

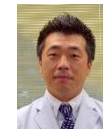


この資料はこんな本達を参考に作成しました



医学研究初心者のための やっぱりわかりにくい統計道場

Shingo Hatakeyama



統計の難しさ①

- ▶ なぜ、難しいのか？

それは **言葉** と **式** が難しく「意味不明」だからです。



- ▶ 正規分布、分散、標準偏差、対応のある・・・、パラメトリック・・・などなど、**???**の嵐ですね。
- ▶ しかし、今の医学には統計はなくてはならない手段です。
- ▶ 理解しようとする、戦意喪失しますが、道具として利用する分には、統計はそんなに難しいものではありません

統計の難しさ②

- ▶ まずは見た目からやる気を失う「統計用語」を受け入れ、便利な道具と認識しましょう。
- ▶ 深入りせず簡単に覚えるには、このデータにはこの解析、と1:1対応で暗記することです。
- ▶ そして最も重要なのは「必要に迫られること」です。
 - ▶ これを書いている私も、数年前までは統計音痴でしたが、必要に迫られ、あれこれ本を買いあさりました。
 - ▶ 統計マニアではありませんので、多少の間違いはあると思いますが、その時はお許してください。

統計解析で言えること、とその限界

- ▶ よく言われるように、統計は **マジック** です。
- ▶ ふさわしい解析方法を選ぶことが重要ですが、どんなデータにどんな解析がいいのか？それを知ることが先決です。
- ▶ 同じデータでもちょっと解析を変えると有意差が出たりします。
- ▶ その有意差を信じていいのか？、を考えられるようになれば統計の限界と使い方がわかります。

統計解析で言えること、とその限界

- ▶ 患者のデータを扱う上で最も注意すべきことは、統計とは「**それが真実かどうか、をもっともらしく数学的に説得する方法**」ではないところです。
- ▶ 数学的に正しい、と医学的に正しいはイコールではありません。
- ▶ 細胞やマウスの実験では統計は力を発揮しますが、人体実験をしない限り人での真実はわかりません。

- ▶ しかし、現在のところこの方法しか説明する手法がないので使っています。「**数学的**」には間違いではありません。

臨床データにおける統計的有意差の意味

- ▶ 仮に、体重を減らすA薬があったとします。極端な例ですが、臨床試験で表のような結果になったとします。
- ▶ この2群間には統計的に有意差が出たとしても、残念ながら、体重をたった0.1kg減らす薬剤は実際あまり意味のない薬＝臨床的意義のない薬です。他の因子の影響（交絡因子：Bias）も十分考えられます。
- ▶ 数学的な意味を臨床現場での意味に変換する作業が我々にとって重要であり、そのための道具の一つが統計です。統計的有意差のマジックに注意しましょう。

	プラセボ	A薬	<i>P value</i>
体重平均	50 ± 1kg	49.9 ± 1kg	<i>P</i> < 0.05

統計解析と実臨床のギャップ

- ▶ **数学的に正しい** \neq **医学的に正しい**
- ▶ 統計は実学：現実で起きていることをすべて数式で証明できるほど、我々は世の中を知っていません
- ▶ もっとも良く当てはまる統計モデルを使用しましょう

観察結果（データ）の考え方

- ▶ 若手の先生たちによく聞かれます。

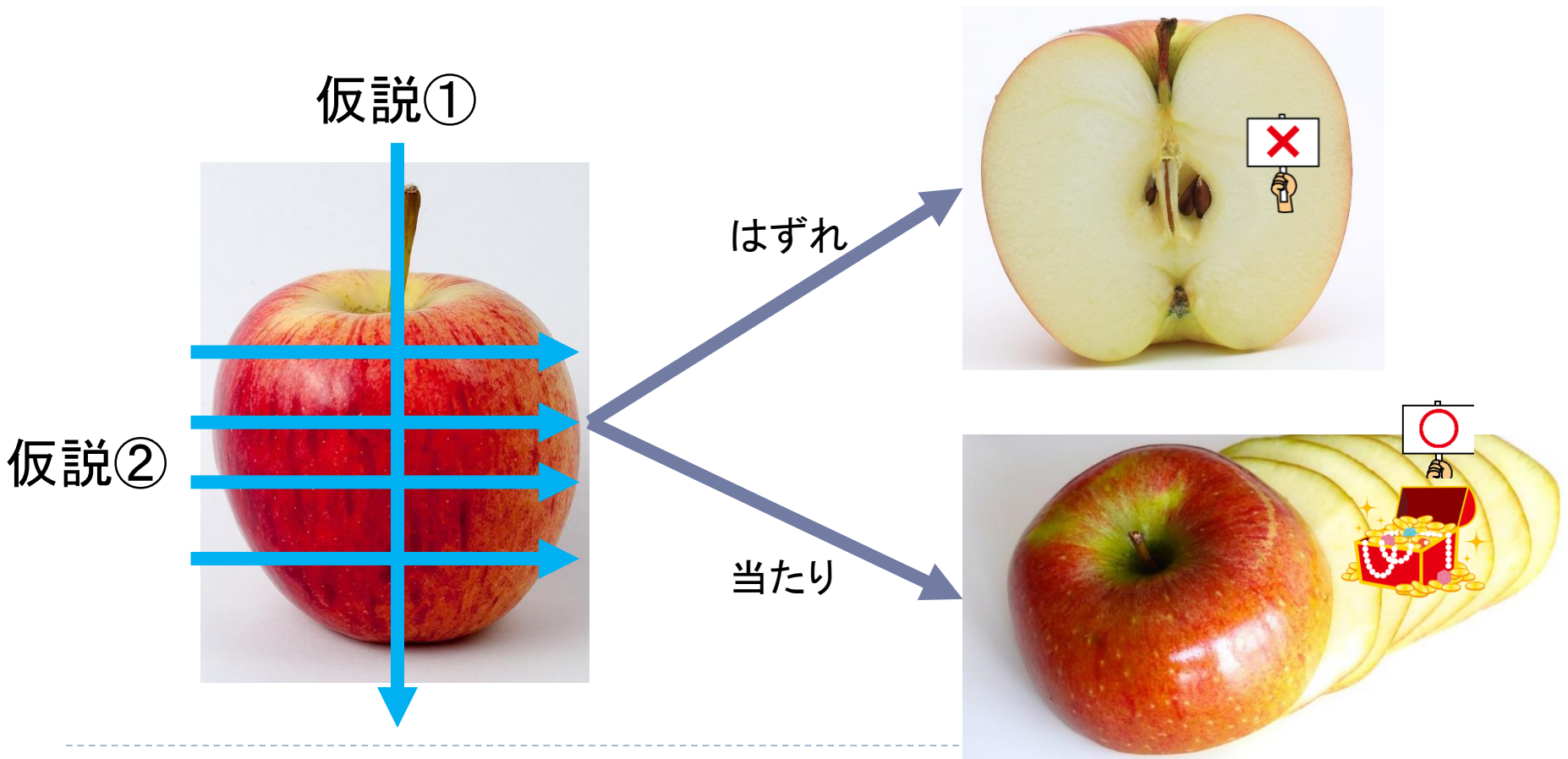
P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
0	2.816	2.044	2.380	0	1.300	1.589	2.389	1.397	3.313
0	3.567	3.071	2.561	0	1.192	1.576	2.420	2.077	2.949
0	1.174	1.660	0.779	0	1.104	0.712	1.387	1.013	1.888
0	2.511	3.244	1.737	0	1.893	1.241	1.661	1.527	2.857
0	4.530	2.071	3.296	0	1.252	1.584	2.378	2.558	4.654
0	3.531	3.462	2.593	0	1.424	1.425	1.915	2.519	4.083
0	3.305	4.184	4.119	0	1.610	2.087	2.162	4.770	3.796
0	3.331	2.983	2.264	0	1.398	1.882	3.321	4.119	3.066
0	3.885	2.689	3.421	0	1.406	0	2.213	2.968	5.531
0	2.466	1.679	2.078	0	1.159	1.431	2.149	1.735	2.749
0	3.997	2.186	2.182	0	1.555	1.563	2.703	2.514	4.042
0	2.869	3.228	3.250	0	2.485	1.746	2.256	3.533	3.250
0	3.155	2.278	2.901	0	1.214	1.360	1.799	2.912	4.520
0	2.742	3.178	2.335	0	1.502	1.821	2.090	2.159	3.091
0	12.754	4.395	5.705	0	1.548	1.899	3.511	3.729	8.879
0	8.113	2.245	6.902	0	1.328	2.002	2.136	2.080	3.943
0	6.033	3.611	4.467	0	2.160	1.634	3.005	3.202	5.383
0	1.763	3.653	1.735	0	1.865	1.356	1.796	3.488	2.646
0	2.166	3.216	2.155	0	1.589	1.356	2.203	3.442	3.057
0	2.667	2.346	2.314	0	1.643	1.472	1.547	1.324	2.841
0	2.877	1.881	1.434	0	1.115	1.330	1.809	3.360	3.506
0	2.858	1.145	1.681	0	1.384	1.532	1.276	1.525	2.041
0	4.478	3.746	4.186	0	1.657	0	2.356	3.022	6.742
0	1.888	4.152	2.996	0	1.575	1.475	2.399	4.359	3.504
0	2.802	2.174	2.961	0	1.526	1.723	2.072	2.503	2.664

統計的手法はわかりました、
でもデータをどう解析して
いいかわかりませんか???

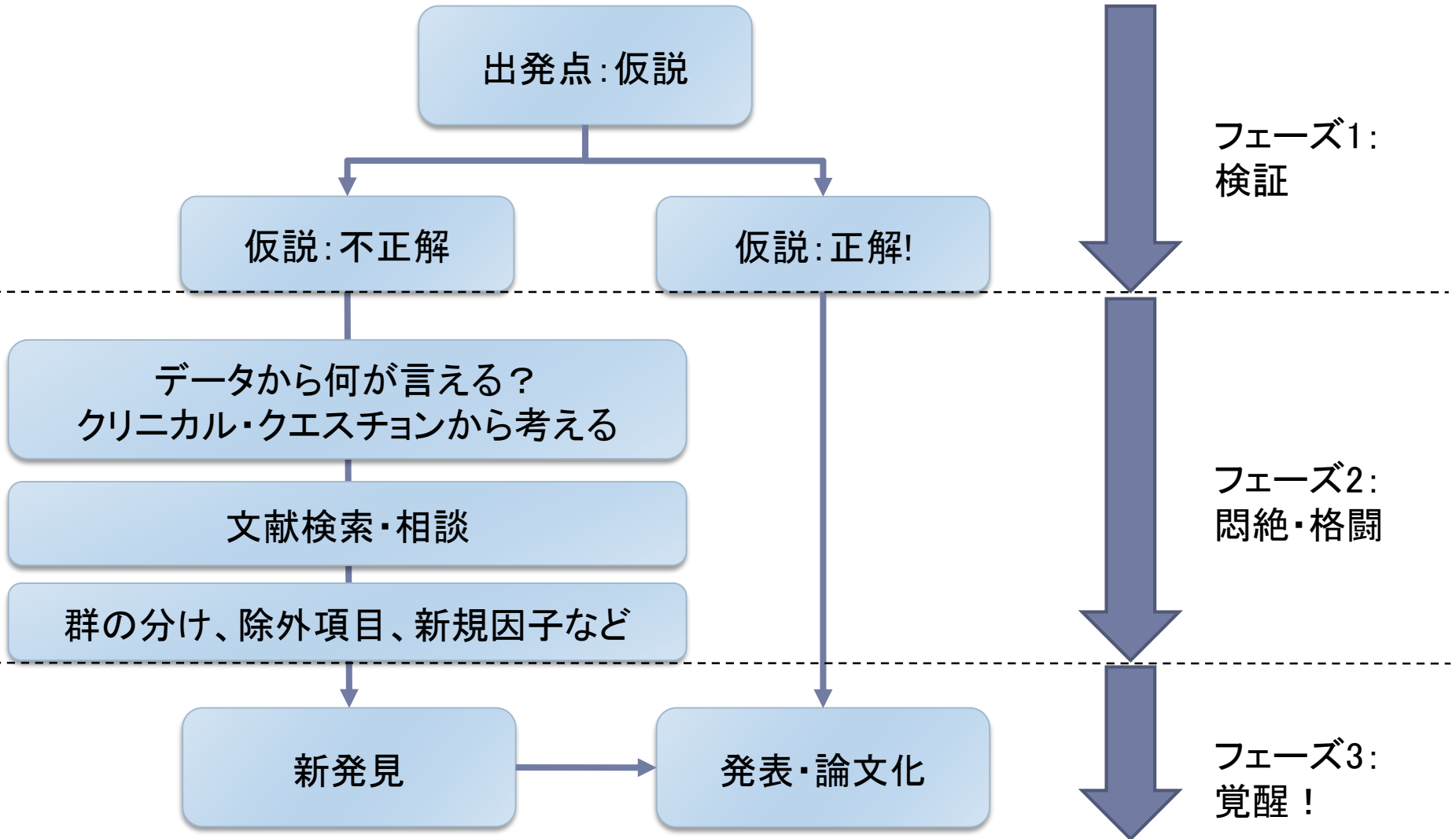


Answer : 試行錯誤して悩みましょう

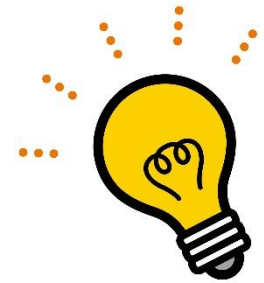
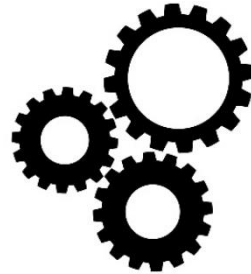
- ▶ データをよく観察し、隠された「答え」を発見する



答えを発見する方法



試行錯誤の積み重ねしかありません！



群間の比較法

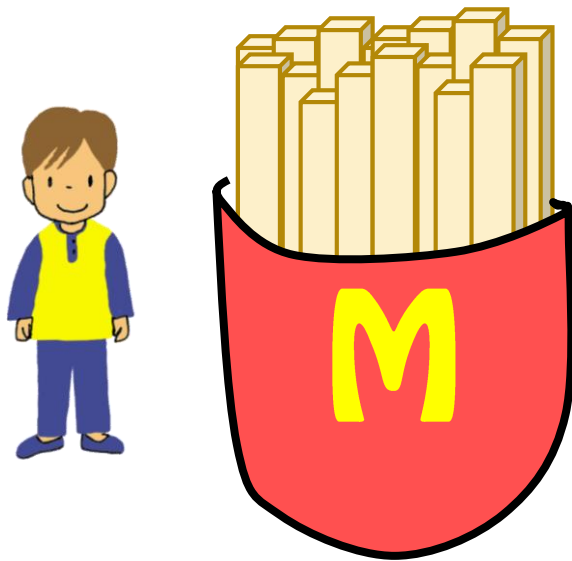
▶ 解析のポイント

比較 : Comparison

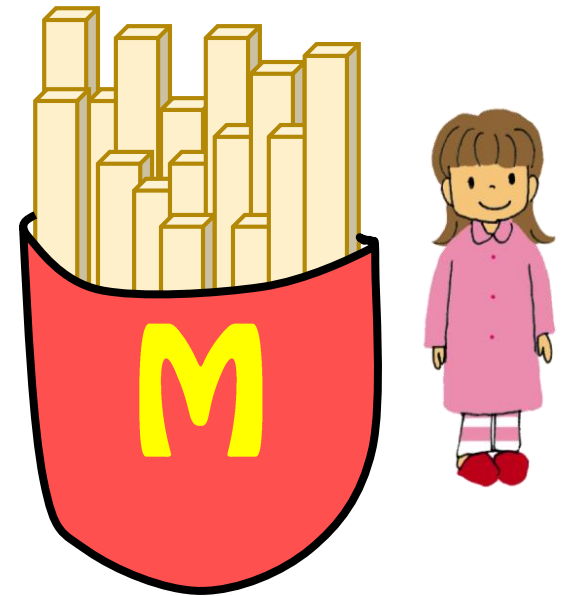
▶ 比較するから、何かと言える

どっちのFried Potatoがお得？（平均とSD）

- ▶ A君とBさんが某M店でポテトを買いました。Bさんの方に長いポテトが多い気がしますが、実際はどちらが長いポテトが多く得をしたのでしょうか？長さを測ってみました。



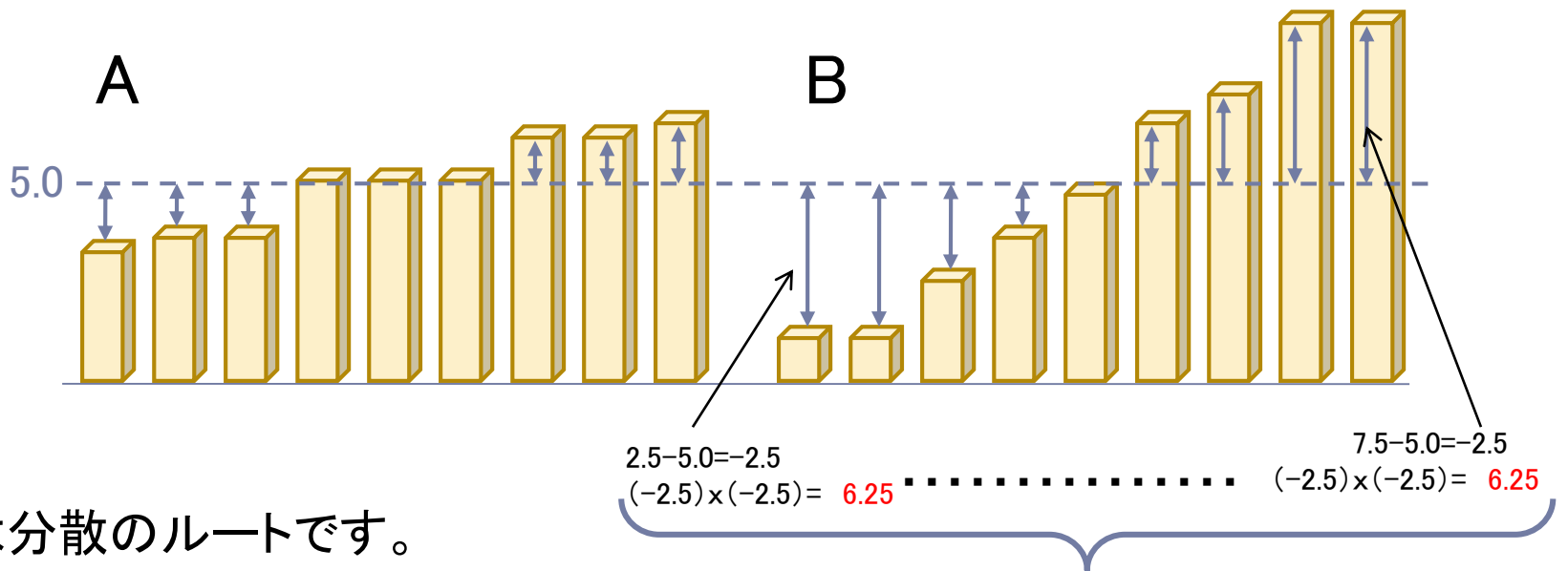
A君	Bさん
7.0	5.0
3.0	7.5
5.0	5.0
4.5	4.5
6.0	8.0
5.0	5.5
4.0	2.5
4.5	2.5
6.0	2.0
5.0	7.5
平均5.0	平均5.0



- ▶ すると、どちらのポテトも平均は同じでした。損得は無い様に見えます。しかし、この2群は数学的には同じとはいえません。バラツキが違うからです。バラツキを表す数値が分散と標準偏差(SD)です

バラツキの指標：分散と標準偏差SD

1. 各ポテトの長さと平均との差を出します。(バラツキを数値化)
2. その差を2乗します(プラスにする)
3. それを合計し(n-1)で割ります(ばらつきの平均値=分散)
4. それをルートして2乗した分を戻します(SD)



- ▶ SDは分散のルートです。
- ▶ 分散は(各数値の平均からの差の2乗の合計)/n-1です。

$$\text{合計} 43.5 / (10-1) = \underline{4.83}, \quad \sqrt{4.83} = \underline{2.20}$$

分散 標準偏差

分散と標準偏差の出し方

▶ 以上をまとめてExcelでやってみると

1. Xの平均を出す (AVERAGE)
2. X-平均を出す
3. それを2乗する
4. それを合計する
5. $(n-1) = 10-1=9$ で割る = **分散**
6. 分散をルートする = **標準偏差** となります。

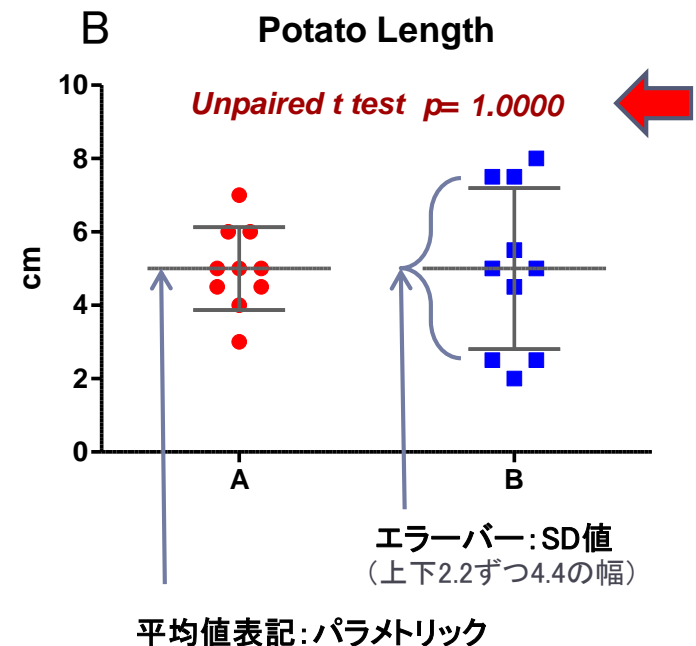
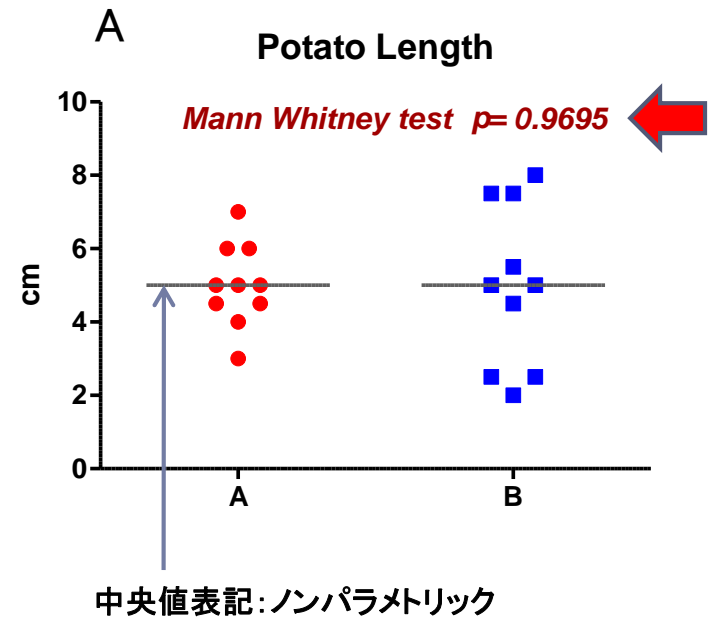
▶ 標準機能で簡単にもできます

1. SDを表示したいマスをクリック
2. 関数からSTDEVを選ぶ
3. SDを出したい範囲を選ぶ
4. リターンをおす、でošimai。

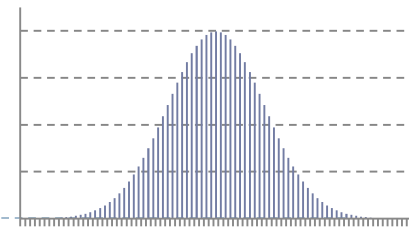
	A			B		
	x	x-5.0	(x-5) ²	x	x-5.0	(x-5) ²
	7	2	4	5	0	0
	3	-2	4	7.5	2.5	6.25
	5	0	0	5	0	0
	4.5	-0.5	0.25	4.5	-0.5	0.25
	6	1	1	8	3	9
	5	0	0	5.5	0.5	0.25
	4	-1	1	2.5	-2.5	6.25
	4.5	-0.5	0.25	2.5	-2.5	6.25
	6	1	1	2	-3	9
	5	0	0	7.5	2.5	6.25
平均	5			5		
合計			11.5			43.5
分散			1.28			4.83
SD			1.13			2.20

標準偏差はばらつき指標

- ▶ 以上の計算により
 - ✓ Aのポテトが持つ情報(平均±SD)は 5.0 ± 1.3 、Bのポテトは 5.0 ± 2.2 となり、Bのポテトの方がばらつきが大きいという結果になります。
- ▶ さて、個々の2つのグループに差があるかどうか、を調べたいとき(検定したいとき)、2群間の比較という方法を行います。
- ▶ この場合はnが少なく母集団が正規分布するかどうかわからないので「対応のないノンパラメトリック検定」(A)を行いました。
 - ▶ 参考までに、「対応のないパラメトリック検定」(B)も記載してあります。



正規分布に従うか、どうか



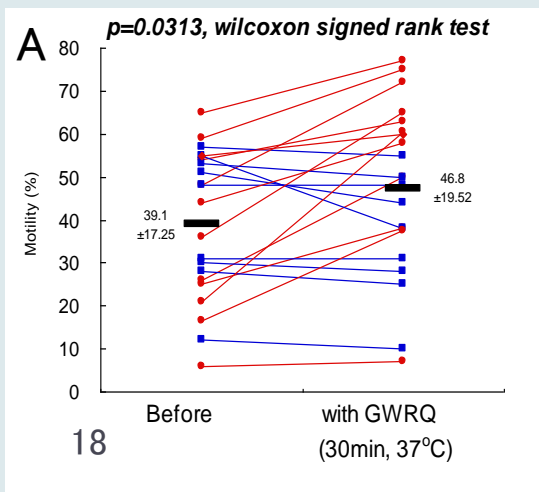
- ▶ パラメトリックは正規分布する、ノンパラメトリックは正規分布しない、という意味です。
- ▶ 厳密にはヒストグラムを描いて正規分布するかどうか、もしくは正規分布の検定をする必要があります。しかし、実際にはデータから大体は予想可能です。
- ▶ **nが少なく、ばらつきが大きければ正規分布しない**
 - ▶ ノンパラメトリックのMann-Whitney 's U testやWilcoxon signed-rank testを選択
- ▶ **nが多くばらつきが少なければ正規分布しやすい**
 - ▶ パラメトリックStudent t-testやPaired t-testが使用可能(有意差がやすい)
- ▶ しかし正規分布に従うか迷う場合はノンパラメトリック解析を選択しましょう。
- ▶ 正規分布するかどうか？の判定は、実際に分布図を書いてみると分かります。こちらにエクセルでの例があります。

http://software.ssri.co.jp/statweb2/sample/example_16.html

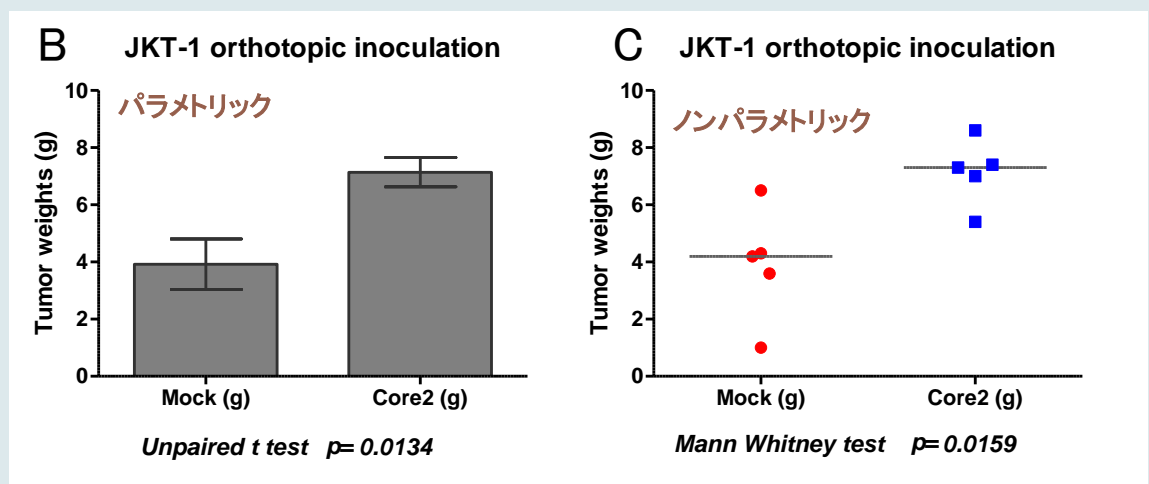
対応のない2群と対応のある2群の意味

- ▶ 「対応のある」「なし」とはどういう意味なのか難しい言葉です。
- ▶ 解りやすく言うと、**同一個体の2種類の観測値を比較検定しているかどうか**、です。していれば「対応がある」ことになります。
 - ▶ A 対応のある2群の例:精子にある薬剤を入れて前後で運動が改善するかを見た実験 (Hatakeyama S, et al. J Urol,2008)
 - ▶ B,C 対応のない2群の例:精巣腫瘍細胞をマウスに植えて大きさを比較した実験 (Hatakeyama S, et al. Int J Cancer, 2008)

対応のある2群の例



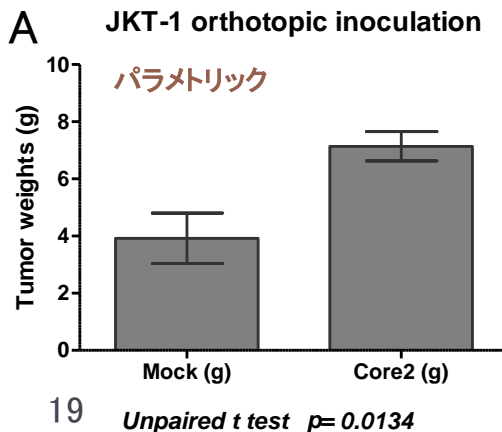
対応のない2群の例(同じデータを2つの方法で解析)



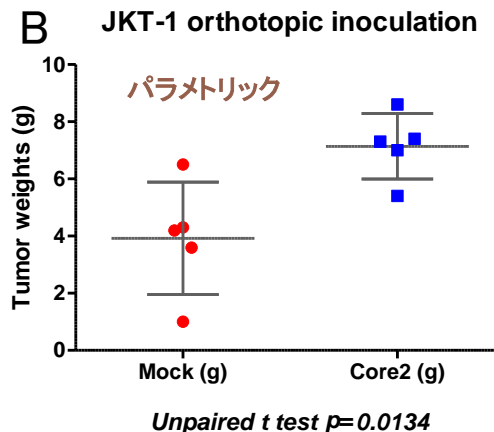
データの表記法について

- ▶ 平均、標準偏差(SD)は正規分布の用語であり、(A)のように棒グラフにエラーバーを表記する場合は集団は正規分布する、という意味なので、パラメトリックのt testが適切です。
- ▶ データのばらつきも表現したいときは(B)のように点グラフにして平均とSDを表示します。
- ▶ データのばらつきが大きく、ある異常値に平均が大きく影響を受けるときは中央値を使います。このときはノンパラ解析をします。
- ▶ 下図の例はどちらでも有意差があり、おそらくnを増やせば正規分布する集団となることが示唆されますが、n=5なので、点グラフ、ノンパラ解析のMWU test(C)が適切です。

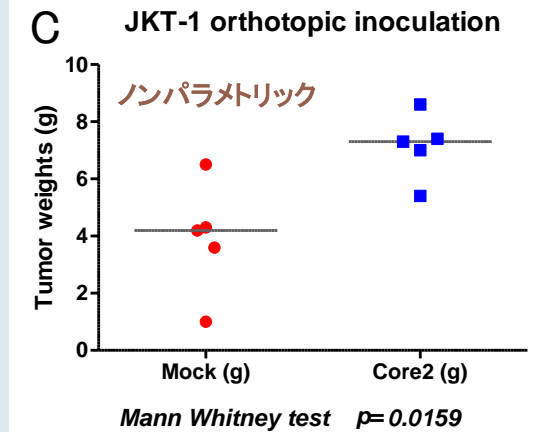
棒グラフのパラメトリック



点グラフのパラメトリック



点グラフのノンパラメトリック



対応のない2群の比較検定

- ▶ 独立した2群のデータに有意差があるか？（棒や点グラフが適切）
- ▶ **Parametric : Student t-test: スチューデントのt検定**
 - ▶ 平均値を比較して検定します。Excel関数で計算可。平均値とSDの棒グラフで表記します。nが多く、ばらつき(2群の分散が一緒)が均一なとき使えます。
- ▶ **Non-parametric : Mann-Whitney 's U test: マン・ホイットニ検定(MWU)**
 - ▶ 中央値を比較して検定します。Excelマクロで計算可。中央値と分布図の点グラフで表記します。nが少なく、ばらつき(2群の分散が一緒)が異なるとき使います。正規分布の適合性が面倒くさいときは、とりあえずこっちで計算できます。

Mann-Whitney's U testを使うとき

- ▶ MWU testは出番が多いのでここで解説。
- ▶ MWU testは、母集団の分布がわからない場合に、データの分布形態を問わずに使うことができる方法です。
- ▶ パラメトリックなデータに対してノンパラ解析を使っても問題はないようです。
 - ▶ MWUはt-testも包括して解析できる方法です。
 - ▶ ただし、データが正規分布とみなすことができる場合は、t-testのほうが、有意差が出やすいようです。
 - ▶ Mann-Whitney 's U testで、「有意差あり」なら、かなりの確率で「有意差がある」と言えるようです。
- ▶ しかし、MWUで「有意差なし」でもt-testで有意差が検出されることがあるので、そういう場合は、母集団の正規分布の検討が必要です。

対応のある2群の比較検定

- ▶ 同一個体に、ある刺激による変化(=差)に有意差があるか？(折れ線グラフが適切)
- ▶ **Parametric: Paired t-test: 対応のあるt検定**
 - ▶ 対応するデータの差の平均値が0からどの程度偏っているかを検定する方法です。Excel関数で計算可。nが多いときには、「対応するデータの差が正規分布」でなくても、使うことができます。極端な値や離散値であり、明らかに前提条件(正規分布に従う連続変数)から離れている場合を除いて、問題が生じることは少ないようです。
- ▶ **Non-parametric: Wilcoxon signed-rank test: ウィルコクサン符号付順位検定**
 - ▶ データの分布形態を問わずに使うことができます。データの分布形態を問わずに使うことができます。しかし、データが正規分布みなすことができる場合は、Paired t-testのほうが、有意差が出やすいようです。n>6は必要。正規分布の適合性が面倒くさいときは、とりあえずこっちで計算できます。

解2乗検定： χ^2 test (chi-square test)

- ▶ 2群間が0-1型の(あり、なし)データの場合、 χ^2 testを用います。
 - ▶ 男女比(男=1、女=0)や免疫染色の結果(陽性=1、陰性=0)など。
- ▶ 2x2分割表に記載できるデータです。
- ▶ Excelマクロでも可能だし、Webでも公開プログラムがあります。
- ▶ お手軽統計マクロ集 Stat macros for Excel(Excel2007でもOKでした)
 - ▶ <http://sci.kj.yamagata-u.ac.jp/~columbo/Stat/>
- ▶ **多機能 WEB 計算機**
 - ▶ http://aoki2.si.gunma-u.ac.jp/calculator/chi_sq_test.html

Control群とAST120群の男女比の検定

男女差	女=0	男=1	合計
Control群	39	66	105
AST120群	37	81	118
合計	76	147	223

~~$p=0.36279$ 、この2群間に男女比の有意差はない~~

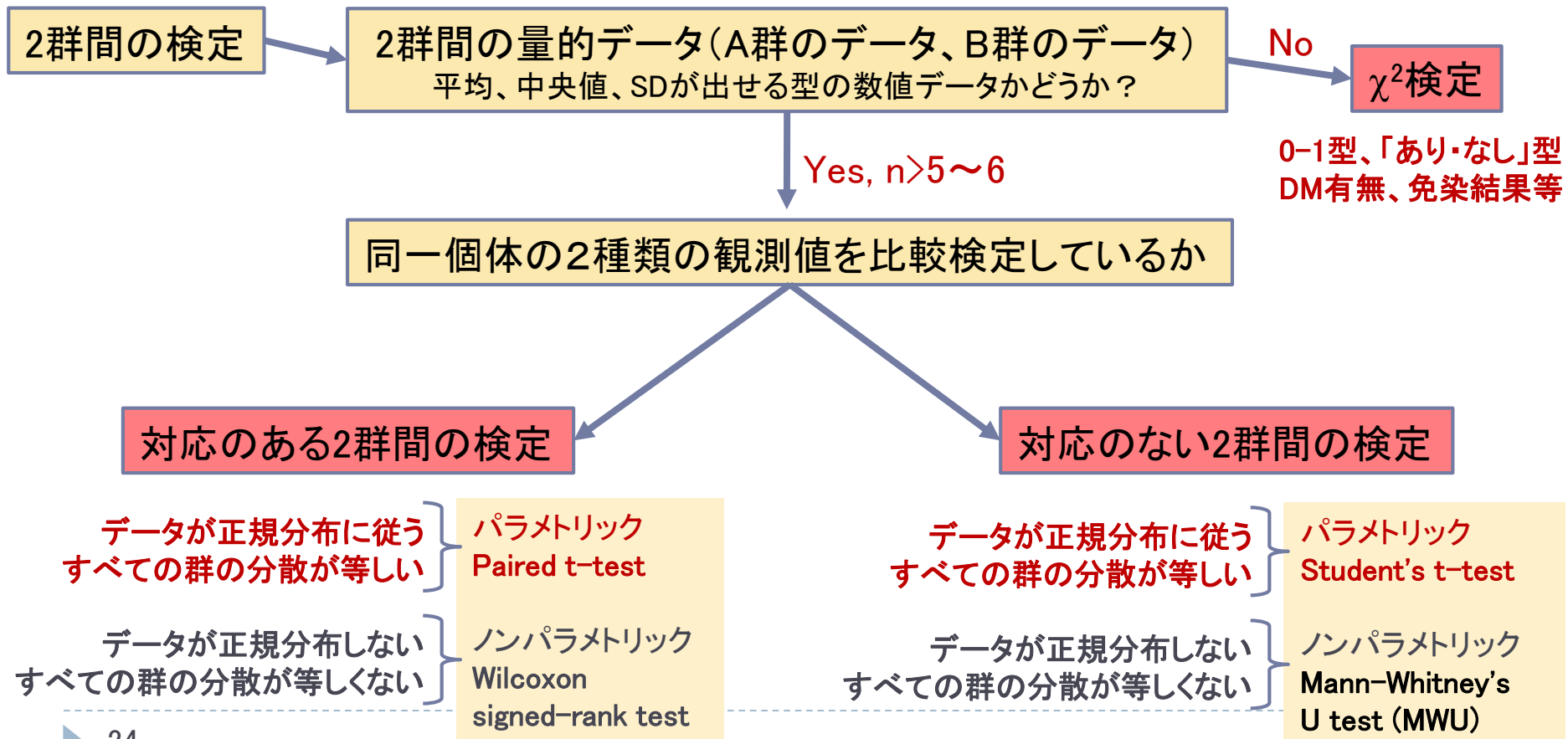
精巣腫瘍におけるStage IとStage II+III のC2GnT1免疫性の検定

免疫染色	陰性=0	陽性=1	合計
Stage I	19	9	28
Stage II+III	3	34	37
合計	22	43	65

~~$p<0.001$ 、この2群間に染色性の有意差はあり~~
= Stage II+IIIでよく染まっている!

2群間の検定法をまとめると・・・

- ▶ 2群間の検定にはデータの種類に応じた解析法があります。以下に模式図として記載します。



3群間の検定

- ▶ 2群間どうしの検定をそれぞれでやってはいけません。
- ▶ 理由は割愛しますが、有意差が出やすくなるからです。
- ▶ 便宜的にやるとすれば、2群同士の検定を各々やり、そのp値を3倍しても $p < 0.05$ なら、有意差があるとされています。
- ▶ きちんとやるには以下の方法があります。
 - ▶ 対応のない3群間の検定
 - ▶ パラメトリック: One way ANOVA
 - ▶ ノンパラメトリック: Kruskal-Wallis test
 - ▶ 対応のある3群間の検定
 - ▶ パラメトリック: One way repeated measures ANOVA
 - ▶ ノンパラメトリック: Friedman test
- ▶ ここではそこまで説明しません。必要なときに調べましょう。

生存分析：Kaplan-Meier法の生存曲線

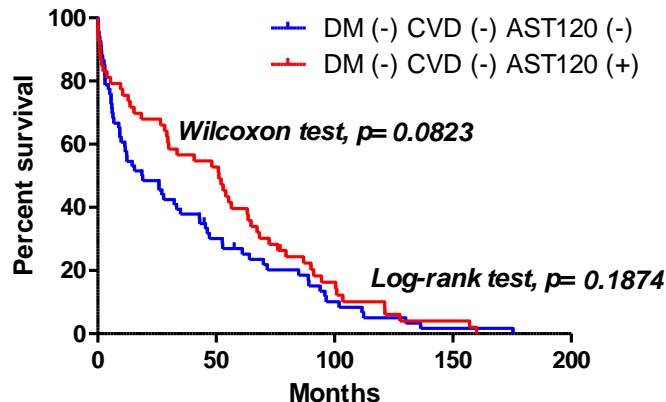
- ▶ 生存分析は、因子の有無と時間の関係を見ることができる統計法です。
- ▶ Kaplan-Meier法の生存曲線は、ある因子の有無で分けた2群において、死亡までの期間(or 観察打ち切りまでの期間)と、その状態変数(0か1のエンドポイント)を入力すれば作成できます(後述)。
- ▶ 死亡(=1)するまでの時間だけでなく、イベントが発生(=1)するまでの時間(癌再発や脳梗塞発生など)にも応用できます。また、打ち切りが扱えるのが生存分析の利点です。
- ▶ 打ち切り例とはエンドポイントに至っていない追跡症例のことで、たとえば
 - ▶ 観察期間を終わった時点で生存している症例
 - ▶ 他の原因で死亡した症例
 - ▶ 消息不明例、など。
- ▶ 打ち切りが多いと問題があり、観察期間が短い例や、他の原因で死亡した症例の場合には問題ないのですが、消息不明例の場合には死亡の可能性も含み、データの信頼性が低くなることがあります。

Log-rank testと一般化Wilcoxon検定

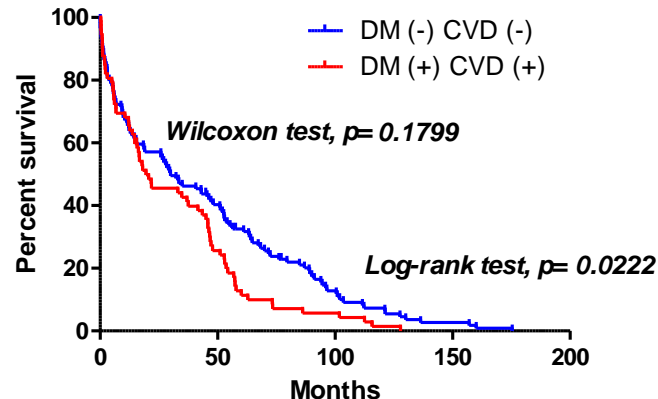
- ▶ Kaplan-Meier法において、2群間の差はLog-rank testか一般化Wilcoxon検定で行われます。
- ▶ Log-rank testは後期の死亡に重みを置き、一般化Wilcoxon検定は早期の死亡に重みを置いて解析しているようですので、目的に合った解析法を選択します。
- ▶ また、比例ハザード性が成立する場合に、つまり、比較する2群のハザード比がどの時間でも等しいとき、最も検出率が高くなるようです。
- ▶ 難しく言うとグループ間の生存曲線が一定の比率で変化している＝簡単に言うと、Kaplan-Meier法でカーブがクロスしていない、ことが必要です。クロスしている場合にはその因子は有意にならないので、クロスしない工夫が必要です。(例:クロスしない時点から解析する、など)

Int J Nephrol. 2012;2012:376128
<http://www.hindawi.com/journals/ijn/2012/376128>

Survival of HD patients (DM-, CVD-, AST120 -/+)



Survival of HD patients (DM +/-, CVD +/-)

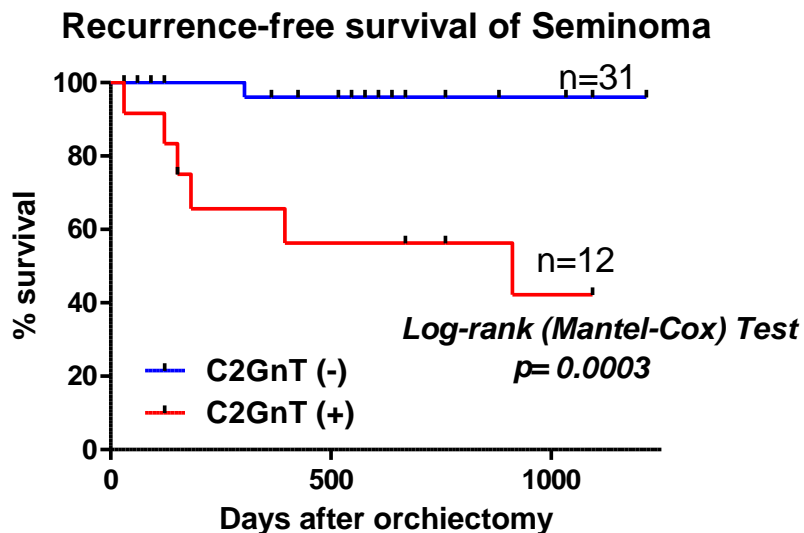


データの入力の仕方

- ▶ GraphPad Prism®における入力法を示します。
- ▶ この場合、精巣腫瘍Stage I 患者の術後再発をイベント発生=1とし、C2GnT1免疫染色の(+, -)で群分けしています。
- ▶ 明らかにC2GnT1陽性例で再発が多い、といえます。

[Int J Cancer. 2010 Sep 1;127\(5\):1052-9.](http://onlinelibrary.wiley.com/doi/10.1002/ijc.25117/abstract)

<http://onlinelibrary.wiley.com/doi/10.1002/ijc.25117/abstract>



X	A	B
Days after orchiectomy	C2GnT (-)	C2GnT (+)
X	Y	Y
121.6	0	
1094.4	0	
1094.4	0	
1094.4	0	
1094.4	0	
30.4	0	
760.0	0	
1094.4	0	
1094.4	0	
1094.4	0	
1094.4	0	
1094.4	0	
304.0	1	
30.4	0	
60.8	0	
608.0	0	
91.2	0	
1033.6	0	
1216.0	0	
1094.4	0	
881.6	0	
668.8	0	
30.4	0	
638.4	0	
608.0	0	
577.6	0	
547.2	0	
516.8	0	
425.6	0	
425.6	0	
364.8	0	
1094.4		0
1094.4		0
1094.4		0
395.2		1
121.6		1
182.4		1
152.0		0
152.0		1
912.0		1
668.8		0
760.0		0
30.4		1

多変量解析 (Multivariate analysis)

- ▶ 「多くの個体について、2つ以上の測定値(身長や体重、年齢、病期、採血値など)がある場合、これらの変数の相互関連を分析する方法の総称」です。
- ▶ 10種類以上の方法があり、データの様式により使い分ける必要があります。
- ▶ 従属変数:yとは結果の値です。例:点数、生死の有無、転移の有無など
- ▶ 独立変数:xとは結果:yに影響を及ぼすと考えられる因子です。

手法 (一部抜粋)	独立変数(x)		従属変数(y)	
	質的データ	量的データ	質的データ	量的データ
重回帰分析	(一部可能)	複数		単数
判別分析		複数	(0-1)の2値型	
ロジスティック回帰分析	複数	複数	(0-1)の2値型	
比例ハザード分析	複数	複数	(0-1)の2値型	
分散分析	複数			単数
主成分分析		複数		
因子分析		(複数)		複数
クラスター分析		複数		

多変量解析の用語

- ▶ 独立変数: x 、従属変数: y という言葉がでてきます。
 - ▶ 独立変数: x とは、学歴、TNM分類、ステージなど結果: y に影響を与える因子のことをさします。
 - ▶ 従属変数: y とは、合計点数、生存の有無、転移の有無など、 x の影響による結果の値、結果の状態をさします。
 - ▶ これらの x と y は知りたい関心に応じて解析者が選ぶものです。
- ▶ 結果の値(従属変数: y)に対して複数の因子(独立変数: x)の影響を知りたい場合に多変量解析を使います。
- ▶ 解析法はデータの様式により使い分けます。
- ▶ 主に使うのは(Cox回帰)比例ハザード分析、ロジスティック回帰分析、重回帰分析などです。

多変量解析の使い分け

- ▶ 時間的要素を考慮しなければならず、従属変数が0-1の2値型の場合は(Cox回帰)比例ハザード分析です。
- ▶ 時間的要素がなく、従属変数が0-1の2値型の場合はロジスティック回帰分析です。
- ▶ 時間的要素がなく、従属変数が点数、身長、採血値などの量的データ、独立変数も量的データの場合は重回帰分析です。
- ▶ という具合に、データの様式により使い分けます。これ以上の説明は成書を参照してください。私も説明不可…。

データ尺度の扱い方：質的と量的データ

- ▶ **名義尺度(質的データ=カテゴリーデータ)**
 - ▶ 質的データとは男=1、女=0や生存=0、死亡=1などダミー変数へ変換したデータをさす。カテゴリーデータとも言う。数値の計算は意味を持たない。
- ▶ **順序尺度(質的データ=カテゴリーデータ)**
 - ▶ 数値が大小関係のみを表す。T分類でT1~4の大小関係が $1 < 2 < 3 < 4$ と保障されている時、T1=1、T2=2、T3=3、T4=4と割り当てれる。数値の計算には意味がなく、順序にのみ意味がある。
- ▶ **間隔尺度(量的データ)**
 - ▶ 測定対象における量の差を表す尺度。例として、年齢、温度など。
- ▶ **比率尺度(量的データ)**
 - ▶ 間隔尺度に似ているが、原点(0値)が定まっているものをさす。長さcm、重さkg、時間minなどである。
- ▶ **尺度の扱い方で意味が変わる(測定者次第です)**
 - ▶ A:鉛筆、B:筆、C:万年筆としたとき、長さをA=16cm、B=15cm、C=14cmとした時は比率尺度、長い順にA=1、B=2、C=3、としたら順序尺度、名前でA=1(鉛筆)、B=2(筆)、C=3(万年筆)としたら名義尺度である。年齢も年代(10代、20代...)とするとカテゴリーとなり質的データとなる。
- ▶ 「名義と順序」尺度を質的データ、「間隔と比率」尺度を量的データとして扱う

解析法の選択法：Cascade Figure

従属変数:y に対する複数の因子の影響を見たい
従属変数:y の数は1つ?それ以上?

1つ

2つ以上

従属変数:yは、
量的データか?
2値型のダミー変数か?

正準相関分析

量的データ
(0-1以外)

重回帰分析

質的データ
(0-1の2値型)

従属変数:yは、
時間要素を含むデータか?

時間依存性なし
(横断データ)

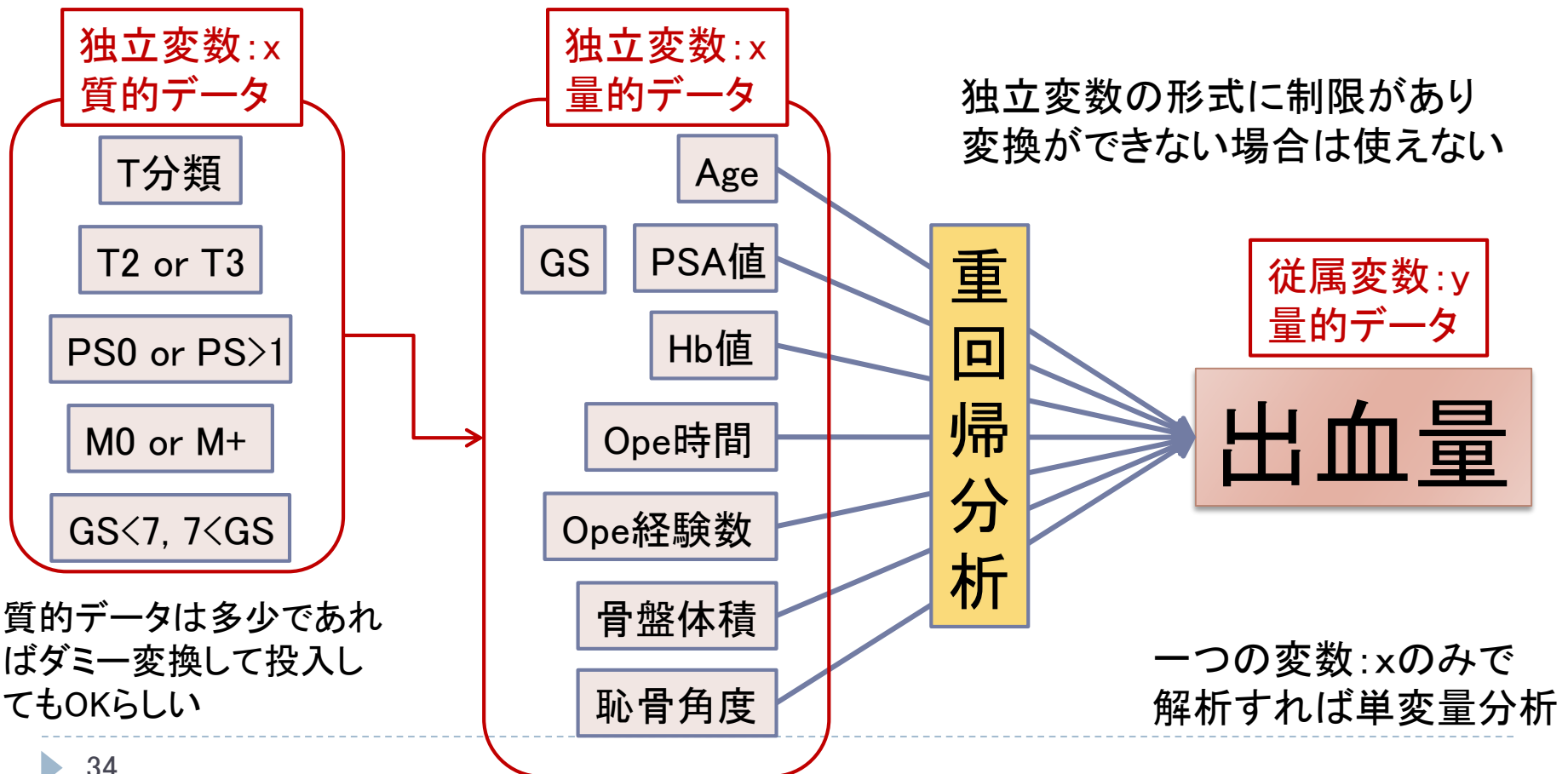
ロジスティック回帰分析

時間依存性あり
(縦断データ)

(Cox回帰) 比例ハザード分析

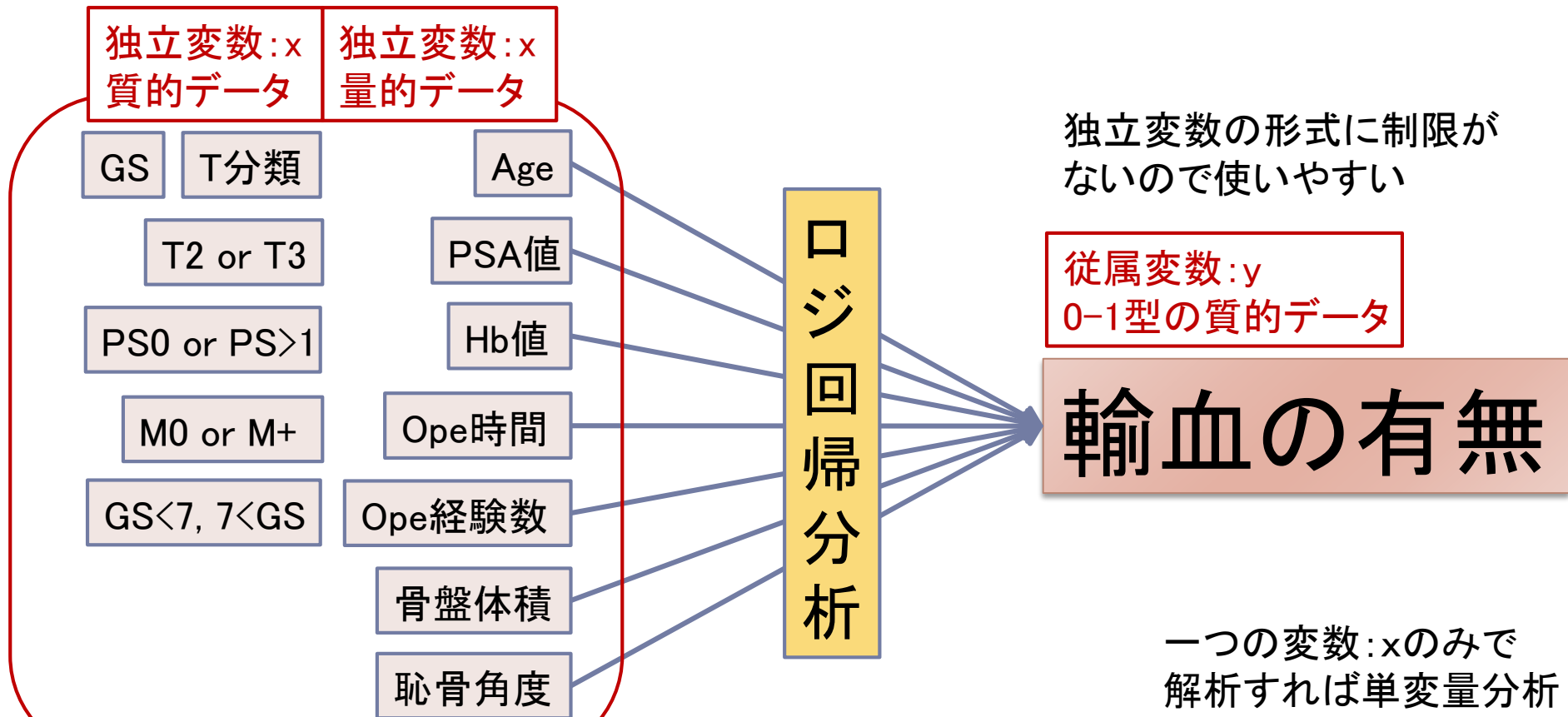
重回帰分析:前立腺癌編

- ▶ 1つの従属変数: y (量的データ)に対して複数の独立変数: x (量的データ)の影響度合いを解析する方法



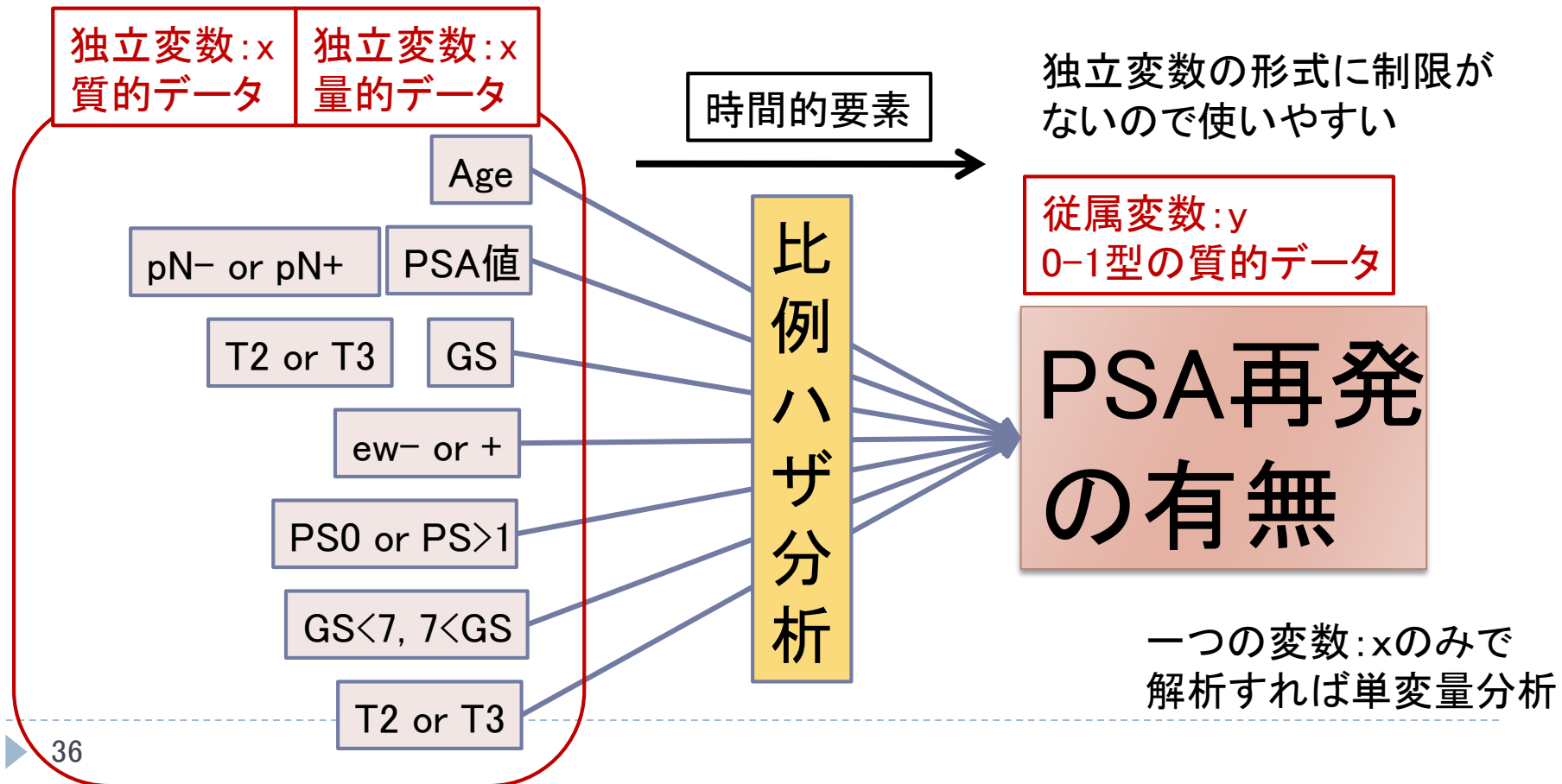
ロジスティック回帰分析:前立腺癌編

- ▶ 1つの従属変数: y (0-1型データ)に対して複数の独立変数: x (質・量的データ)の影響度合いを解析する方法



Cox回帰比例ハザード分析:前立腺癌編

- ▶ 従属変数: y (0-1型データ): イベントが起こった群(1)と起こらない群(0)の2群: に対して、**時間的要素も考慮**して複数の独立変数: x (質・量的データ) の影響度合いを解析する方法



95% CIの意味（オッズ比、ハザード比）

- ▶ 95%の確率で母集団の平均値が含まれているような範囲を95%信頼区間(95% CI)という
- ▶ ロジスティック分析ではオッズ比、比例ハザード分析ではハザード比という言葉がでてきます。
- ▶ オッズというのは、事象がどのくらい確実に起こるかの度合いを表現する方法で(詳しくは割愛)、ある疾患などへの罹りやすさを2つの群で比べる統計学的な尺度となります。
- ▶ オッズ比やハザード比が1とは、ある疾患への罹りやすさが両群で同じということであり、1より大きいとは、疾患への罹りやすさがある群でより高いことを意味します。逆に比が1より小さいとは、ある群において疾患に罹りにくいことを意味します。
- ▶ 信頼区間に1が入るということは、その比率が1 = 同じということもありうる、という意味になるので、有意差はなくなります。

単変量と多変量の使い分け

- ▶ 多変量の独立変数: x は何でもかんでも投入すればいい訳ではありません。「なるべく少ない変数: x を投入」が原則です。
- ▶ よくある手法としては、まずは単変量解析で独立変数: x 1つ1つの有意差を検定します。
- ▶ その後、影響があると思われる独立変数: x 数個を多変量解析に投入します。
- ▶ 例: 透析導入を遅らせる因子の解析(後ろ向き観察研究)

Cox回帰比例ハザード分析

[Int J Nephrol. 2012;2012:376128](http://www.hindawi.com/journals/ijn/2012/376128)

<http://www.hindawi.com/journals/ijn/2012/376128>

/

因子: x	ハザード比	95%CI	P value
ACEI/ARB	1.275	0.957 - 1.698	0.097
AST120	1.415	1.073 - 1.867	0.014

因子: x	ハザード比	95%CI	P value
Gender	1.115	0.843 - 1.474	0.447
Age	0.990	0.978 - 1.003	0.128
DM	0.831	0.634 - 1.089	0.180
CVD	1.179	0.902 - 1.541	0.277
ACEI/ARB	1.343	1.012 - 1.783	0.041
AST120	1.467	1.116 - 1.93	0.006

Winner!

独立変数の選択：症例数の問題

- ▶ 影響ある独立変数の選択には注意が必要です。多変量解析は、イベント発生症例数に応じて、投入因子数が決まります。
- ▶ 一般に2群比較で、1群当たり「因子数 × 10～15例」が、イベント発生の少ない群に必要な数とされます。
- ▶ **ロジステック回帰分析：因子数 × 10例**
- ▶ **コックス比例ハザードモデル：因子数 × 10例**
- ▶ **線形回帰モデル：因子数 × 15例**
- ▶ 年齢、性別などの因子は、結果に大きな影響力がありますので外せません。この2因子のみでも、(1群あたり10例 × 2=20例) × 2群=40例が必須になります。これ以下ですと、統計量不足で測定できないことになります。

必要症例数：発生率からの検討

- ▶ 前立腺癌術後にPSA再発しやすいリスク因子の検討
- ▶ 前立腺癌の術後癌再発率20%
- ▶ 独立因子：
年齢、性別、PSA、T分類、グリソンスコアの5因子
- ▶ 再発群に5因子 × 10例 = 50例必要
- ▶ 全体では 20% : 50例 × 5 = 250例が必要 (つまり非再発群
200例 : 80%)

必要症例数：イベント発生数からの検討

- ▶ ハイリスク前立腺癌に対する術前療法が術後PSA再発予防効果があるか検討したい場合
- ▶ 前立腺癌術前の治療介入群：500例、非介入群：500例
- ▶ 術後PSA再発の発生数：
治療介入群：50例(10%)、非介入群：100例(20%)
- ▶ 目的とするアウトカムの少ない方の群のイベント数=50なので、投入できる因子数は5個です。
- ▶ もし10因子で検討したければ、10因子×10例=100例、発生率10%より1000例の治療介入群が必要です。

では、どうやって独立変数を選ぶのか？

- ▶ 現実には検証可能な症例数は限られているので、多数因子の解析は不可能です。
- ▶ 因子を増やすと、統計モデルが不安定となり、統計量不足で測定できないこととなります。
- ▶ そこで、現状では、先行論文や臨床経験等で重要な因子を選択し、最適なモデルを作ることが重要です。
- ▶ 「検証的研究」の場合は因子が既知であるため、必要症例数には十分な配慮が必要です。
- ▶ ただ「探索的研究」の場合は、必ずしも必要症例数を満たす必要はありません。その点を上手に記載することが重要になります。

サンプル・サイズ

- ▶ 臨床研究の質を決める重要な要素です
- ▶ サンプル数が多ければ多いほど解析の精度が上がる、それに伴いP値は小さくなります。
- ▶ 臨床的に意味のないどんなに小さな差でも、サンプル数を増やせばいつかは統計的には有意となります。
- ▶ しかし、安全性とコストの点から“必要最低限”のサンプル数を見積もることが必要です。

サンプルサイズを考えるツール

▶ アルファ・エラー (α)

- ◆ 一般に「1型エラー」、「有意水準」とも呼ばれ、通常5%を使用 ($\alpha = 0.05$)

▶ 検出力 (Power、 $1 - \beta$) 本当に差があるときに正しく“差がある”と判断する確率

- ◆ パワーは大きいほどよく、慣習的に0.8-0.9とする

ポイント;

$\alpha = 0.05$, $\beta = 0.20$, 検出力=0.80の設定が一般的

α エラーについて

- ▶ 臨床試験：新薬vsプラセボ
- ▶ 新薬に効果が「ない」のに誤って「ある」結果になってしまうエラーの確率のことを「 α エラー」
- ▶ そんな確率は低い方がいい。
- ▶ 慣習的なルールで、このエラーが5%未満程度であれば許容する
- ▶ =間違いを5%まで許した設定

βエラー

- ▶ 新薬に効果が「ある」のに誤って「ない」結果になるエラーは「βエラー」。
- ▶ これも低い方がいい。
- ▶ $\beta = 0.20$ としたとしたら、新薬に効果が「ある」のに、「ない」といってしまう確率が20%
- ▶ 20%の確率で見逃すのを許容した設定

- ▶ $1 - \beta$ を統計学的パワー（検出力）という。 $\beta = 0.20$ に設定すると、80%の確率で、新薬の差を検出ことができる

サンプルサイズの計算

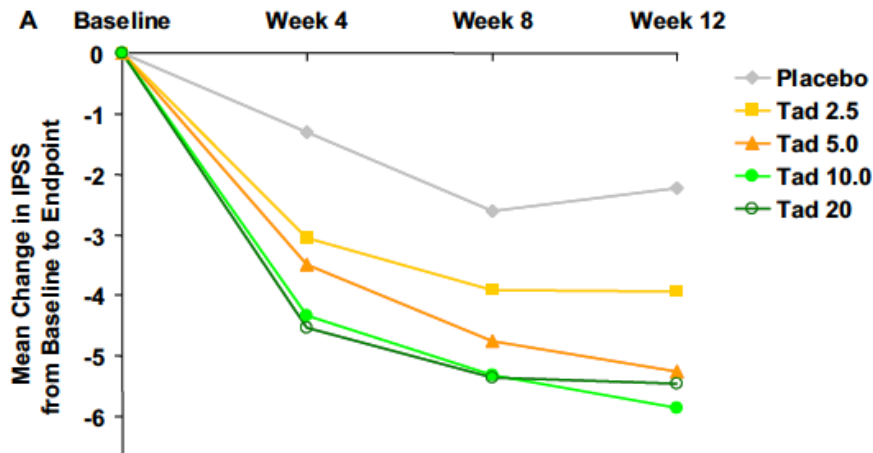
- ▶ 既存の試験から、どのくらいの効果が期待できるか?を推定

Tadalafil Administered Once Daily for Lower Urinary Tract Symptoms Secondary to Benign Prostatic Hyperplasia: A Dose Finding Study

J Urol 2008

Claus G. Roehrborn,* Kevin T. McVary, Albert Elion-Mboussa and Lars Viktrup

From the Department of Urology, University of Texas Southwestern Medical Center at Dallas (CGR), Dallas, Texas, Department of Urology, Northwestern University (KTM), Chicago, Illinois, and Lilly Research Laboratories, Eli Lilly and Co. (AEM, LV), Indianapolis, Indiana



TadalafilのIPSS変化

- Tadalafil 5.0mg: 17.3から12.1点 (30.0%)
- プラセボ: 17.1から14.7点 (13.9%)

プラセボでも **13.9%**の排尿改善が得られる

計算

	A	B	C	D	E
1	Sample Size: Comparing Two Proportion				
2	*Chi-square testまたは Fisher's exact probability testを用いる場合				
3				プラセボ 13.9%	
4	対照群有効率(π_1)	0.139			
5	治療群有効率(π_2)	0.3		Tadalafil 30.0%	
6	症例数の比($n/m = \psi$)	1			
7	α エラー*	0.05		$\alpha : 0.05$	
8	Power	0.8		パワー: 0.80	
9	対照群症例数(m)**	103	カイ2		
10	治療群症例数(n)	103			
11	総症例数(N)	206			
12					

- ▶ 同じ様な研究をするなら、必要症例数は206例となる

Webでも計算可能

サンプルサイズ設計

二群比較

- MMRMに対するサンプルサイズ設計
- 二群の生存関数の検定のサンプルサイズ設計 (年次生存割合から計算)
- 二群の生存関数の検定のサンプルサイズ設計 (MST から計算)
- 二群の生存関数の検定のサンプルサイズ設計 (MST+HR から計算)
- 二群の母平均の差の仮説検定に対するサンプルサイズ設計 (群間差から計算)
- 二群の母平均の差の仮説検定に対するサンプルサイズ設計 (二群の平均から計算)
- 2×2 分割表のカイ二乗検定に対するサンプルサイズ設計

二群比較 (非劣性)

- 非劣性仮説のlog-rank検定に対するサンプルサイズ設計 (年次生存割合から計算)
- 非劣性仮説のlog-rank検定に対するサンプルサイズ設計 (MST から計算)
- 非劣性仮説のlog-rank検定に対するサンプルサイズ設計 (MST+HR から計算)
- 2×2 分割表のカイ二乗検定に対するサンプルサイズ設計 (非劣性)

▶ <http://www.nshi.jp/contents/>

▶ 非劣性試験は、プラセボデータから非劣性マージンの設定が必要です(1.2~1.25の範囲が一般的の様です)

アプリ

入力

標準治療群の有効確率 (p_1)

試験治療群の有効確率 (p_2)

標準治療群に対する試験治療群の割付比

α

$1 - \beta$

仮説
下側 (H1: p2-p1<0)
上側 (H1: p2-p1>0)

計算結果

合計サンプルサイズ	206
標準治療群のサンプルサイズ	103
試験治療群のサンプルサイズ	103
実際の検出力	0.802

計算条件

標準治療群の有効確率 (p_1)	0.139
試験治療群の有効確率 (p_2)	0.3
標準治療群に対する試験治療群の割付比	1
α	0.05
$1 - \beta$	0.8
仮説	両側
ログ	2015/12/27 22:34:56

感度・特異度・ α エラー・ β エラー

N=1000	生検陽性	生検陰性	
PSA>4	真の陽性100	偽陽性300 α エラー = 300/890 (0.34)	陽性適中率 = 25% $\frac{100}{100+300}$
PSA<4	偽陰性10 β エラー = 10/110 (0.09)	真の陰性590	陰性適中率 = 98% $\frac{590}{10+590}$
	感度 = 91% $\frac{100}{10+100}$	特異度 = 66% $\frac{590}{300+590}$	

感度とは、ある検査について「陽性と判定されるべきものを正しく陽性と判定する確率」
 特異度とは、ある検査について「陰性と判定されるべきものを正しく陰性と判定する確率」
 前立腺癌がないのに、癌あり、とまちがう確率 = α エラー
 ▶ 前立腺癌があるのに、癌なし、と間違える確率 = β エラー



この資料はこんな本を参考に作成しました



臨床研究初心者のためのやっぱり わかりにくい臨床研究デザイン

その簡単な理解のための要点集

臨床研究はデザインですべてが決まる

- ▶ 臨床研究デザインの型は偉い先人のおかげですすでに確立しています。我々はそれを選ぶだけです。たとえば、
 - ▶ 観察するのか、介入するのか
 - ▶ 過去にさかのぼって調べるのか、これから調べだすのかなどなどさまざまあります。
- ▶ また、自分が組む組まないにかかわらず、臨床研究デザインを知ることは論文を読むときに深い理解ができるようになります。
- ▶ 他人の仕事がいい仕事なのか解るためにも、基本的なことだけでも理解しましょう。統計よりは解りやすいです。

観察研究と介入研究

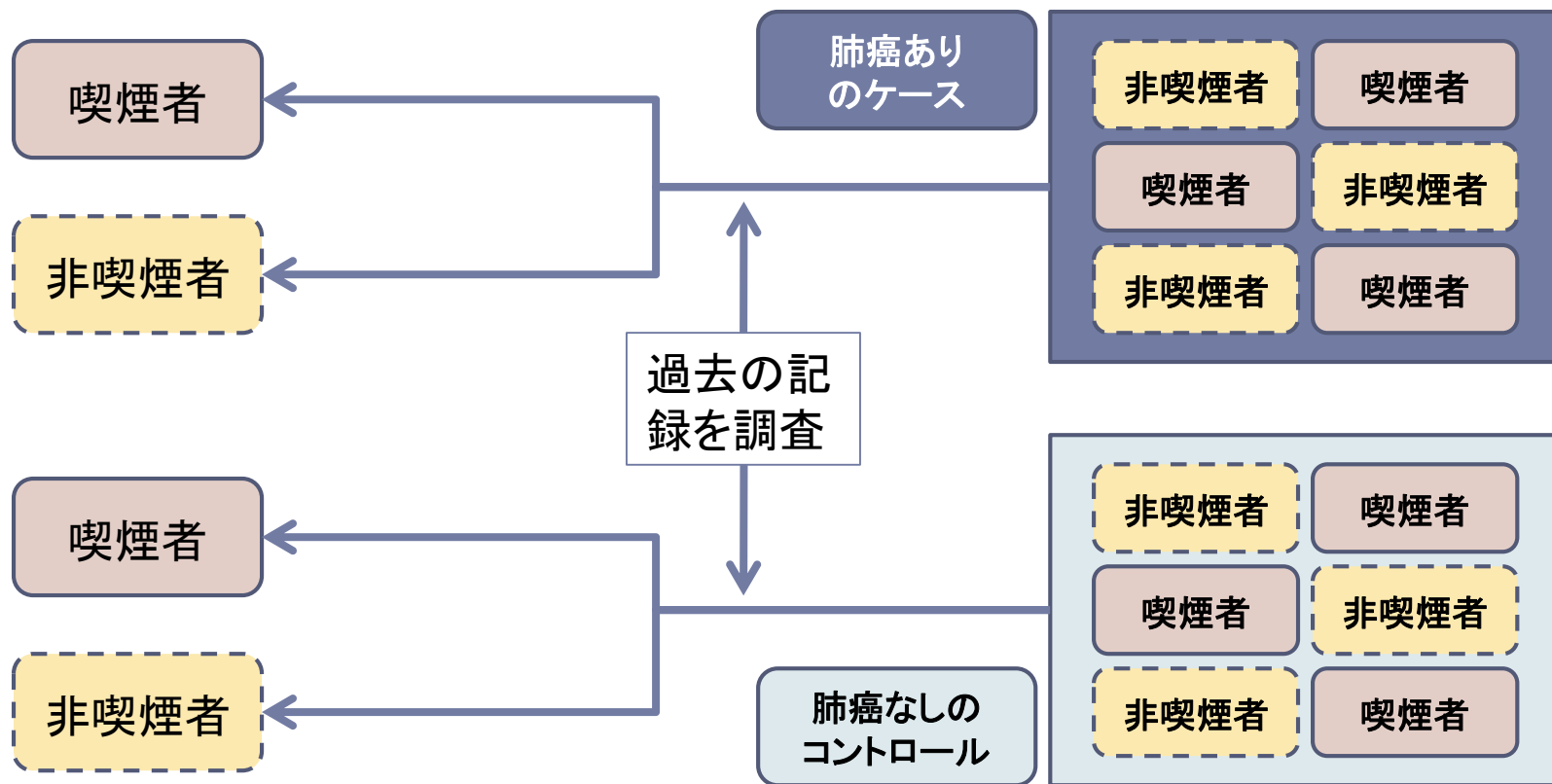
- ▶ 大きな分類として観察か、介入か、に分けられます。じっと見つめて観察するだけか、何か薬を飲ませて介入するかの違いです。
- ▶ 観察研究は仮説を形成するのに向いている、介入研究は仮説を検証するのに向いています。
- ▶ 観察研究はやりやすい利点がありますが、こじつけが可能な点からEvidence Levelは低くなります。
- ▶ 介入研究は比較試験です。最強なのはランダム化比較試験(RCT)ですが、そう易々とできるものではありません。NEJMなどでは1000人規模でのRCTの結果が華々しく一世を風靡しています
 - ▶ 今はこれをやらないと効果を語れない時代になっています。

観察研究	介入研究
横断研究(時間経過なし)	クロスオーバー研究(前向き)
症例対照研究(後ろ向き) (ケースコントロール研究)	ランダム化比較試験(前向き)
コホート研究(前向き)	

横断研究のエッセンス

観察研究	Evidence Level: 記載なし
特徴	現時点でのデータを集めるタイプ 時間経過を伴わない
目的	現状把握ができる 何らかの因果関係が見いだせる
利点	長期の追跡がいらないので簡単、 気軽にできる、お金がかからない
難点	因果関係の検証はできない 思いこみがバイアスになる可能性あり 医学研究には不向き
例	内閣支持率、国勢調査、 インフルエンザの感染率、日本人の平均寿命、 etc...

ケースコントロール研究（後ろ向き・観察研究）



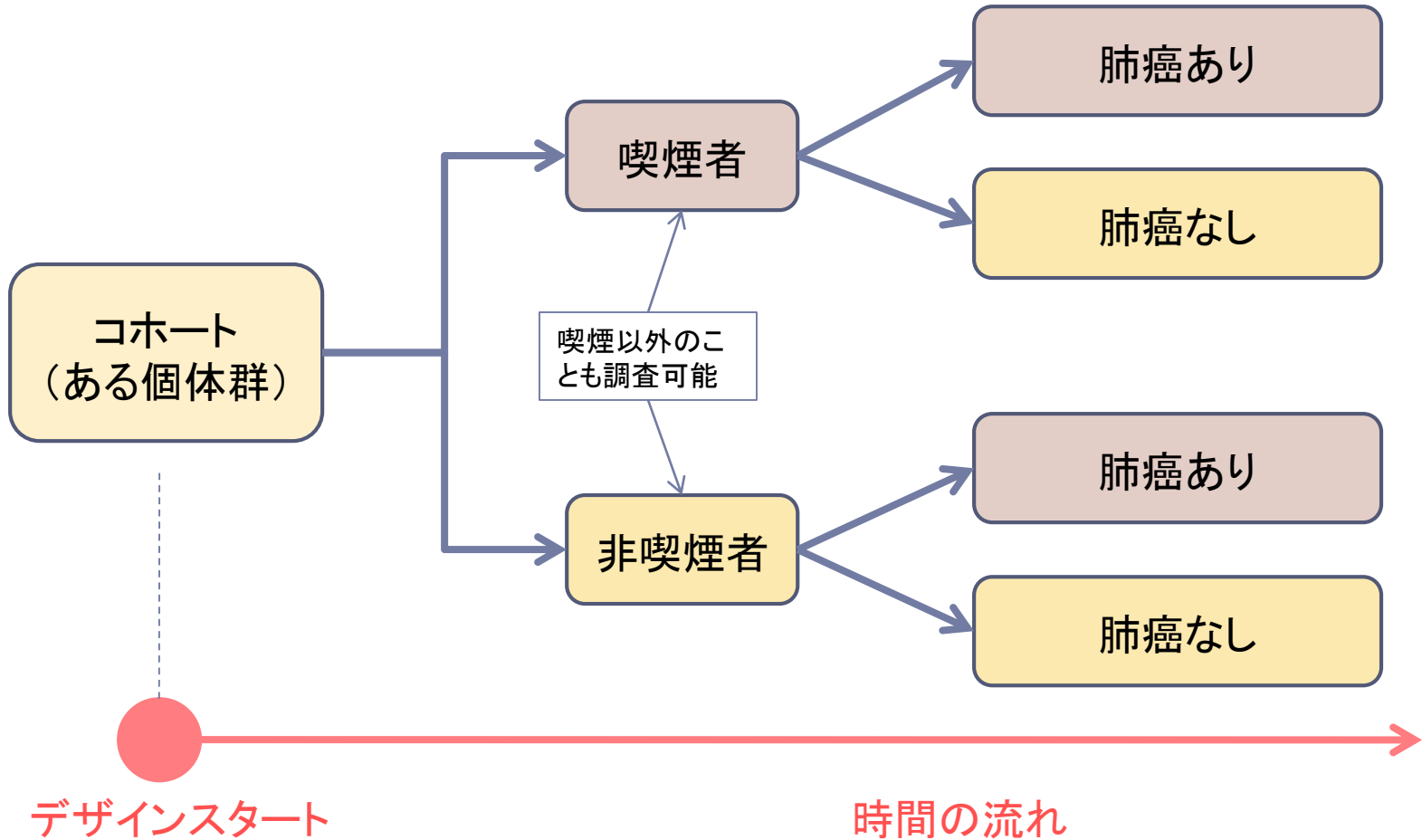
時間の流れ

デザインスタート

ケースコントロール研究のエッセンス

観察研究	Evidence Level: III~IV
特徴	現時点の患者に対し、その原因を過去にさかのぼって調査する(後ろ向き)
目的	原因不明な因果関係を見いだす
利点	カルテを見返すだけなので簡単、気軽にできる、お金がかからない
難点	過去の記録に頼るしかなく、過去のカルテ記載にバラツキがあるとアウト コントロールの選択にバイアスがかかる可能性あり、結果をこじつけることができる
例	癌の原因、まれな疾患の原因究明、コホートやRCTを組むための動機付け

コホート研究（前向き・観察研究）

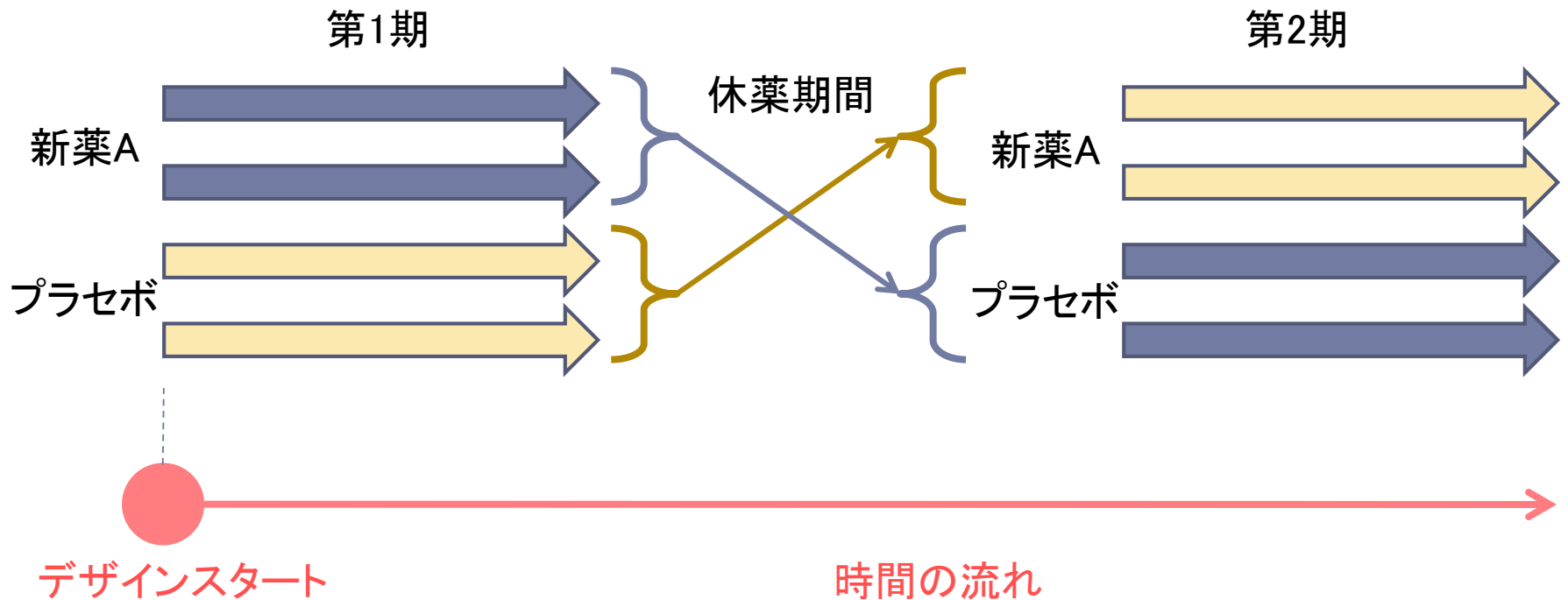


コホート研究のエッセンス

観察研究	Evidence Level: III~IV
特徴	ある個体群を対象に、時間の流れに従って追跡調査をしていく観察研究(前向き)
目的	特定の因子がある病気のRisk Factorかどうかを見いだす。
利点	広く情報を集めることができる。倫理的に安全である。ケースコントロールと比してバイアスが少ない。
難点	時間もかかるしお金がかかる。患者の脱落がおこる。病気になったかどうかわからないことがある。長いので調査の質を保つのが難しい。結果のこじつけが可能。
例	癌などの疾患の原因究明、RCTを組むための動機付け

クロスオーバー研究（前向き・介入研究）

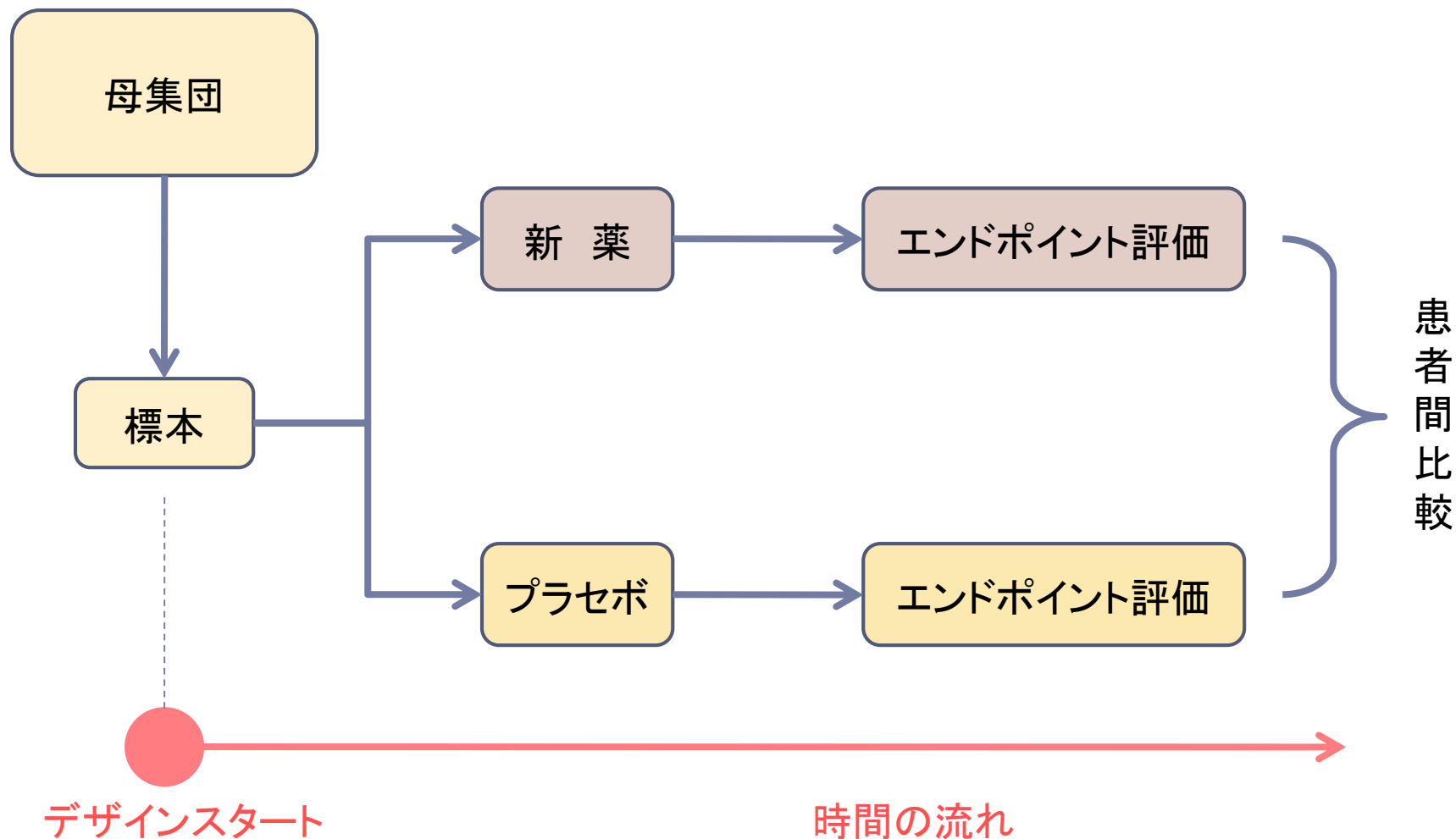
Good; サンプル数が少なくても数が稼げる
Bad; 治る病気には使えない



クロスオーバー研究のエッセンス

介入研究	Evidence Level: II~III
特徴	比較したい介入を期間を入れ替えて調査する介入研究(前向き)
目的	個人差の大きい因子の調査に効果的。新薬の第1相試験でよく使う。
利点	標本数が少なくて済む。患者内比較なので誤差が少ない。説得力がある。
難点	治る病気には使えない。死亡の調査には使えない。Washoutの時間が必要。持ち越し効果があるとバイアスになる。
例	新薬開発の第1相試験(副作用のチェック)

ランダム化比較試験（前向き・介入研究）



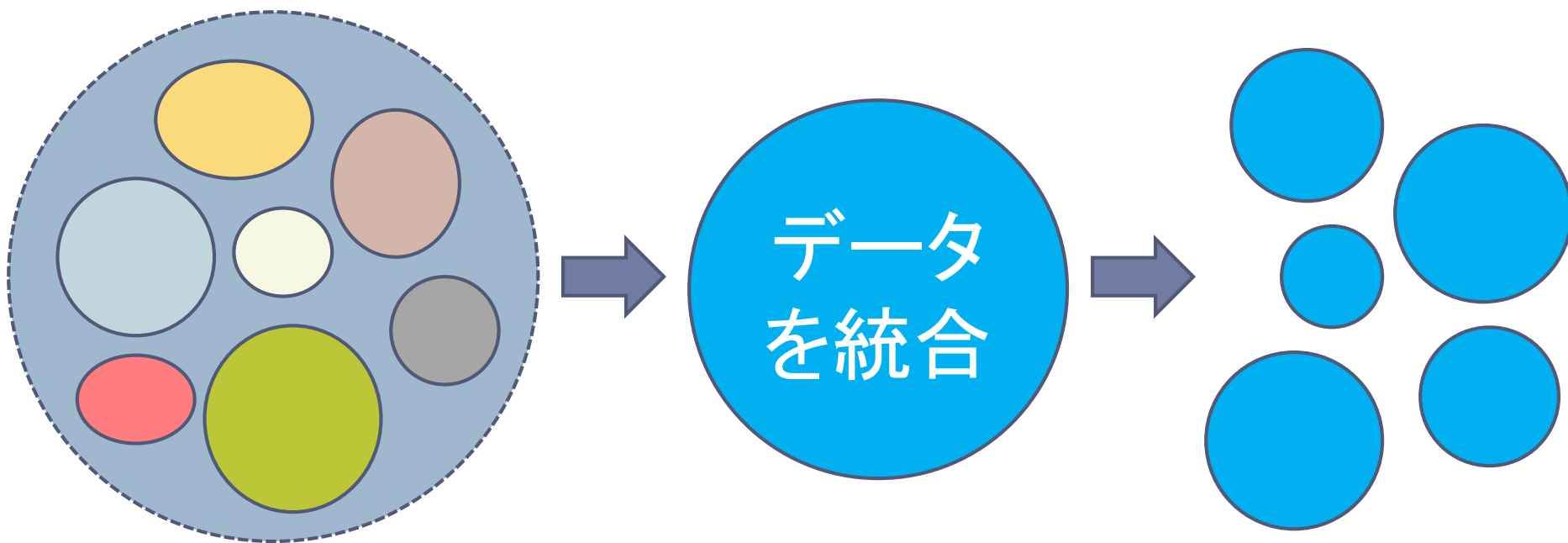
ランダム化比較試験のエッセンス

介入研究	Evidence Level: I~II
特徴	比較したい介入を2つのグループにランダムに分けて調査する介入研究(前向き)
目的	治療の効果を検証するのに最適である
利点	統計分析に非常に強い
難点	コストがかかり過ぎる。ランダム化やマスク化に手間がかかり過ぎる。ベストとは限らない。介入研究に共通する倫理的問題が大きい。
例	新薬開発の第3相試験(効果のチェック)など

メタアナリシス

- ▶ 過去に独立して行われた臨床試験のデータを掘りなおしてまとめて解析する方法です。生データを使ってやることもできるし、代表値(症例数、平均値、SDなど)だけでも可能です。
- ▶ データさえそろえば比較的簡単でEvidence Levelは高いのですが、限界もあります。
- ▶ データ、方法論、結果の均一性、同質性の点検が必須
- ▶ 過去のデータのまとめなので「後ろ向き」である
- ▶ 後ろ向きは、都合のいい論文を集め、後付け解析を100もやって、都合のいいデータだけを論文にできてしまう、という欠点があります。
- ▶ 弊害をなくすため「前向き」のメタアナリシスもありますが、WHOと国際高血圧学会主導というレベルでしかできないのが現状です。

メタアナリシスのイメージ



いろいろな臨床試験

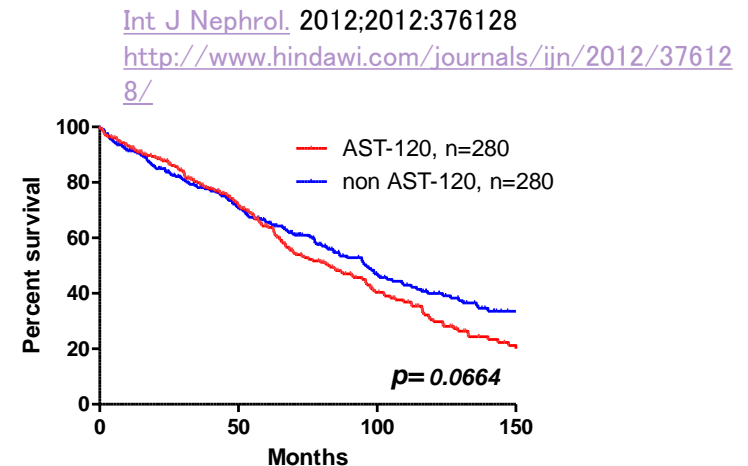
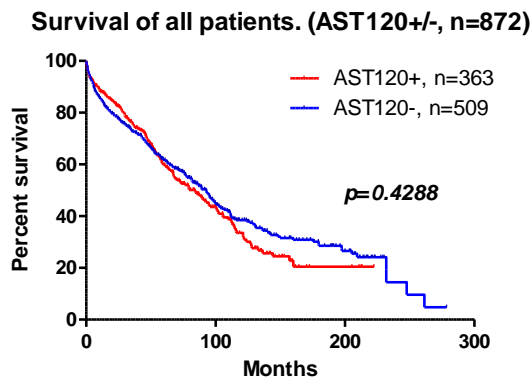
サブグループ解析

Propensity Score Matchingについて

- ▶ 患者さんを2群に分けるときの、患者選択バイアスが生じます。たとえば、若い患者が大手術を受け、高齢者は非手術療法を選択することが多くあり、この2群で治療成績を比較しようにも、年齢差があり結論は出せません。
- ▶ 実際、医療で予後因子を出すとき、「年齢」は非常に大きな因子です。
- ▶ このような背景の調整を行うため、90年代後半に医学領域におけるPropensity Scoreを用いた解析が行われるようになりました。
- ▶ 介入・暴露に対する傾向スコア(Propensity Score)を算出し、観察研究データをpseudo-randomize(偽無作為化)することで、医師によるIndication Biasの調整を行い、最終的に観察研究試験を無作為化比較試験の精度に近づける事を目指している統計解析手法です。
- ▶ 有名な研究報告に β -Blockerを術前に用いるか否かというリサーチクエスチョンを観察研究データにより証明した論文NEJM 2005 28;353(4):349-361やSwan-Ganzカテーテルを用いた重症症例に対する右心室モニタリングの生命予後に対する影響: JAMA 276(11):889-897, 1996があります。

Propensity Scoreの算出法

- ▶ 治療(介入)有り群を1、無し群を0として、Logistic 回帰分析でスコアを算出します。
- ▶ 算出されたPropensity Scoreを認容誤差範囲: 一般的には±0.03に準じて有り無し各群から1:1でピックアップしてMatchingを行います。Matchingをした因子はほとんど有意差がなくなります。1:1ですので必ず偶数になります。
- ▶ 選ばれた症例で予後等の比較を行います。



Propensity Score Matching利点と弱点

- ▶ 倫理面への配慮から、無作為化比較試験の実施が非常に難しい分野において有効である。
- ▶ 弱点は、症例数が約半分になるため、症例数が多くないと説得力がない。
- ▶ さらに、測定していない因子、マッチングできない因子については調整できない。統計的調整で完全に恣意性を排除困難。この点でも到底RCTにかなうものではない。
- ▶ 参考資料
 - ▶ 医学的研究のための多変量解析(メディカル・サイエンス・インターナショナル) p157-p161
 - ▶ SPSSで学ぶ医療系多変量データ解析(東京図書株式会社) 対馬栄輝 第6章:多重ロジスティック回帰分析の実際“予測値(Propensity)”の設定に関する記載(P126)

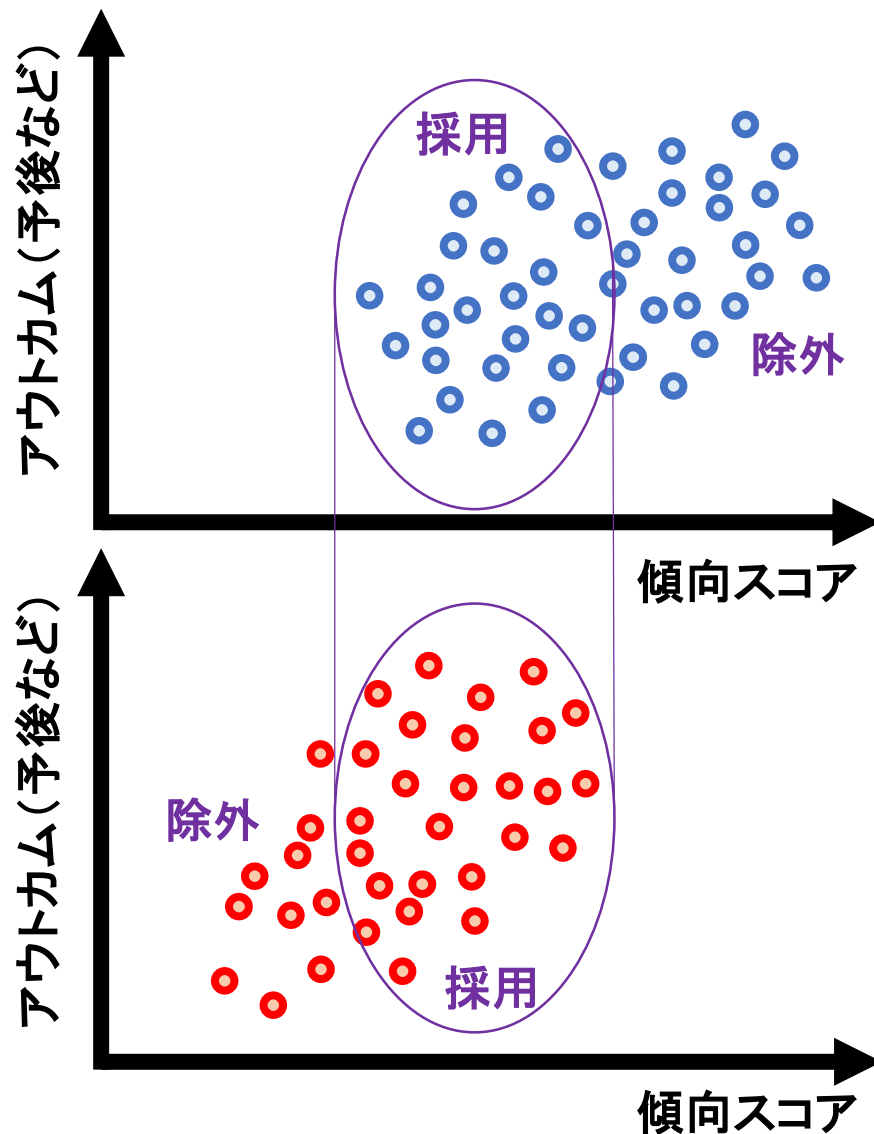
傾向スコア(Propensity score)による

- マッチング(1:1)
- 統計的補正・直接法(回帰分析)
- 統計的補正・逆数補正法(回帰分析)

傾向スコアによるマッチング

- 癌患者さん1000人、標準治療を手術と過程した場合
 - 手術を受ける患者800人
 - 術前化学療法 (Neoadjuvant: NAC) をして手術を受ける患者200人
- 当然2群間には患者背景の差があります
 - 若い、元気=NAC+手術
 - 高齢、元気ない=手術単独
- 予後も差があります
 - 若い、元気=長生き
 - 高齢、元気ない=予後短い

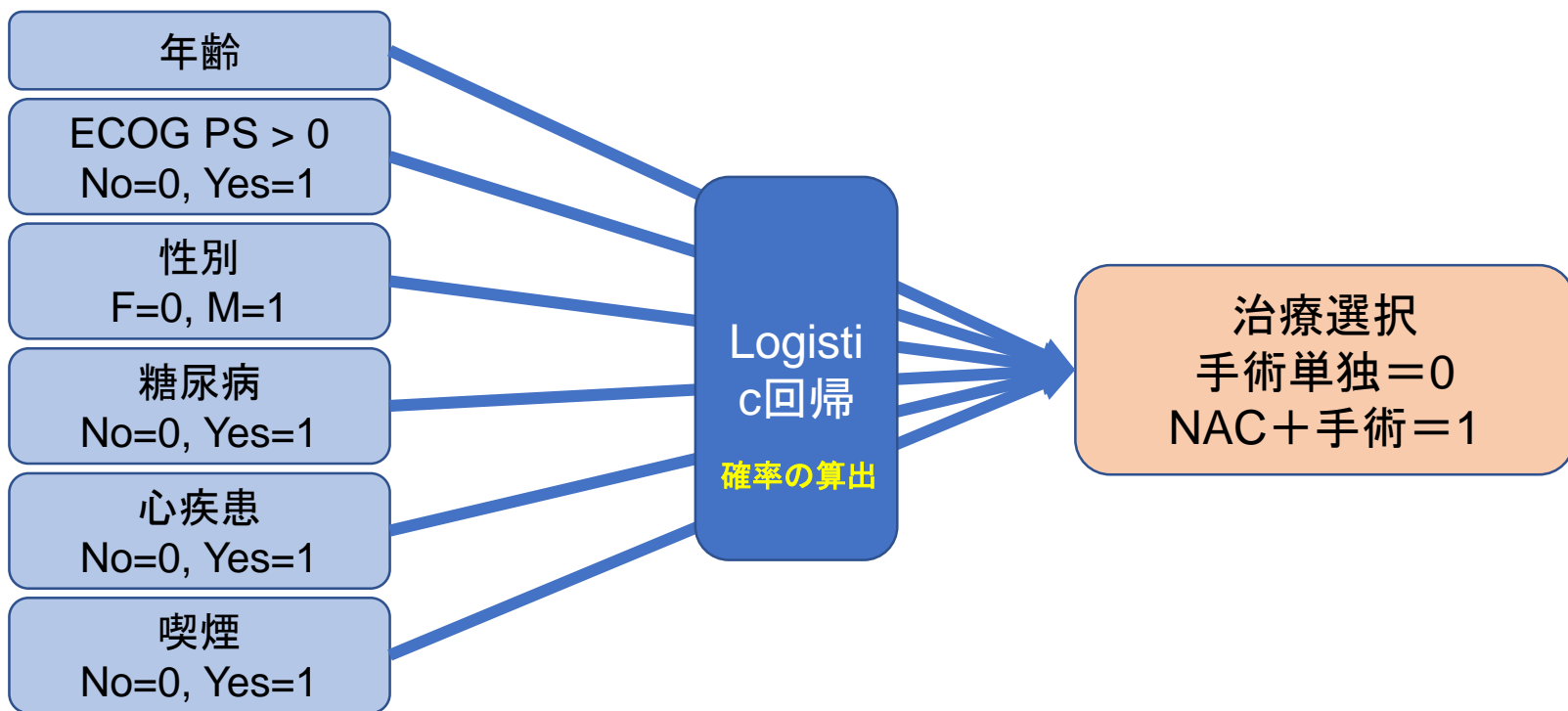
マッチングのイメージ図



- 患者背景を統一した数値に変換し、同じような背景を持った患者群を選択する
- 同じ数値の患者は同じ年齢、性別、基礎疾患をもつ
- 弱点: 患者数が減る

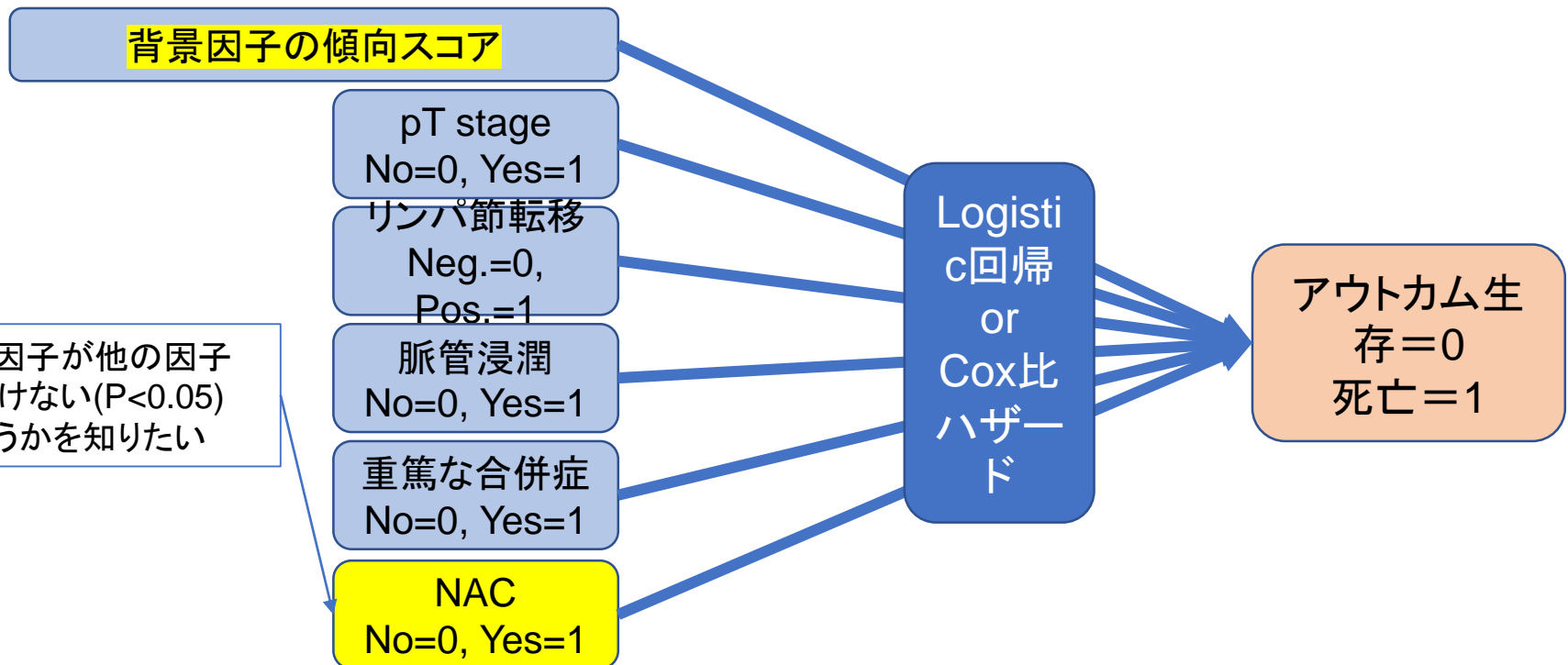
傾向スコアの作り方

- Logistic回帰分析(一般線形化モデル・二項Logisticモデル)で算出
- 交絡する可能性のある説明変数(年齢・性別・既往症など)をすべて加える
- 説明変数は入れすぎてもよい、らしい...



直接補正法

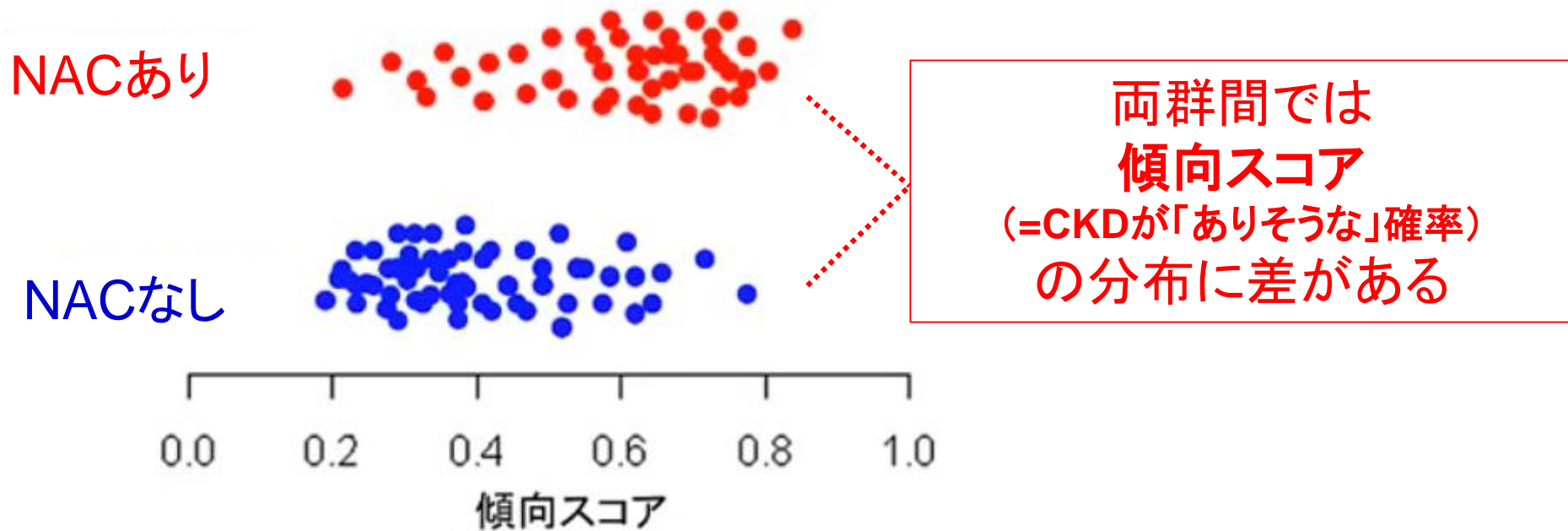
- 治療前因子を傾向スコアとして1因子にまとめる
(最終解析に投入因子数を減らす)
- Logistic回帰、Cox比例ハザードモデルに直接傾向スコアを投入する



傾向スコア逆数重み法

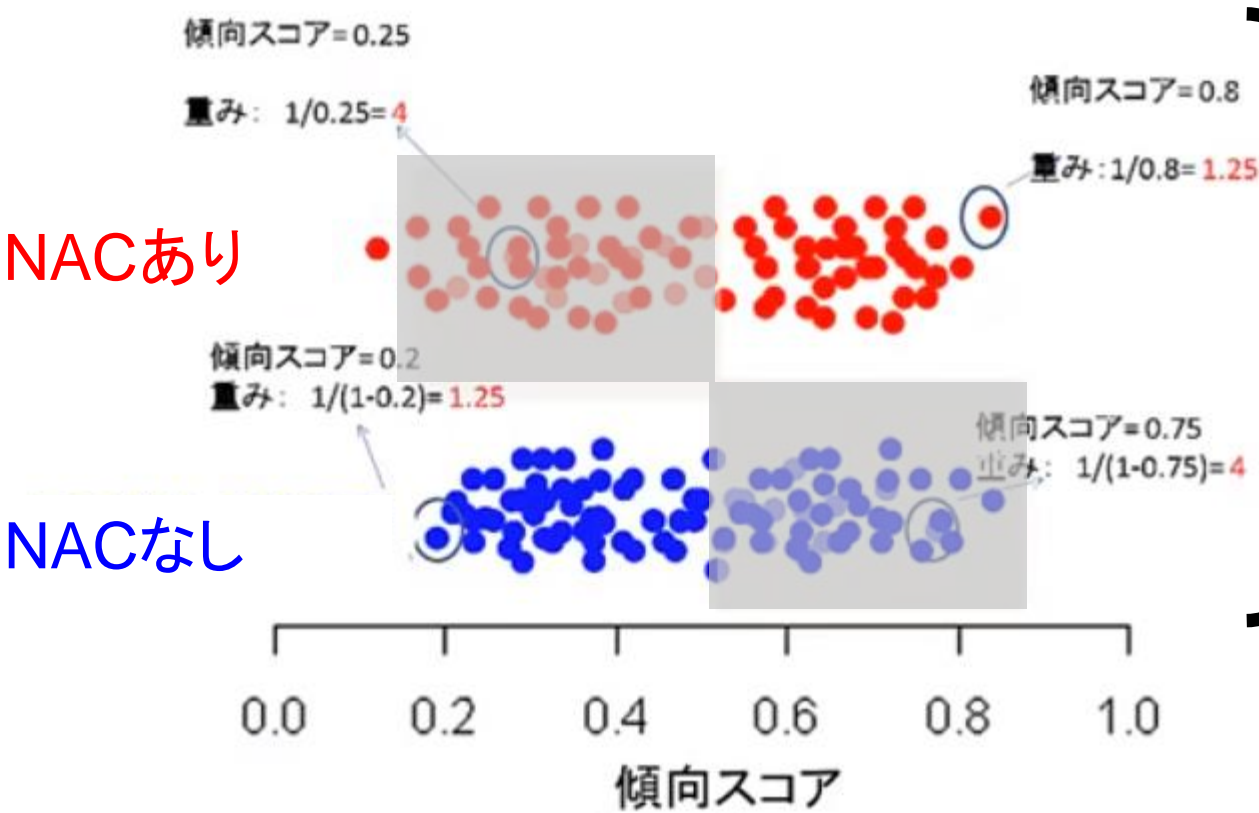
(Inverse probability of treatment weighing: IPTW法)

<模式図>



新谷歩の今日から使える医療統計学ビデオ講座：
傾向スコアの使い方とコンセプトより抜粋

傾向スコア逆数重み法



重み付けの結果
両群間の分布は
擬似的に一致



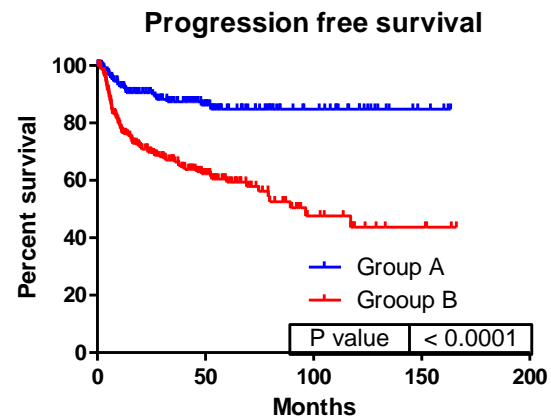
ロバスト法 + Cox回帰分析

傾向スコアの利点

治療選択
手術 = 0
放射線 = 1



アウトカム
生存 = 0
死亡 = 1



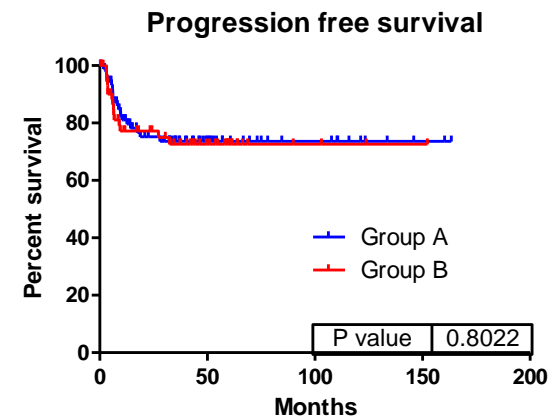
治療選択
手術 = 0
放射線 = 1



傾向スコア



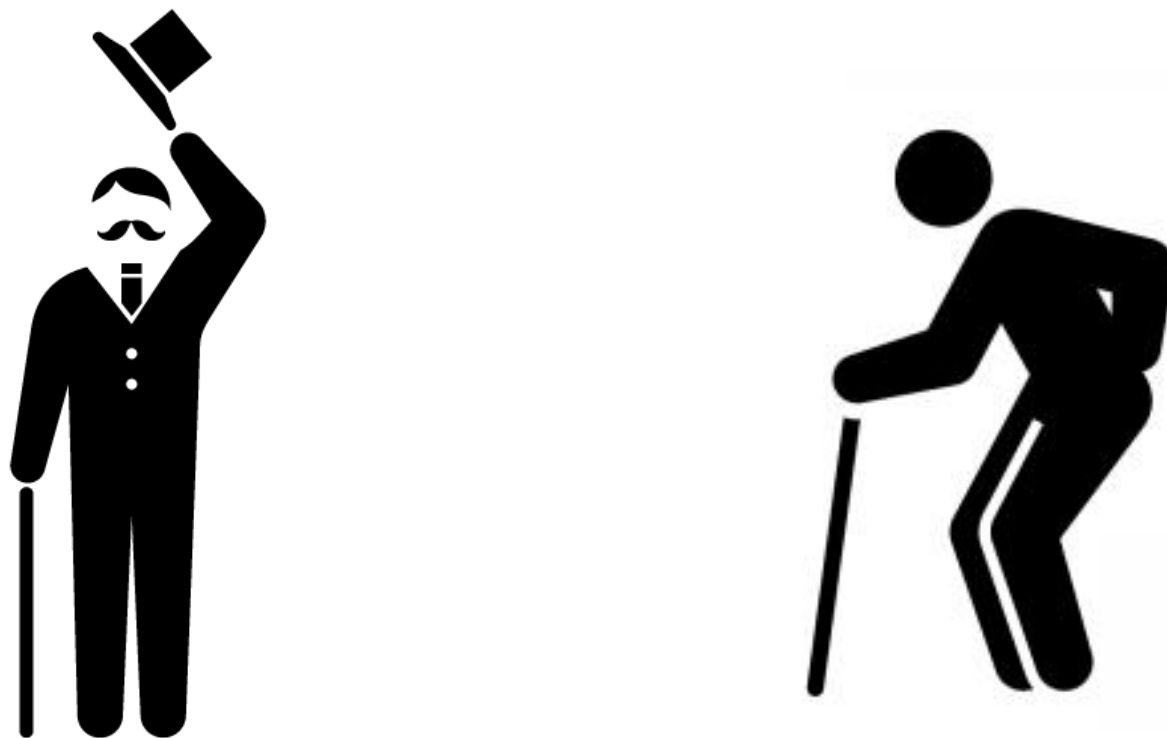
アウトカム
生存 = 0
死亡 = 1



測定できた因子は調整可能

傾向スコアの弱点

測定できない因子は調整不可



例：手術適応？放射線？

傾向スコアの使いどころ

- ランダム化可能な因子: RCTが理想的
 1. 標準治療 vs. 新規薬剤の効果
- ランダム化困難なクリニカル・クエスチョン: 傾向スコアが現実的
 1. 希少疾患の検討
 2. 標準治療がない疾患
 3. 患者選択バイアスが大きくRCT困難な治療の効果
(ex. RT vs. surgery)

Propensity Score 参考資料

- ▶ 医学的研究のための多変量解析(メディカル・サイエンス・インターナショナル) p157-p161
- ▶ SPSSで学ぶ医療系多変量データ解析(東京図書株式会社)対馬栄輝 第6章:多重ロジスティック回帰分析の実際“予測値(Propensity)”の設定に関する記載(P126)
- ▶ 医学的研究のための多変量解析(メディカル・サイエンス・インターナショナル) p157-p161
- ▶ SPSSで学ぶ医療系多変量データ解析(東京図書株式会社)対馬栄輝 第6章:多重ロジスティック回帰分析の実際“予測値(Propensity)”の設定に関する記載(P126)

終わりに お詫び

- ▶ 最後まで読んでいただきありがとうございます。
- ▶ この資料を信用し過ぎないでください。
- ▶ 統計、臨床研究について色々まとめましたが、素人の私が、自分と後輩のために作った資料ですので、正しいか？と言われると自信がありません。
- ▶ 間違いもあると思いますので、あらかじめお詫びいたします。そのときは、正しい答えを“そっと”教えてください。訂正させていただきます。
- ▶ この資料が、誰も教えてくれない「統計」という手法を受け入れる第一歩になっていただければ幸いです。
- ▶ さらなるレベルアップの時は、ご自身で成書を御購入の上学習くださるようお願いいたします。