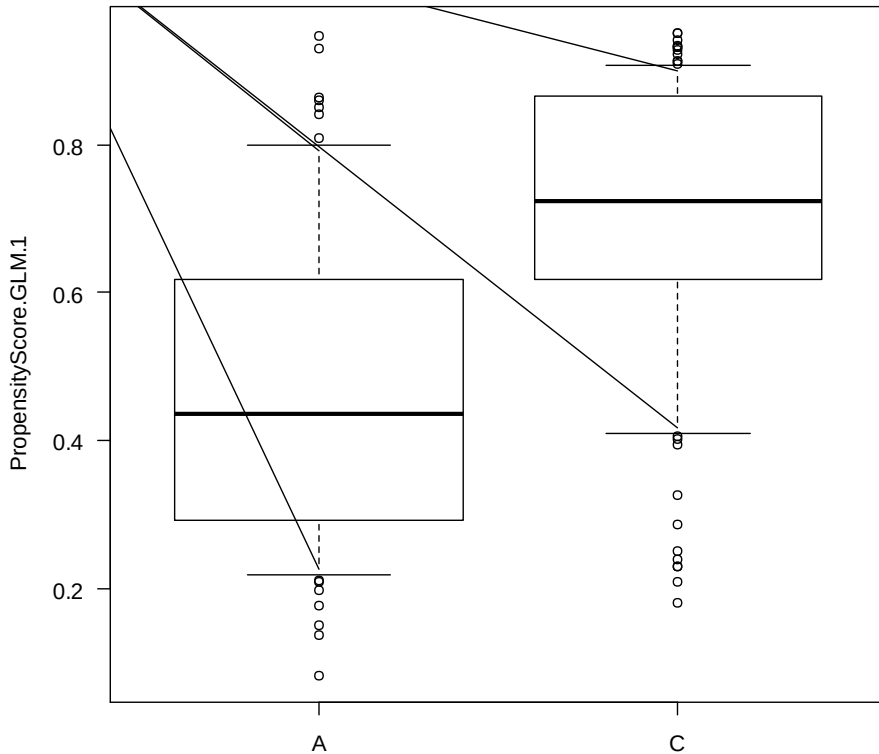
- control
- Dataset
- Match.data
- odds
- TempDF
- treat

OK キャンセル

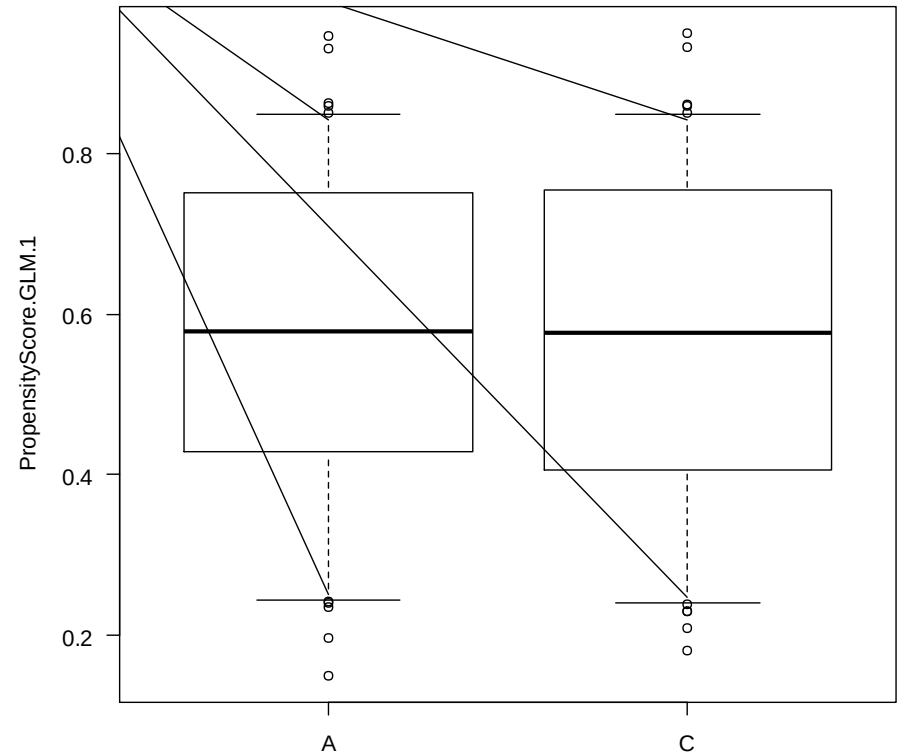
Match.dataを選択してOKボタンを押す

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

主要評価項目の評価

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 ▶	市
検査の正確度の評価 ▶	信頼区間の計算
マッチドペア解析 ▶	比率の検定
メタアナリシスとメタ回帰 ▶	率の差の信頼区間の計算
必要サンプルサイズの計算 ▶	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)

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

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。