

1981年6月5日号

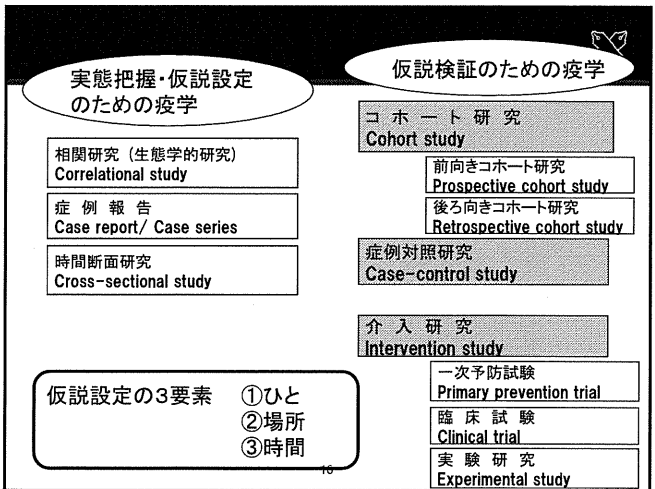
1980年10月から1981年3月の間、カリフォルニア州、ロサンゼルス市の3つの病院で5人の若い男性が *Pneumocystis carinii*肺炎の治療を受けた。

5人はすべて活動的な男性同性愛者であった。診断は生検で確認されている。5人の患者はサイトメガロウイルス感染およびカンジダの粘膜感染の既往があるかまたは現在その病気に罹っている。

5人の症例報告は、下記のとおりである、.....

臨床医学の場で出合った問題に対する疫学の原理と方法の適用 (Fletcher & Fletcher)

II
臨床研究

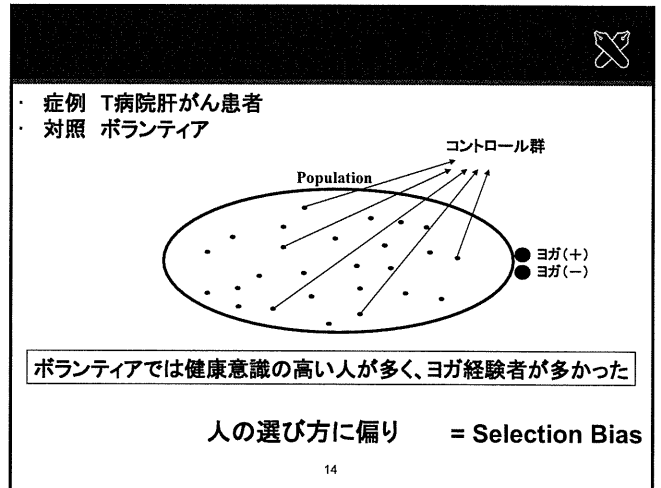
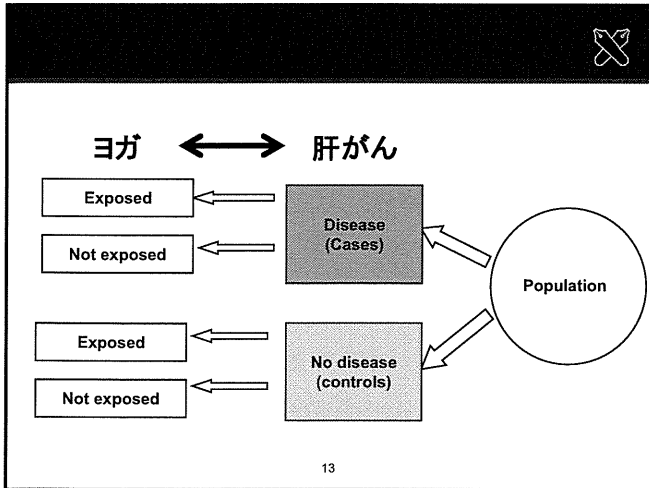


たしかにp値は大切ですが、..

その前に.....

〇〇新聞ニュース 「肝がん予防に“ヨガ”が有効」

- △△大学研究者によると
- 症例対照研究の結果、肝がん患者20名ではヨガの経験者は5名(25%)であったが、対照群では20名中12名(60%)がヨガ経験者であり、統計学的に有意な差(p=0.03)がみられたという...
- 研究者らはヨガによる肝がん予防のメカニズムについて...



いくら p値 が正しくても、もともとの研究デザインがおかしければ意味がない！

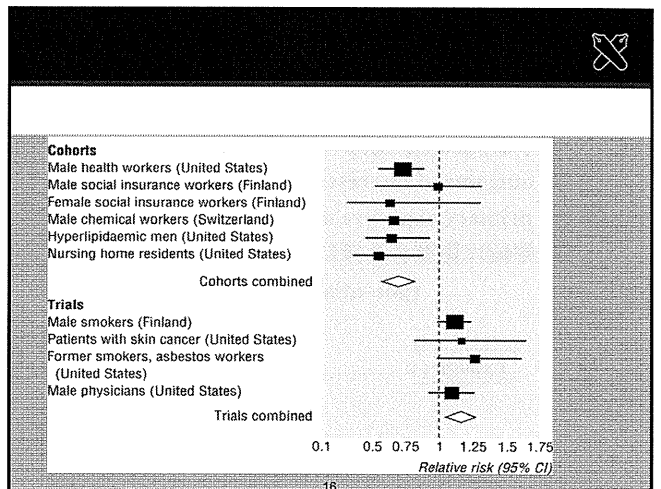
↓

研究が終了し、解析する段階であわてても手遅れ

↓

研究を始める前が肝心

15



デザイン 統計

デザイン 統計

デザイン 統計

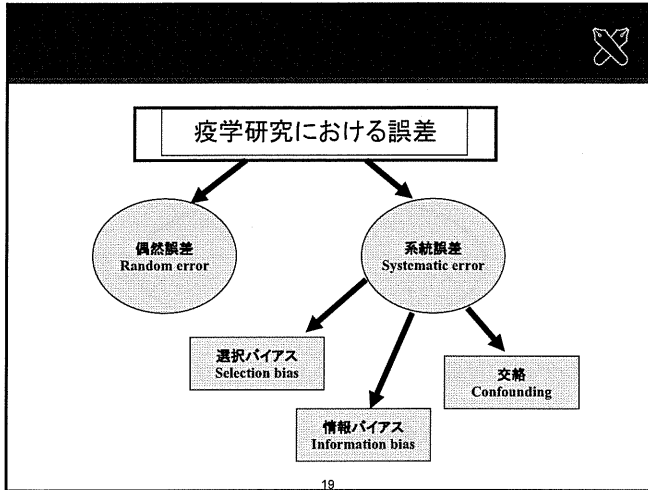
デザイン 統計

デザイン 統計

17

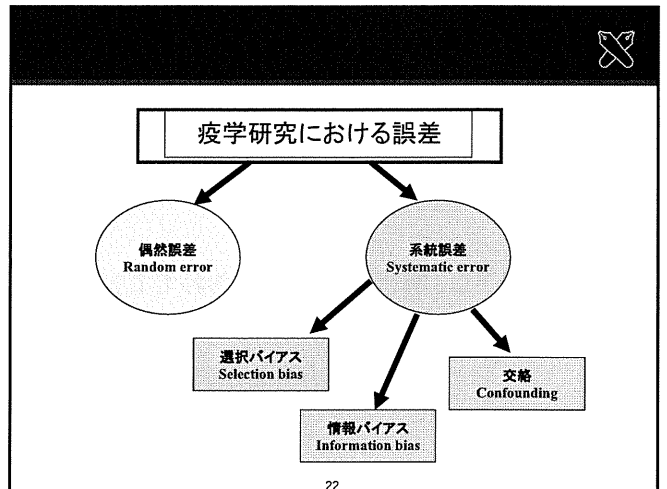
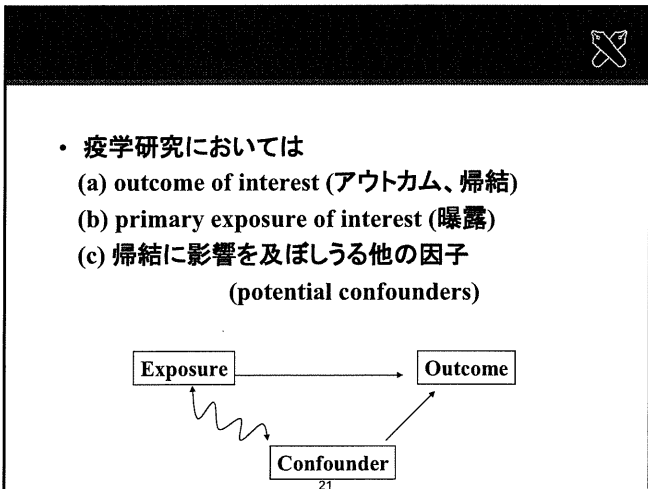
疫学研究につきものの
 [] の制御に尽きる

18



研究計画立案・研究計画書作成に
不可欠な以下の三つの概念を理解する

- ◆ Chance(偶然)
- ◆ Bias(バイアス)
情報バイアス・選択バイアス
- ◆ Confounding(交絡)



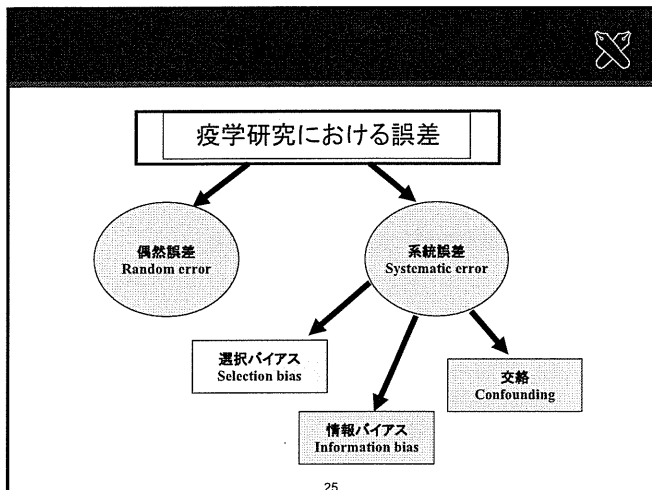
真値との差に方向性のない誤差

例1: サンプルングによる誤差
例2: 実測値の偶然のゆれ

- ◆ 偶然誤差が少ない研究を精度の高い研究と呼ぶ
- ◆ 何らかの「サンプリング」「測定」をすれば、必ずそこには偶然誤差が発生する。
 - 十分なサンプル数を準備する
 - データのばらつきを小さくする工夫

偶然誤差 (Random error) 精度 Precision

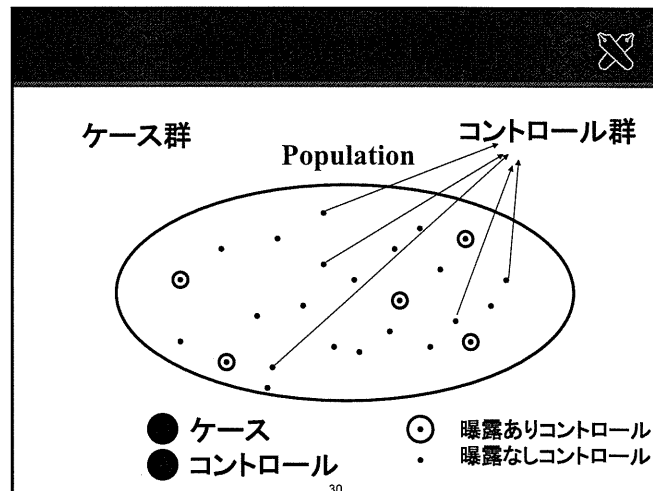
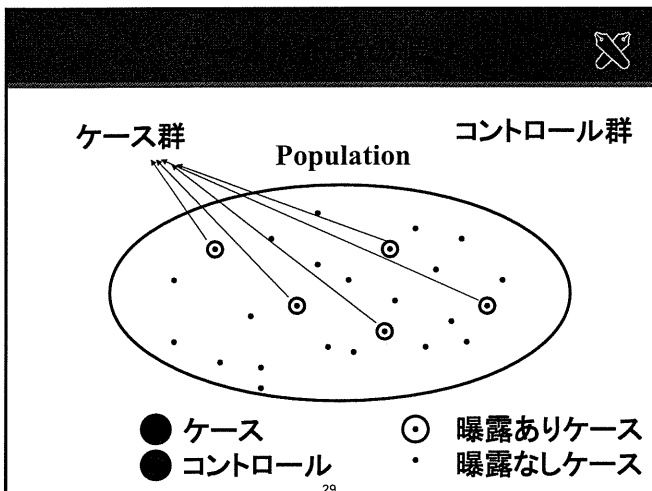
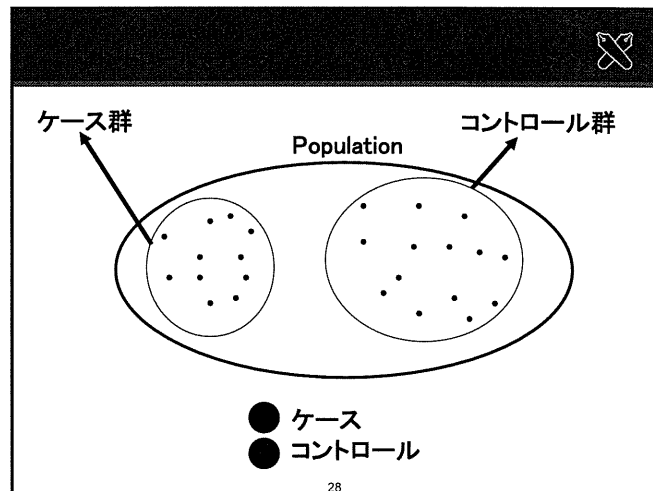
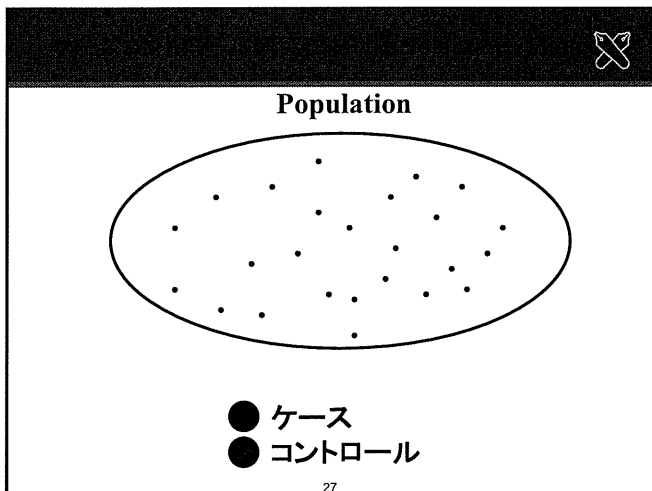
系統誤差 (Systematic error) 妥当性 Validity



• 研究対象者がどのように選ばれたか？

コホート研究：曝露群 vs 非曝露群
 ケースコントロール研究：ケース群 vs コントロール群
 介入研究：介入群 vs 非介入群

26





ケースコントロール研究 コントロールを選択する際のバイアス

喫煙と肺がんの関連を調べるためのケースコントロール研究である。
ケースは、2003年から2004年までの間に慶應病院で肺がんと確定診断を下された患者である。
一方コントロールは、同時期に慶應病院の呼吸器内科で慢性気管支炎と診断された患者である。

31



- ケースが抽出された母集団が不明確なことが多く、Selection biasがおきやすい

例) ケース: 慶應病院に入院した肺がん患者

コントロール:

もし肺がんになったら慶應病院に入院するであろう(慶應病院に入院する確率がケースと等しい)母集団から選ぶべき……

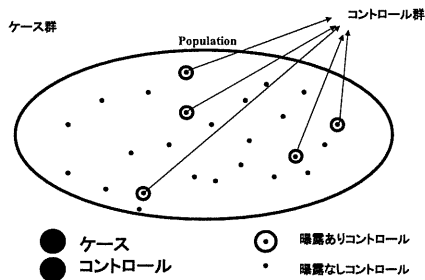
32



・他の疾患で入院している人

しかし、当該曝露と関連のある疾患では……

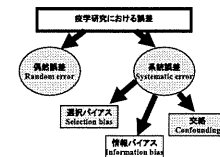
(慢性気管支炎も喫煙と関連がある)



33



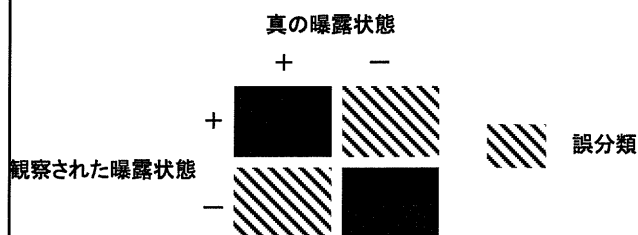
- 得た情報が正しくないために起こる偏り
- 曝露の情報、アウトカムの情報に基づいて、観察対象者を分類し解析するので、これらの情報の誤りは分類の誤りmisclassificationを引き起こす。



34



例 ケースコントロール研究



35



Non-differential misclassification

Differential misclassification

曝露(アウトカム)に依存しない
アウトカム(曝露)の誤分類

曝露(アウトカム)に依存した
アウトカム(曝露)の誤分類

曝露(+)群と曝露(-)群で同じように起こるアウトカムの誤分類
アウトカム(+)群とアウトカム(-)群で同じように起こる曝露の誤分類

曝露(+)群と曝露(-)群で異なって起こるアウトカムの誤分類
アウトカム(+)群とアウトカム(-)群で異なって起こる曝露の誤分類

36

✕

例： 緑茶の飲用と胃がんの関連に関する
ケースコントロール研究

曝露の測定：「1日2杯以上飲む」
「1日2杯未満」

誤分類：

- ・あいまいな記憶
- ・自己判断のばらつき (例：器のサイズ、ペットボトルのお茶は含めるか)
- ・記入間違い(質問票)

37

✕

Case(胃がん) Control

真の曝露

	+	-		+	-
質問票に 基づく曝露	+	+	+	+	+
	-	-	-	-	-

Case群とControl群で同じように起こる曝露の誤分類

38

✕

・ 効果の指標 (measures of effect) は、nullに近づく
(bias toward the null)

Odds ratio / Risk ratio / Rate ratio

2.4 ⇨ 1.0

0.1 ⇨ 1.0

39

✕

比較する群間で異なって
起きる誤分類

Recall Bias

Observer bias

40

✕

例： 人工流産の回数と乳がんリスクに関するケースコントロール研究

曝露の測定： 人工流産の回数をインタビューにより尋ねる

群間で異なっておきる誤分類：
 乳がん患者ではしっかりと思い出す
 コントロールの女性は回数の申告に消極的

	+	-		+	-
Case	+	+	+	+	+
	-	-	-	-	-
Control	+	+	+	+	+
	-	-	-	-	-

インタビューに
基づく曝露

41

✕

・ 効果の指標 (measures of effect) は、真値から遠ざかる
 ・ しかし、その方向(上がるか、下がるか)はケースバイ
 ケース

Odds ratio / Risk ratio / Rate ratio

3.6 ← 2.4 → 1.4

42

• アウトカム：死亡率
• 曝露：膝の痛み

The diagram shows '疫学研究における誤差' (Errors in epidemiological research) branching into '偶然誤差 Random error' and '系統誤差 Systematic error'. 'Systematic error' further branches into '選択バイアス Selection bias', '情報バイアス Information bias', and '交絡 Confounding'. To the right is an X-ray image of a knee joint.

43

結論：膝が痛い人は長生き(死亡率が低い)

本当か？

膝が痛い人X..... 死亡率

性別

交絡 というのは、他の因子で説明できること！！

The diagram shows '膝が痛い人' (People with knee pain) and '死亡率' (Mortality rate) connected by a dashed arrow with an 'X' through it, indicating a spurious correlation. A wavy arrow points from '性別' (Gender) to both '膝が痛い人' and '死亡率', indicating that gender is a confounder that explains both variables.

44

• 疫学研究においては

(a) outcome of interest (アウトカム、帰結)
(b) primary exposure of interest (曝露)
(c) 帰結に影響を及ぼしうる他の因子 (potential confounders)

The diagram shows a box for 'Exposure' with an arrow pointing to a box for 'Outcome'. A wavy arrow points from a box for 'Confounder' to the arrow between 'Exposure' and 'Outcome', indicating that the confounder influences the relationship between exposure and outcome.

45

1. アウトカムの独立したリスク因子である(関連がある)
2. 曝露とあいだに関連がある
3. 曝露とアウトカムの causal pathway 上にない

The diagram shows 'Exposure' leading to 'Outcome'. A 'Confounder' box has arrows pointing to both 'Exposure' and 'Outcome'. Two callout bubbles are present: one pointing to the arrow between 'Exposure' and 'Confounder' with the text '統計学的な関連でOK (分布に差がある)', and another pointing to the arrow between 'Confounder' and 'Outcome' with the text '統計学的な関連でOK (分布に差がある)'.

46

The diagram shows a box for '曝露 アルコール摂取' (Exposure Alcohol consumption) with an arrow pointing to a box for '肝硬変' (Liver disease), which in turn has an arrow pointing to a box for 'アウトカム 肝がん' (Outcome Liver cancer).

47

● 研究デザイン時

- 限定 Restriction
例1: 性別が交絡...男性もしくは女性のみ
例2: 年齢が交絡...一定の年齢層
- マッチング Matching
コホート研究(ケースコントロール研究?)
交絡因子となりうる因子をマッチ(適合)するように対象者を選んでいく
- Random assignment (randomization)
無作為化、無作為割り付け(介入研究)

48

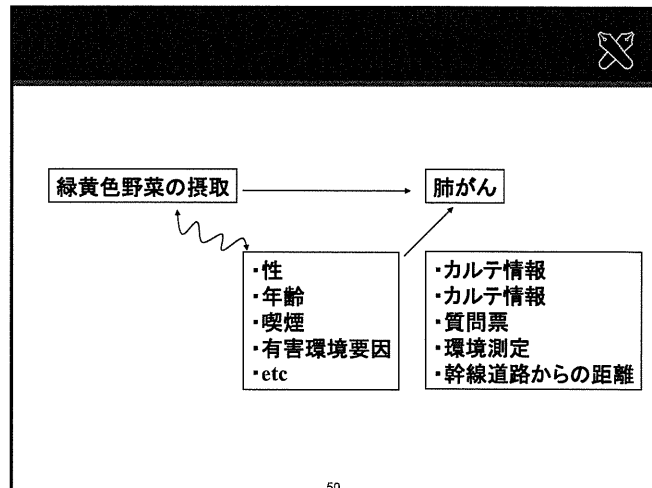
● 研究実施時

- 「交絡しうる」因子の情報収集
 - ・生物学的常識に基づく判断
 - ・文献・先行研究から
 - ・自分の研究で実際に「交絡」しているとは限らない

● 解析時

- 層化
- 多変量解析(重回帰分析、ロジスティック回帰分析、etc)

49



異なるのは、

- バイアスは、曝露や影響（結果）の直接関連して発生する系統的エラーであるが、交絡因子は、曝露や影響と関連のある第3の因子である
- バイアスは、一度発生したら、決して取り除いたり修正したりすることはできない。交絡因子は、研究に入り込んでいても、後で（解析時に）補正することが可能。

→ バイアス対策は、研究デザイン時にできるだけ発生しないように工夫すること（のみ）。

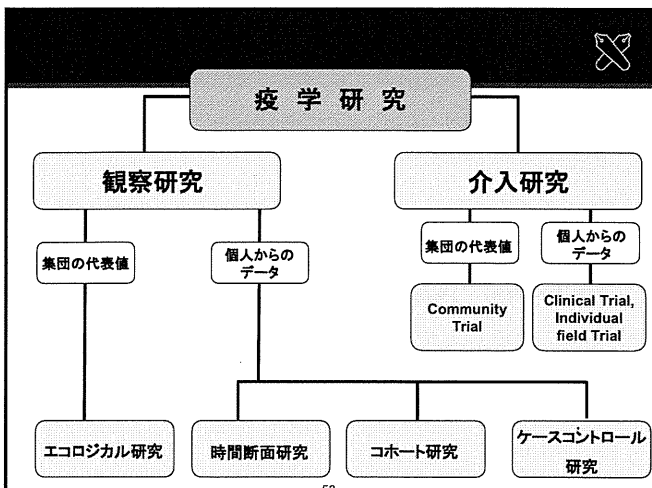
→ 交絡因子対策は、デザイン時にできるだけ発生しないように工夫すると同時に、解析時に交絡因子を発見し補正するために、「交絡因子になりうる要因」の情報を集めるようにすること。

51

【臨床疫学Ⅱ】
研究デザインの種類とその長所・短所

東邦大学 衛生学
西脇祐司

52



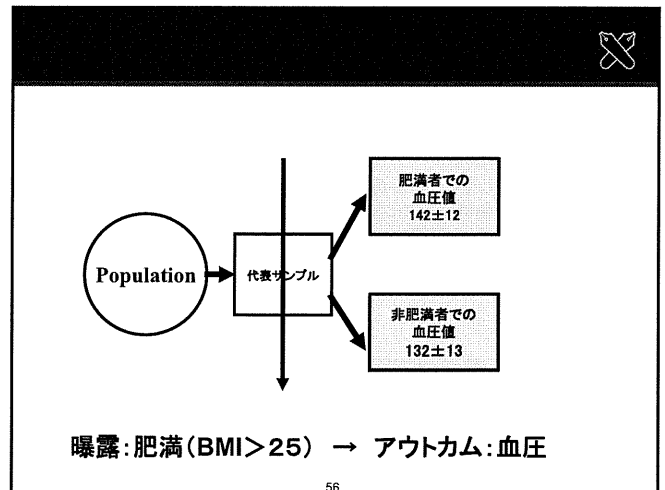
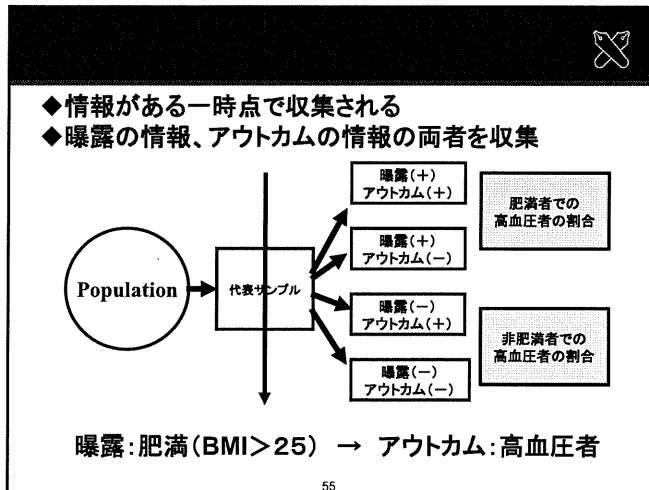
1. 病因解明のための臨床疫学研究方法論

- ◆ 時間断面研究
- ◆ コホート研究
- ◆ 症例対照研究

2. 診断・治療効果判定のための臨床疫学研究方法論

- ◆ 介入研究(臨床試験)

54



長所

- 複数の曝露を検討可能
- 複数のアウトカムを検討可能
- 仮説の設定

短所

- 関連の時間性の推定がときに困難
(どちらが原因で、どちらが結果か?)

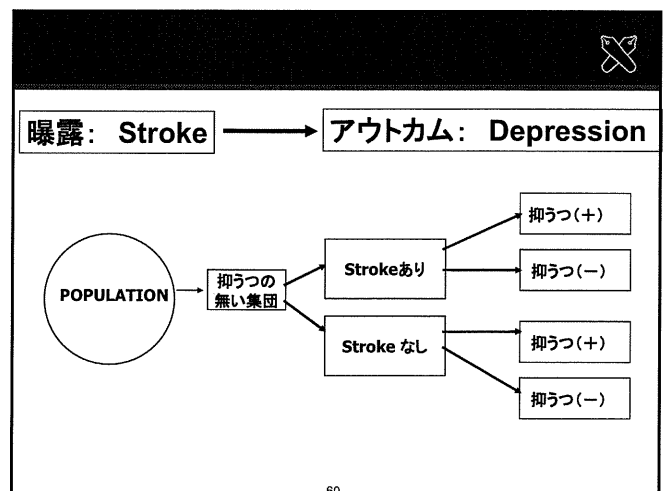
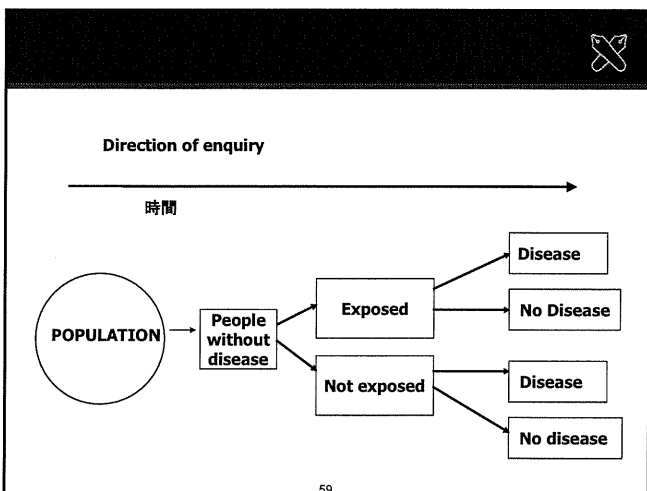
Ex. 喫煙者にパーキンソン病が少ない

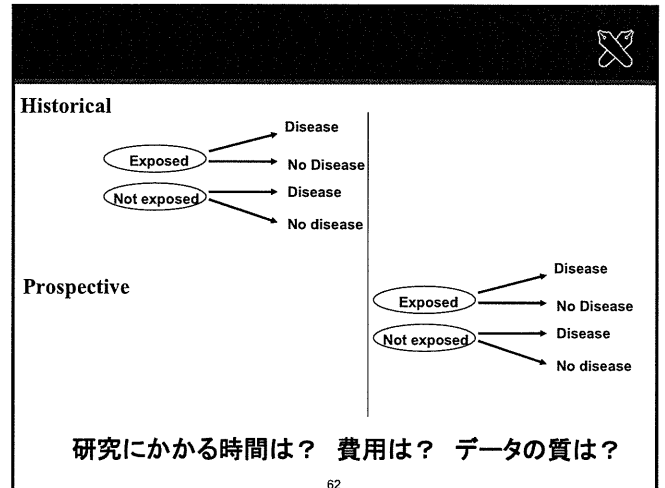
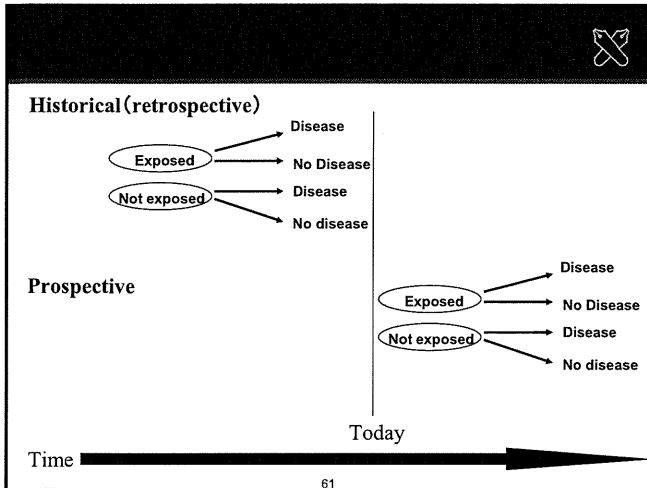
57

何らかの共通特性(例えば、同じ住所地、同じ職業、同じ学校、同一の曝露要因など)を持った集団を追跡し、その集団からどのような疾病・死亡が起こるかを観察し、要因と疾病との関連を明らかにしようとする研究 (疫学、南江堂)

Follow-up study 追跡研究
Longitudinal study 縦断的研究

58

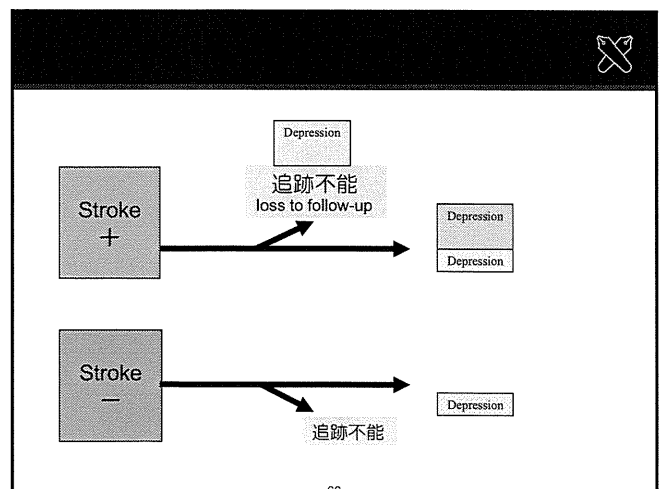


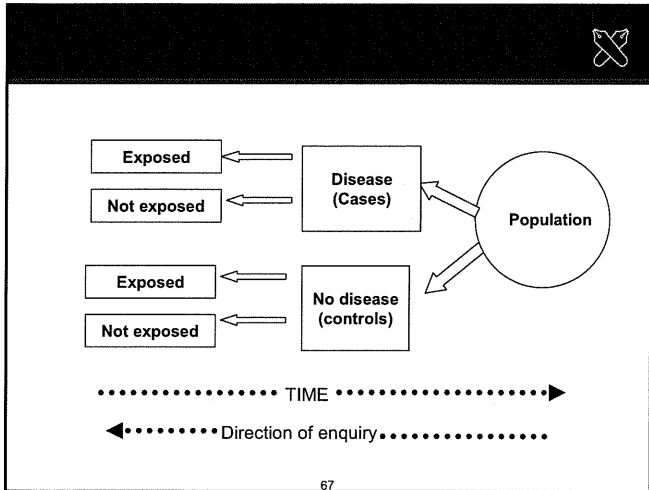


- 研究対象を選ぶ
 - 曝露の定義と測定
 - アウトカムの定義と測定
 - 交絡要因の測定
 - データの解析
 - 結果の解釈
- 63

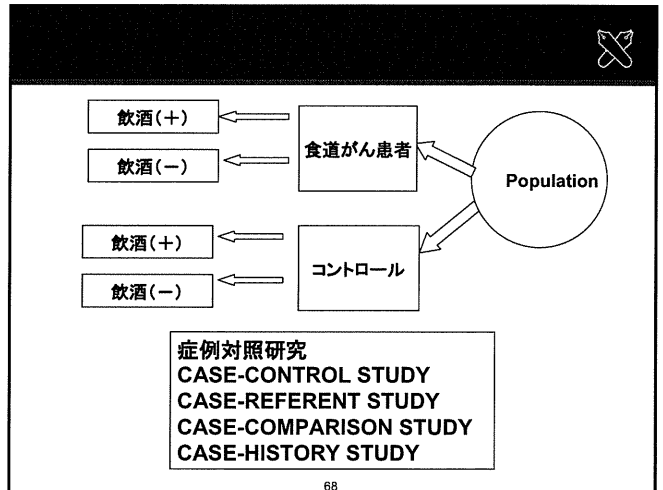
- 1) 比較的可能な曝露について研究可能
例: 特殊な農業と関節リウマチ
 - 2) 複数のアウトカム
例: 厚生省多目的コホート研究(がん、循環器、etc)
 - 3) 曝露がアウトカムより時間的に前に発生
・・・時間断面研究では？
 - 4) 交絡要因に関する情報も入手可能
例: アスベストと肺がん ・・・喫煙
 - 5) 量反応関係の評価が可能
例: 病気のgrade (I、II、III度)
アルコール摂取 (2合以上、1~2合、1合以下、飲まない)
 - 6) 罹患 (病気の新規発生) の情報が直接求まる
- 64

- 1) まれなアウトカムには適さない
 - 2) prospective cohort: 時間、費用がかかる
 - 3) historical cohort: 適切なデータが入手可能か
 - 4) Loss to follow-upによる研究妥当性の低下
- 65





67



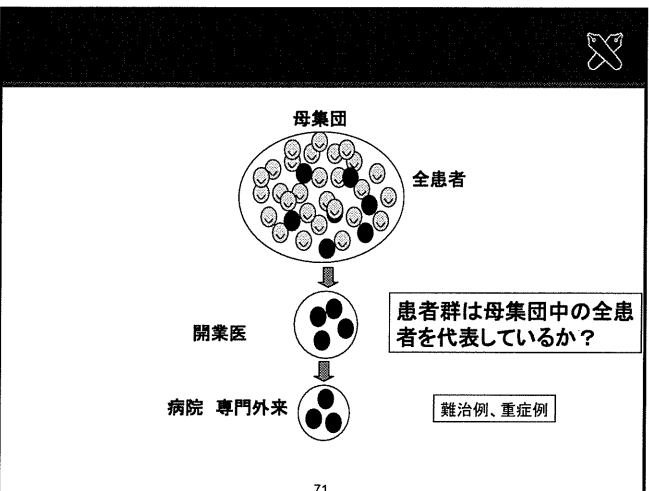
68

- 当該疾病やアウトカムをもつCase(患者群)の同定
- 当該疾病やアウトカムをもたない、適切なコントロール(対照群)の同定
- 曝露の prevalence(or level)を測定
- case群とコントロール群で曝露のprevalence(or level)を比較

69

- Definition**
- clear & precise (客観指標 例: 病理所見)
 - apply to all cases
 - inclusion and exclusion criteria (診断時の年齢、居住地)
- Source**
- 病院からのcase
 - 一般集団からのcase
- 集団の全ケース(患者)が選ばれないかぎり注意が必要！！**

70



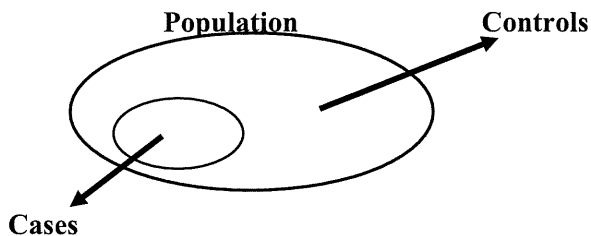
71

- Incident vs Prevalent cases**
- prevalent cases: survival case? (回復、死亡、移動例が除外)
 - 思い出し(recall)が不正確
 - 病気の結果、行動が変化? 例: がんと栄養
- Incident caseの方が望ましい?**

72



患者群が選ばれた母集団から、対照群も選ばれることが重要！！



73



マッチング(例:年齢、性)

- マッチングの目的は研究効率を上げること

有効な手段だが……

- matched analysisが必要
- マッチさせた因子については検討できない
- オーバーマッチングに注意
(曝露要因と密接な関係にある因子でのマッチング)

74



・コントロール群は何群必要か？

二つ以上のコントロール群 魅力的、でも危険！

・患者群とコントロール群のサイズの比は？

- 統計学的検出力は、コントロール：患者の比が増すにしたがって増加する。
- しかし、4：1 以上に増やしても、ほとんどメリットがない。

75



- 医療記録から曝露情報を入力
- データベースのリンク
- 生物学的サンプルの採取
- 参加者にインタビュー
- 参加者に郵便で質問
- 親戚に質問
- 医療関係者、その他に質問
- その他

76



ケースコントロール研究の長所

- 費用、時間的に有利(コホート研究に比べて)
- 同時に複数のリスク要因(曝露)を検討可能
- 必要であれば、リスク要因を詳しく検討可能
- まれな疾患、潜伏期間の長いに有用

ケースコントロール研究の短所

- selection bias, information biasに注意
- まれな曝露には不向き
- 通常、罹患率の推定はできない
- 関連の時間性が時に不明確(例:がんと栄養)

77



疫学的手法の中で研究担当者が人に“操作”を加える方法を“介入研究”と呼ぶ。

「対象者の状態のある側面を意図的に変容させるための研究である。例えば、予防または治療計画を導入し、仮説上の関連を検証するための研究。」
(疫学辞典)

介入の内容: 疾病予防や予後の改善に有利に作用する方向に目的とする要因への曝露状態を変更する。

例1: 血中コレステロールを下げるような食事指導で冠疾患の発生率を下げる事が可能か?(一次予防試験)

例2: 新薬が従来の薬より優れているか?(臨床試験)

実験疫学(Experimental epidemiology)ともいう。

78

臨床試験 (clinical trial)

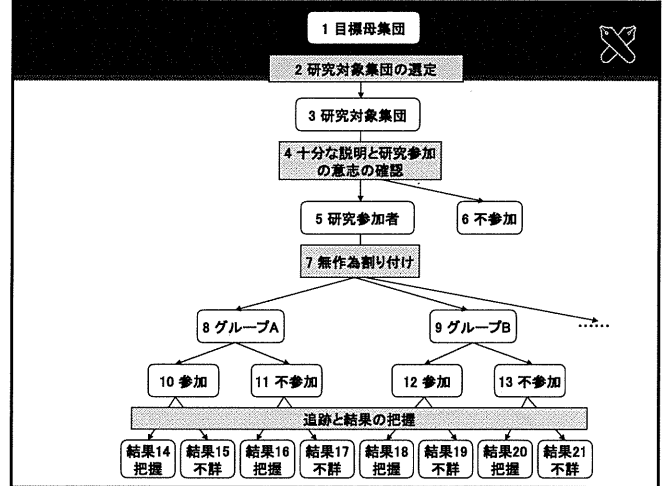
患者を対象とした介入研究。薬剤の治療効果、手術の効果などを評価する。対象者の選定にあたっては厳密な診断基準が必要で、対象に加えらるる症例は一定の水準で均一性を保つことが望まれる。(治療試験 therapeutic trial)

野外試験 (field trial)

対象者が目的とする疾患に罹患していない者、すなわち患者ではないことが臨床試験と異なる。疾病予防のための生活習慣の改善、検診、予防接種などの効果判定を行う。大規模な一般人口集団が対象になる場合が多い。大人数、長期追跡で、費用がかかる。

地域介入研究 (community intervention trial)

介入の対象が個人単位でなく地域単位の場合をいう。



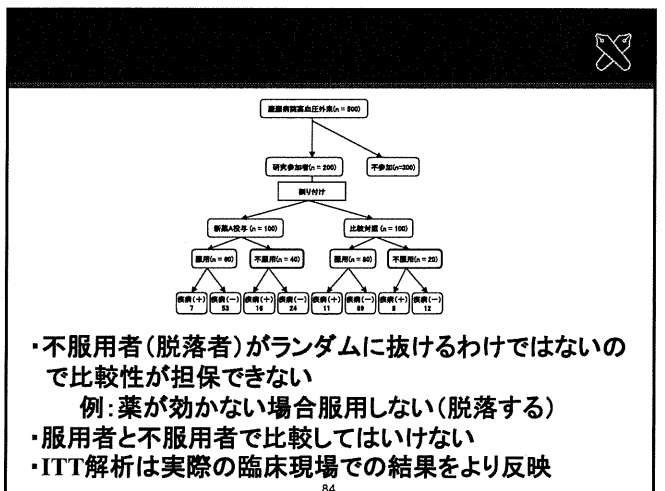
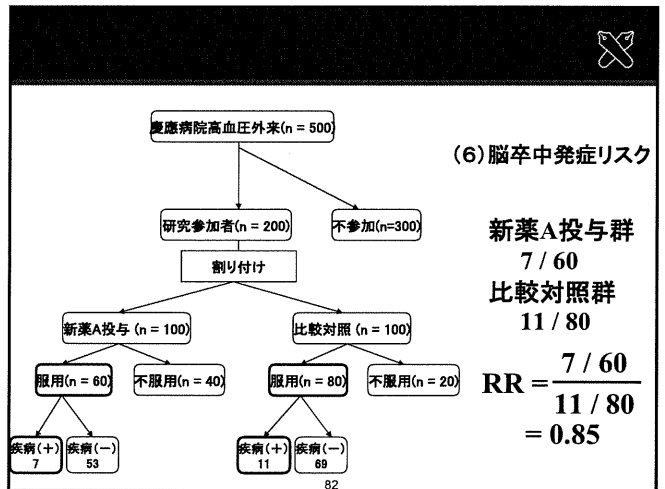
どういふ症例を解析する？

症例データの取扱い

薬を飲んでいない症例は？

途中でこなくなった症例は？

A剤に割り付けられたのにB剤を飲んだ症例は？





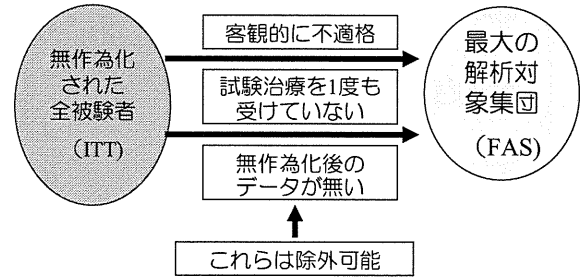
割り付け重視の原理 intention-to-treat principle

当該治療を受けたか否か、もしくは完了したか否かにかかわらず、それぞれの治療方式に割り当てられたすべての患者がその治療群を代表するものとして解析されるもの。

85



もっとも厳しいのは
割付通りの全被験者の完全な追跡結果



86

慶應義塾大学医学部CCR 教育研修プログラム

臨床研究講習会

臨床統計

2011年 6月18日
慶應義塾大学医学部
クリニカルリサーチセンター (CCR)
阿部 貴行

講義内容

- 試験デザインと統計解析
- 正確度と精度
- 統計的推測 (検定と推定)
- 症例数の設計
- 臨床試験データの解析事例

慶大医CCRセミナー2011 2/49

試験デザインと統計解析 design & analysis

慶大医CCRセミナー2011 3/49

様々な研究デザイン

“無作為化”できない
→ 偏りを減らす努力が必要

“無作為化”できる
→ 結果の解釈が容易

観察研究

介入研究

主に、

- ケースコントロール研究
- コホート研究
- 無作為化並行群間比較試験
- クロスオーバー試験

“無作為化”の可否は研究デザインの重要な点!

慶大医CCRセミナー2011 4/49

こんな場合も

これだけデータの数があれば十分?

ある疾患の患者集団の既存データ (n=3500)

データの適格性を検討すると...

- 基準A 適格
 - 基準B 適格
 - 基準B 不適格
- 基準A 不適格
 - 基準B 適格
 - 基準B 不適格

...

基準Z 適格 (n=120)

観察研究で、解析可能データを集めるのは一苦労!

慶大医CCRセミナー2011 5/49

医学部にて

- ある日、仕事が忙しかった医師Bは、研究デザインを知らせずにデータのみを統計家Aに送り、2標本 t 検定を依頼した
- 統計家Aは解析結果を医師Bに連絡した (「有意差ありです!」)

Q: ここでの問題は何だろう?

慶大医CCRセミナー2011 6/49

試験デザインと統計解析

- “試験デザイン”と”データ”は、セットではじめて意味をもつ
 - ケース1 : 無作為化比較試験
 - ケース2 : 観察研究
 - ケース3 : クロスオーバー試験

ケース1：無作為化比較試験

- Randomized controlled trials (RCT)

(a) 無作為化臨床試験のデータ

Placebo (n=10)		Drug A (n=10)
156	比較可能な集団 (comparable) ←→	73
171		81
133		103
...		...
130		108
105		99

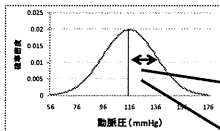
2標本 t 検定を用いて、平均の比較を行うことは適切!

無作為化

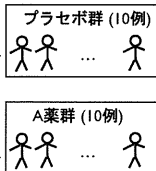
- 20世紀頃から積極的に活用



Sir Ronald A. Fisher
(1890-1962)
英国の統計学者
Wikipedia より



試験に参加した集団の背景因子の分布 (イメージ図)



- ポイント
- 2つの群は1つの集団を2つに分けたもの
- 背景因子に偏りがなく比較可能!

ケース2：観察研究

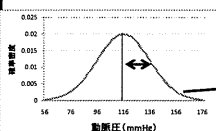
- Observational studies

(b) 観察研究のデータ

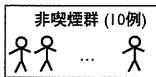
喫煙者 (n=10)		非喫煙者 (n=10)
156	交絡因子に注意! ←→	73
171		81
133		103
...		...
130		108
105		99

- 2標本 t 検定のみによる、平均の比較は不適切!
- 2群が “比較可能” であるかが重要: (例) “交絡因子” の偏り

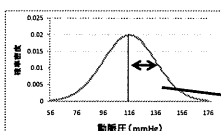
無作為化ができない場合



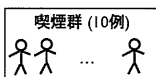
非喫煙者の背景因子



無作為化をしていないと系統的な偏りがあるかもしれない!



喫煙者の背景因子



ケース3：クロスオーバー試験

- Crossover trials

実は10例分のデータ!

表：擬似データ

Placebo (n=10)		Drug A (n=10)
156	同一被験者 ←→	73
171		81
133		103
...		...
130		108
105		99

- 対応のあるデータに2標本 t 検定は使えない!
- 検定の前提条件 (データの “独立性”) が不成立

試験デザインについてのまとめ

- 統計家との打ち合わせでは、研究デザインの詳細の話をしましょう
- デザインを知らずにデータ解析してはいけません！

正確度と精度

accuracy & precision

正確度と精度

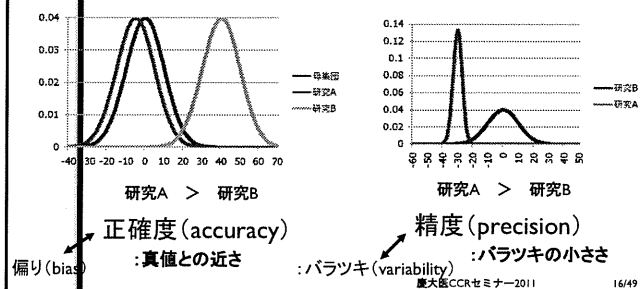
- 英語では、accuracyとprecision
- 日本語だと同じようにみえる
- 妥当性と信頼性に対応

Q: それぞれ何を測っているか？

- ヒント
 - accuracyの例：診断薬Aの“感度”は98%
 - precisionの例：血小板数(μL)の“SD”は...

正確度と精度

目標：偏りが少なく精度が高い研究



解析対象集団では？

Intention-to-Treat (ITT)

= Full Analysis Set (FAS; 最大の解析対象集団)

- 完全な集団
- 除外による偏りが無い
- 日常の医療を反映
- 検証試験

Per Protocol Set (PPS)

= 試験実施計画書に適合した集団

- 効果を検出しやすい
- ワクセン開発
- 探索的な試験

試験結果の安定性の確認が重要！

主要な解析対象集団を事前に明記。
→ “偏り”, 試験結果の “保守性”などを勘案

臨床研究では？

- 偏り
 - 適切な“選択/除外基準”
 - 無作為化
 - 二重盲検化
 - マッチング
- 精度
 - 試験実施手順の標準化(国, 医療機関間の変動)
 - 試験実施者のトレーニング

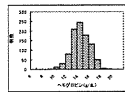
正確度と精度のまとめ

- accuracy と precision の区別は重要
- 違う意味をもつ専門用語
- accuracy と precision の観点から医学研究を吟味してみよう
 - 自分の研究の結果をまとめる際
 - 他の研究者の論文を読む際

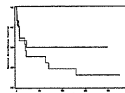
統計的推測 statistical inference

評価項目の型

- 連続型データ
 - (例) 血圧, LDL-C, 症状点数
- カテゴリカルデータ
 - 2値変数 (例) 有効/無効
 - 名目変数 (例) 性別
 - 順序変数 (例) 重症度
- 生存時間データ
 - 死亡までの日数など



	有効	無効	
プラセボ	20	80	100
A薬	60	40	100

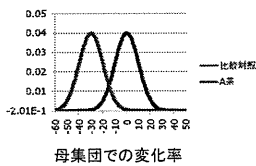


統計手法も様々

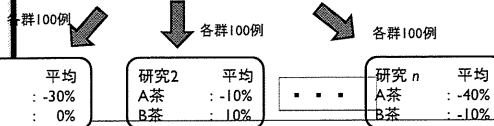
- 分散分析モデル
- ノンパラメトリック法
- 分割表データ解析
- 生存時間データ解析
- 一般化線形モデル
- 混合効果モデル
- 多重比較法, 中間解析
- 欠測データの解析
- ベイズ流の解析

• 医学研究(特にデザイン作成)では、臨床家と生物統計家の共同作業が必須!

母集団と標本：サンプリング誤差

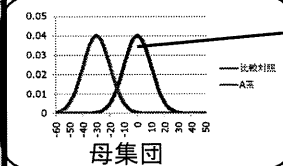


母集団 (population)
(例) LDL-C が高い日本人の集団全体



データ(標本)は、“サンプリング誤差”をもつ

母集団と標本：統計的推測



パラメータ：
分布を決める量
(例：平均, 分散)

統計的推測

- データから母集団のパラメータに関する推論
- “推定”や“検定”

データ (n=100)	標本平均
	A茶 : -20%
	B茶 : 5%

“不確実性”を含む
“確からしい”命題を導く

検定のロジック

- 平均値を2群間で比較したい
(差があるかないか?)

表：ある時点の平均動脈圧 (mmHg) ; RCTの結果

Placebo (n=11)	Drug A (n=11)
156	73
171	81
133	103
...	...
130	108
105	99

P=0.0051

Glantz (2005) P.123より引用

検定のロジック

仮説検定の手順

(1) 帰無仮説と対立仮説を設定

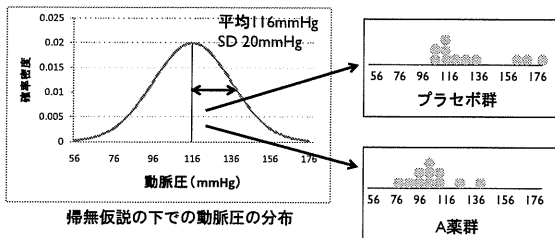
H_0 (差がない): A薬 = プラセボ
 H_1 (差がある): A薬 \neq PBO 又は (A薬 > PBO)

(2) 検定手法を選び、検定統計量を計算

(3) P値から有意性を判定

帰無仮説のイメージ図

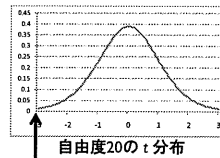
- 帰無仮説 (差がないという仮説)
→ 1つの分布から無作為サンプリング



P値の計算

- P値 (P-value)
 - 差がない時に、試験結果 (検定統計量) よりも極端な結果となる確率
 - 試験結果と帰無仮説との整合をみている

帰無仮説の下で、
検定統計量とその
分布を考える



試験結果から
計算したt値 = 3.14

P値
片側 P = 0.0026
両側 P = 0.0051

解析結果 (動脈圧データ)

• 2標本t 検定の結果

t検定: 等分散を仮定した2標本による検定

	プラセボ	A薬
平均	128.9	102.5
分散	498.7	281.9
観測数	11	11
プールされた分散	390.3	
仮説平均との差異	0	
自由度	20	
t	3.14	
P(T<=t) 片側	0.0026	
t境界値 片側	1.72	
P(T<=t) 両側	0.0051	
t境界値 両側	2.09	

結論:
群間に
有意差あり!

検定の前提条件

- t 検定の前提条件
 - データはランダム標本である
 - 各データは独立である
 - データは正規分布に従う
 - データのバラツキが2群で等しい

症例数設計

慶大医CCRセミナー2011

31/49

医学論文でよくある記述

- Hay E.M. et al. (2005). *Lancet*. より抜粋

Statistical Analysis

The sample-size calculation was based on change in RMDQ at 12 months after randomisation. A previous study in primary care showed a mean reduction in RMDQ score of 5.3 (SD 5.8) over 3 months. To detect a clinically significant difference of 2 points between the two treatment groups, with a significance level of 5% (2-tailed) and 90% power, 180 patients were needed in each treatment group.

2処置を比較した
無作為化臨床試験

慶大医CCRセミナー2011

32/49

症例数設計

- 検定における2種類の過誤
 - ・ 差がないときに差があるとする (α)
 - ・ 差があるときに差がないとする (β)
- デザインの段階でこれらを適切に制御する必要がある
 - ・ 検証的な試験では莫大なコストがかかる
 - ・ β が50%のような試験は実施不可能

慶大医CCRセミナー2011

33/49

検定：2種類の過誤

		真実の状態	
		差がない (H_0)	差がある (H_1)
検定の結果	H_0 を棄却	α	$1 - \beta$
	H_0 を棄却せず	$1 - \alpha$	β

第一種の過誤

(例) 無効な薬を有効であると結論づける誤り

第二種の過誤

(例) 有効な薬を有効とは言えないと結論づける誤り

慶大医CCRセミナー2011

34/49

症例数設計

2群の比較 (連続型の評価項目)

必要なパラメータ

- 評価項目と検定手法
- 効果の大きさ (ES)
- 有意水準
- 検出力

過去のデータなどからESを精度よく見積もることが重要!

通常, 両側5%

通常, 80%以上

慶大医CCRセミナー2011

35/49

効果の大きさ (effect size)

- 連続量データ

$$ES = \frac{\text{平均値の群間の差}}{\text{標準偏差}}$$

- 2値データ
 - 2群の割合 (差, odds ratio and relative risk)
- 生存時間データ
 - ハザード比, イベント発生率

慶大医CCRセミナー2011

36/49