

交絡 Confounding

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！！

37

Exposures and Outcomes

- 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)

38

交絡因子

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

39

Causal pathway

40

交絡要因に対する配慮

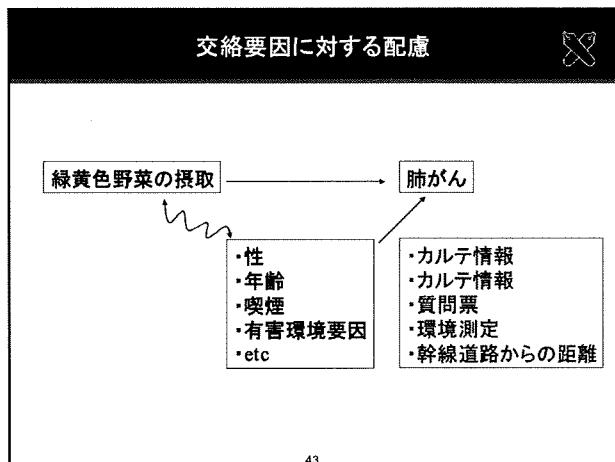
- 研究デザイン時
 - 限定 Restriction
 - 例1: 性別が交絡... 男性もしくは女性のみ
 - 例2: 年齢が交絡... 一定の年齢層
 - マッチング Matching
 - コホート研究(ケースコントロール研究?)
 - 交絡因子となりうる因子をマッチ(適合)するように対象者を選んでいく
 - Random assignment (randomization)
 - 無作為化、無作為割り付け(介入研究)

41

交絡要因に対する配慮

- 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)

42



＜バイアスと交絡因子＞

- 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）。

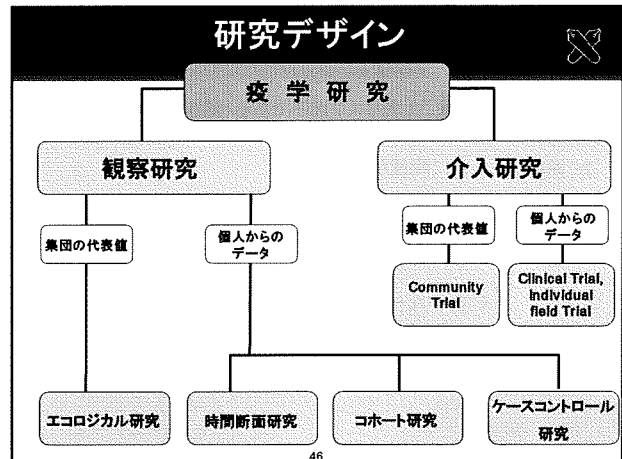
→交絡因子対策は、デザイン時にできるだけ発生しないように工夫すると同時に、解析時に交絡因子を発見し補正するために、「交絡因子になりうる要因」の情報を集めるようにすること。

平成21年度クリニカルリサーチセンター教育研修プログラム
臨床研究入門コース 2009/12/05

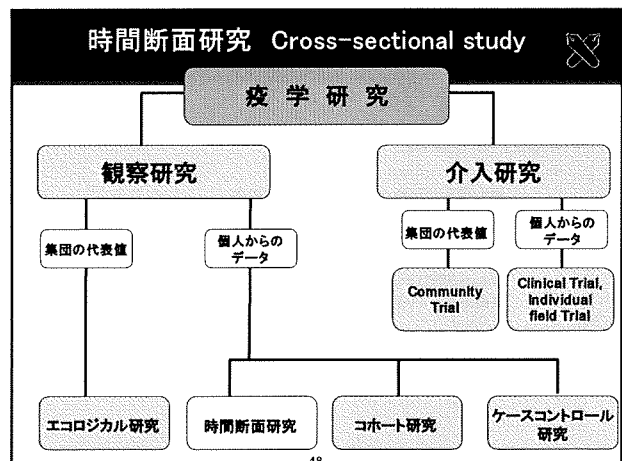
【臨床疫学Ⅱ】 研究デザインの種類とその長所・短所

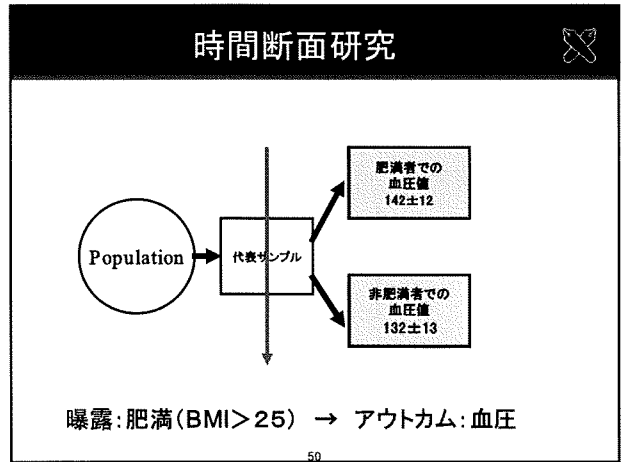
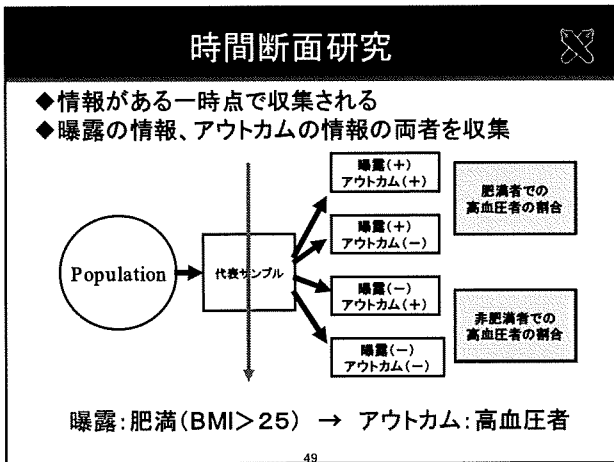
衛生学公衆衛生学
西脇祐司

45



- ### 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)
- 47





時間断面研究の長所・短所

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

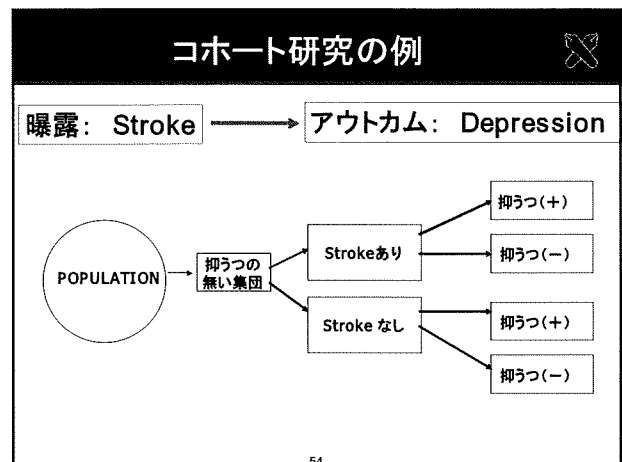
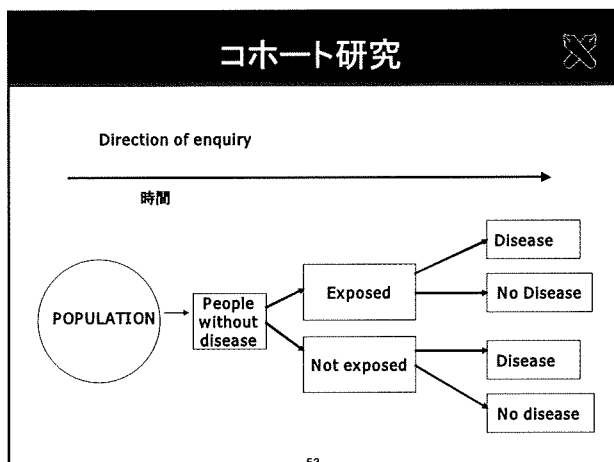
51

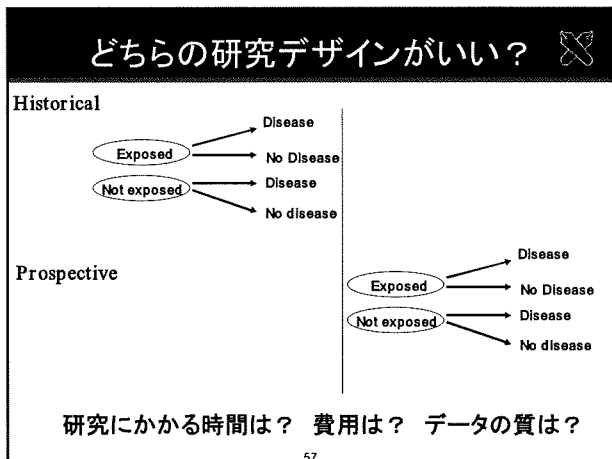
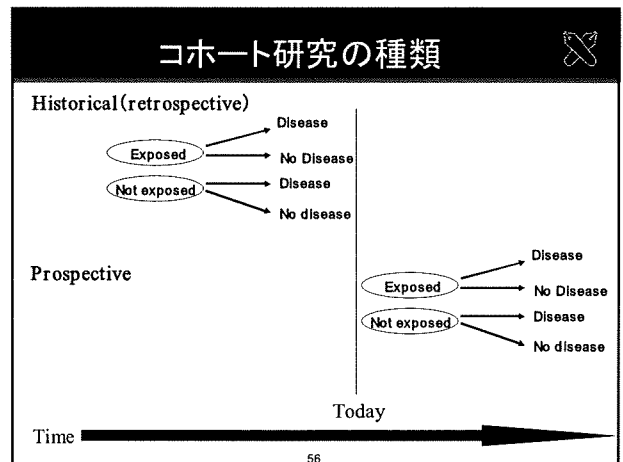
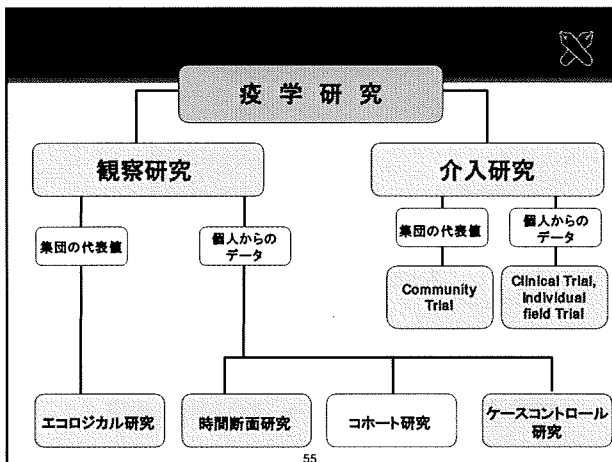
コホート研究 Cohort Study

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

52





- ### コホート研究の重要な要素
- 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
- 58

- ### コホート研究の長所
- 1) 比較的可成りな曝露について研究可能
例: 特殊な農薬と関節リウマチ
 - 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
- 59

- ### コホート研究の短所
- 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
- 60

コホート研究におけるバイアス

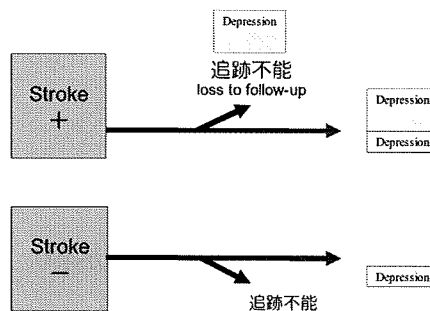


stroke+群とstroke-群の比較性comparabilityが保たれている

- 結果要因の情報 = depressionの評価プロセス
 - 患者が、depressionの原因をstrokeと思いこんで解答
 - 主観的な診断基準で、strokeの有無がdepressionの評価に影響
- 曝露要因の情報 = 研究参加者の確定プロセス
 - 将来のdepression発症傾向を知らなからリクルート?
 - コホート(追跡)研究ではselection biasはおこりにくい

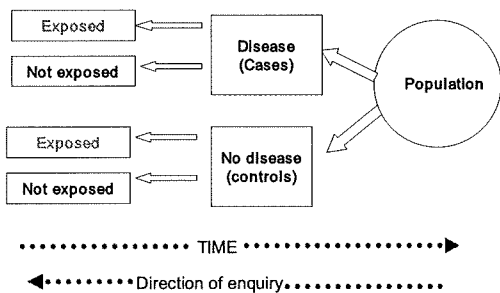
61

コホート研究の選択バイアス



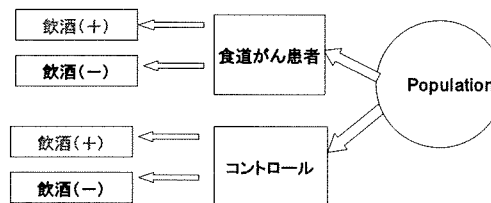
62

ケースコントロール研究



63

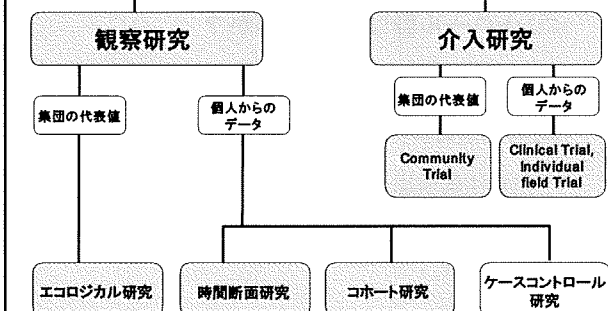
ケースコントロール研究



症例対照研究
CASE-CONTROL STUDY
CASE-REFERENT STUDY
CASE-COMPARISON STUDY
CASE-HISTORY STUDY

64

疫学研究



65

ケースコントロール研究

- 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

66

ケース(患者)群の選定

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)

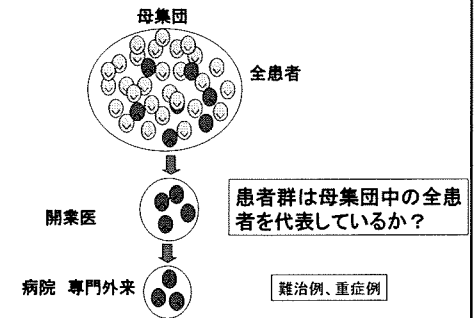
Source

- 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！！

67

ケース(患者)群の選定



68

ケース(患者)群の選定

Incident vs Prevalent cases

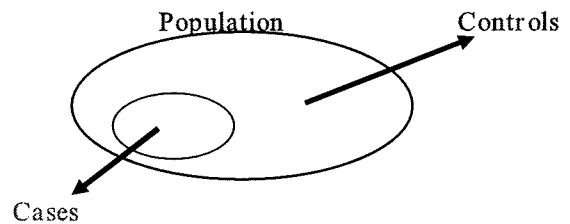
- prevalent cases: survival case? (回復、死亡、移動例が除外)
- 思い出し(recall)が不正確
- 病気の結果、行動が変化?
例: がんと栄養

Incident caseの方が望ましい?

69

コントロール(対照)群の選定

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！！



70

誰をコントロール群にすべきか?

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)

71

誰をコントロール群にすべきか?

- 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。

72

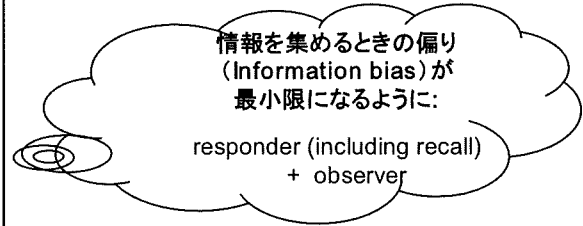
ケース群とコントロール群で 曝露prevalenceを評価



- 医療記録から曝露情報を入手
- 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

73

ケース群とコントロール群で曝露の prevalence (or level)を比較



74

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)

75

介入研究 (Intervention study) とは？

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。

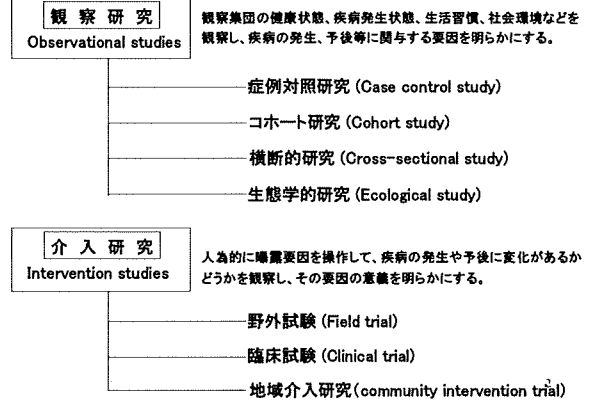
「対象者の状態のある側面を意図的に変容させるための研究である。例えば、予防または治療計画を導入し、仮説上の関連を検証するための研究。」(疫学辞典)

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のが可能か？(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。

疫学研究方法の分類



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、対象に加えられ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。

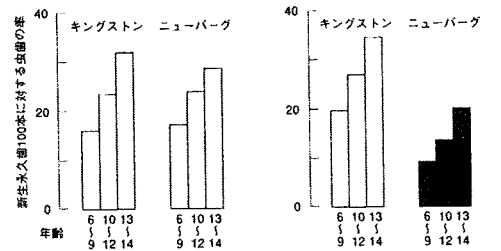
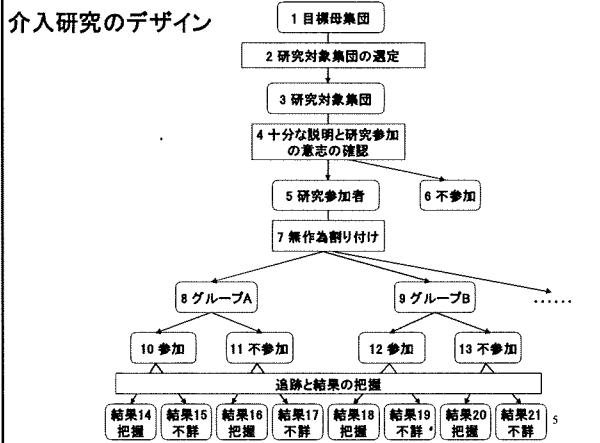


図 1.5 虫歯発生に対するフッ素の効果 (Morris, 1957)

介入研究のデザイン



◇ Sample size (サンプル数)

1. Effect size
2. Alpha error ($P\alpha$)
3. Beta error ($P\beta$)
4. Characteristics of the data
 - nominal scale: the rate of events
 - continuous variable: variation

		検定の結果	
		差を認める	差を認めない
母集団の真の姿	差がある	○	β error
	差がない	α error	○

Table 9.2
Determinants of Sample Size

	Determined by	
	Investigator	The Data
N varies according to	$\frac{1}{\Delta, P_{\alpha}, P_{\beta}}$	and $\frac{V}{I}$ or $\frac{1}{P}$

Where n = number of patients studied; Δ = size of difference in outcome between groups,
 P_{α} = probability of an α (Type I) error, i.e., false-positive results; P_{β} = probability of a β
(Type II) error, i.e., false-negative result; V = variability of observations (for interval data);
and P = proportion of patients experiencing outcome of interest (for nominal data)

慶應義塾大学 医学部
臨床研究講習会
『臨床統計』

2009年 12月 5日
クリニカルリサーチセンター 客員講師
阿部 貴行

慶應大・医 CCR教育研修プログラム 2009 1/41

本日の内容

例題を用いた解析結果の読み方

- 統計学の基本知識
- 医学研究の例題
 - (1) アレルギー性鼻炎の臨床試験
 - (2) 心疾患イベントのリスク因子に関する研究
 - (3) 骨髄腫の臨床試験

慶應大・医 CCR教育研修プログラム 2009 2/41

前回(入門コース)の内容

仮説の設定 → 臨床試験のデザイン → 臨床試験の実施 → 統計解析・結果の要約

臨床研究のプロセス

研究デザイン作成段階

- ・ Fisherの3原則
- ・ 無作為化
- ・ 母集団と標本
- ・ 解析対象集団
- ・ サンプルサイズの計算

解析結果の解釈

- ・ 推定と検定
- ・ SDとSE
- ・ 検定のロジック

慶應大・医 CCR教育研修プログラム 2009 3/41

医学研究の例題を用いた 解析結果の読み方

慶應大・医 CCR教育研修プログラム 2009 4/41

CCRの統計相談の内容

表: CCRでの統計相談内容の要約

反応の型	相談内容	頻度
連続型データ (連続変量)	分散分析モデルの構築 共分散分析の結果の解釈 ノンパラ検定	中～高
2値データ (離散変量)	ロジスティック回帰モデルの構築 統計出力(オッズ比)の解釈 オッズ比の信頼区間の構成法	高
生存時間データ	Cox回帰モデルの構築 統計出力(ハザード比)の解釈 生存時間解析の例数設計(ログランク検定)	低～中

慶應大・医 CCR教育研修プログラム 2009 5/41

よくある解析結果

表: 鼻症状点数の治療期2週後のベースラインからの変化量
(解析対象集団: FAS)

投与群	ベースライン		変化量(2週間)		投与群間の差 (95% CI) 新薬 - プラセボ	P値*
	N	Mean	LSmean	SE		
プラセボ	314	1.53	-0.37	0.02	-	-
新薬 5mg	318	1.54	-0.47	0.02	-0.10 (-0.16, -0.04)	0.001
新薬 10mg	310	1.52	-0.47	0.02	-0.11 (-0.17, -0.05)	0.001

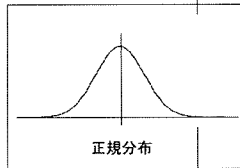
K. Okubo et al (2008) より抜粋
#: Comparison with Placebo using the ANCOVA model which contains treatment group, study site as factors, and baseline as a covariate.

慶應大・医 CCR教育研修プログラム 2009 6/41

統計学の基本知識

○ 推定と検定

- 標準偏差(SD)と標準誤差(SE)
- 信頼区間の構成法
- 検定の手順(ロジック)



慶應大・医 CCR教育研修プログラム
2009

7/41

推定と検定

○ 推定(estimation)

- 効果(パラメータ)の大きさを推定
- 95%信頼区間の幅 → 研究の精度
- P値だけでなく、信頼区間を示すことが推奨

○ 検定(test)

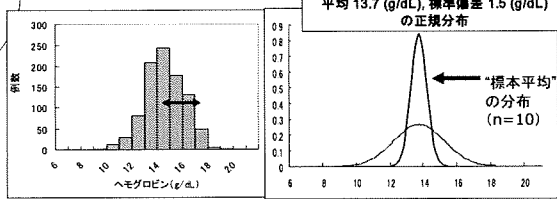
- 仮説を設定し、仮説の尤もらしさを判定する方法
- P値の計算(有意差あり・なし)

慶應大・医 CCR教育研修プログラム
2009

8/41

標準偏差(SD)と標準誤差(SE)

○ ヘモグロビン(g/dL)データ



標準偏差(SD): データ自体の分布のバラツキ
標準誤差(SE): “標本”平均の理論的なバラツキ

$$SE = \frac{SD}{\sqrt{N}}$$

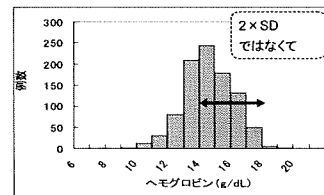
統計量

慶應大・医 CCR教育研修プログラム
2009

9/41

平均値の95%信頼区間

- 信頼度95%で真値を含む範囲
- (例) ヘモグロビン(g/dL)の平均値



慶應大・医 CCR教育研修プログラム
2009

例: 95%CI: (12.7, 14.7)
→ 標本平均 ± 2 x SE

検定の手順

(1) 仮説の設定

例: 優越性, 非劣性など

○ 帰無仮説: 差がないという仮説
○ 対立仮説: 差があるという仮説

片側検定: 新薬 > プラセボ
両側検定: 新薬 ≠ プラセボ

(2) 検定統計量の選択

例: t検定, ノンパラなど

● データから検定統計量を計算
例: t値, F値など

(3) 有意性の評価

例: P値の計算

● P値を有意水準と比較
片側検定: $\alpha = 0.025$
両側検定: $\alpha = 0.05$

慶應大・医 CCR教育研修プログラム
2009

11/41

検定: 2種類の過誤

		真の状態	
		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 を採択	$1 - \alpha$	β

第一種の過誤

(例) 無効な薬を有効であると結論づけてしまう誤り

第二種の過誤

(例) 有効な薬を有効ではないと結論づけてしまう誤り

慶應大・医 CCR教育研修プログラム
2009

12/41

検定：用語の整理

- P値 (P-value)
 - 差がないと仮定した時、その試験結果の出る確率
 - $P < 0.05$ → 仮定が正しくないと考え、有意差あり
 - $P \geq 0.05$ → 有意差なし (≠差がない)
- 有意水準 (significance level)
 - 許容する α : 両側 5% 又は 片側 2.5%
- 検出力 (power)
 - $1 - \text{許容する } \beta$: 80% 以上

慶應大・医 CCR教育研修プログラム
2009

13 / 41

3つの医学研究の例題

慶應大・医 CCR教育研修プログラム
2009

14 / 41

3つの例題

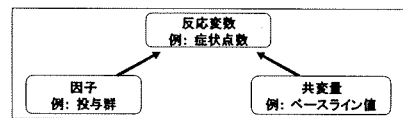
評価項目	例題	目的／統計手法
連続量	アレルギー性鼻炎の臨床試験	新薬とプラセボの比較 ／ANCOVA
2値	心疾患イベントのリスク因子に関する研究	心疾患のリスク因子の探索 ／ロジスティック回帰
生存時間	骨髄腫の臨床試験	薬剤AとBの比較 ／Cox回帰

慶應大・医 CCR教育研修プログラム
2009

15 / 41

(1) 評価項目が連続型の場合

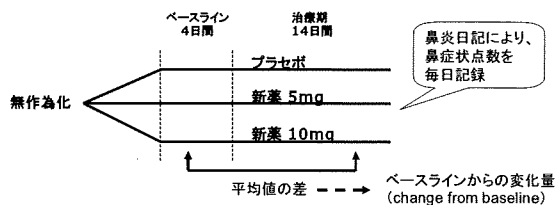
- 連続型の評価項目
 - 血圧, 臨床検査値, 症状評価尺度の点数など
- 統計手法
 - 2標本 t 検定 → デザインベース
 - ANOVA : Analysis of Variance
 - ANCOVA : Analysis of Covariance → モデルベースの解析



16 / 41

例題1: アレルギー性鼻炎患者を対象としたプラセボ対照二重盲検試験

- 研究目的
 - 鼻症状点数の改善を主要評価項目(エンドポイント)として、新薬群とプラセボ群を比較する



慶應大・医 CCR教育研修プログラム
2009

17 / 41

共分散分析(ANCOVA)モデル

$$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \text{誤差}$$

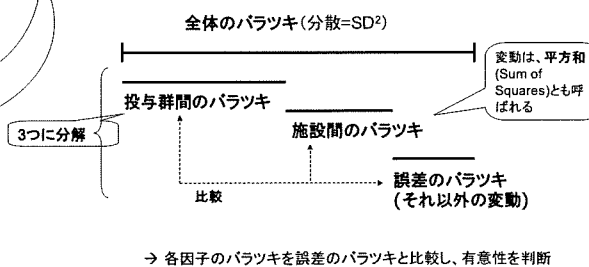
反応: 症状点数の変化量
 切片
 投与群 (x_1, x_2): プラセボ (0, 0); 新薬 5mg (1, 0); 新薬 10mg (0, 1)
 症状点数のベースライン値 (連続型) (x_3)

線形モデル (Linear Model) → 反応をパラメータの足し算で表現

慶應大・医 CCR教育研修プログラム
2009

18 / 41

分散分析(ANOVA)の概略

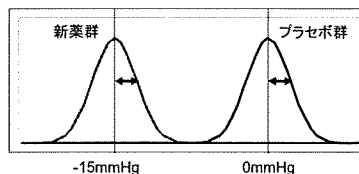


慶應大・医 CCR教育研修プログラム
2009

19 / 41

分散分析での仮定

- データは正規分布に従う → 正規性
- 群間で分散が等しい → 等分散性



慶應大・医 CCR教育研修プログラム
2009

20 / 41

共分散分析では、LSmeanという言葉がよく出てくるが？

表: アレルギー性鼻炎の薬剤の臨床試験成績

投与群	ベースライン		変化量(2週間)		投与群間の差 (95% CI)	P値*
	N	Mean	LSmean	SE		
プラセボ	314	1.53	-0.37	0.02	-	-
新薬 5mg	318	1.54	-0.47	0.02	-0.10 (-0.16, -0.04)	0.001
新薬 10mg	310	1.52	-0.47	0.02	-0.11 (-0.17, -0.05)	0.001

慶應大・医 CCR教育研修プログラム
2009

21 / 41

LSmean とは

- LSmean (Least Squares Mean)

- 最小二乗平均
- 調整済み平均

- なぜ Mean でなく LSmean を用いるのか？

- 偏りを減らす
→ 処置群間で、反応と相関の強い共変量の分布に偏りがある場合、標本平均は偏りを持つ
- 検出力を上げる

慶應大・医 CCR教育研修プログラム
2009

22 / 41

共分散分析の概略

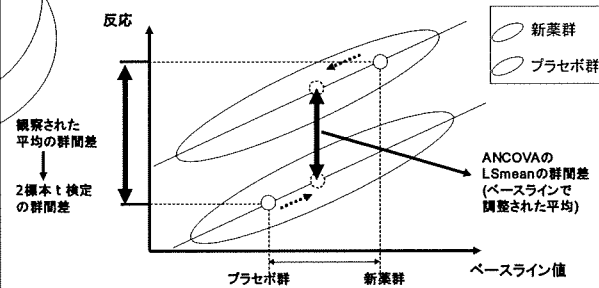


図: 反応とベースライン値の相関

23 / 41

共分散分析の結果

表: 鼻症状点数の治療期2週間後のベースラインからの変化量 (解析対象集団: FAS)

投与群	ベースライン		変化量(2週間)		投与群間の差 (95% CI)	P値*
	N	Mean	LSmean	SE		
プラセボ	314	1.53	-0.37	0.02	-	-
新薬 5mg	318	1.54	-0.47	0.02	-0.10 (-0.16, -0.04)	0.001
新薬 10mg	310	1.52	-0.47	0.02	-0.11 (-0.17, -0.05)	0.001

K. Okubo et al (2008) より抜粋

#: Comparison with Placebo using the ANCOVA model which contains treatment group, study site as factors, and baseline as a covariate

慶應大・医 CCR教育研修プログラム
2009

24 / 41

分散分析表とパラメータ推定値

$$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \text{誤差}$$

変動因	自由度	タイプIII平方和	平均平方	F値	Pr > F
Drug	2	4.5	2.2	8.3	<.0001
Baseline	1	10.5	10.5	10.5	<.0001

パラメータ	推定値	SE	t 値	Pr > t
Intercept	-0.43	2.47	-0.18	<.0001
Drug 5	-3.45	1.89	-1.83	<.0001
Drug 10	-3.34	1.85	-1.80	<.0001
Drug 0	0.00	.	.	.
Baseline	-0.99	0.16	6.00	<.0001

b₀ →
b₁ →
b₂ →
b₃ →

(注) 数値は、擬似的な例 25/41

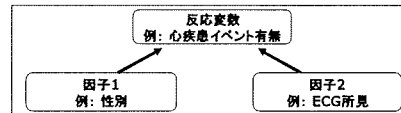
(2) 評価項目が2値の場合

○ 2値の評価項目

- 有効/無効、疾患イベント有無など

○ 統計手法

- χ^2 検定, Fisher の直接確率法 → デザインベース
- ロジスティック回帰分析 → モデルベース



慶應大・医 CCR教育研修プログラム 2009

26/41

データ: 心疾患イベントのリスク因子

表: 心疾患イベント発生例数の要約

性別	心電図異常	イベントあり	イベントなし	合計	オッズ
女性	なし	4	11	15	0.36
	あり	8	10	18	0.80
男性	なし	9	9	18	1.00
	あり	21	6	27	3.50

G. Koch et al (1985) から引用

研究目的: ECG所見とイベント発生の関連性を評価すること
→ ロジスティック回帰モデルを用いてオッズ比を推定

慶應大・医 CCR教育研修プログラム 2009

27/41

オッズ(比)とは?

$$\text{オッズ} = \frac{\text{イベントが起こる確率}}{\text{イベントが起こらない確率}}$$

性別	心電図異常	イベントあり	イベントなし	合計	オッズ	オッズ比
女性	なし	4	11	15	0.36	2.2
	あり	8	10	18	0.80	
男性	なし	9	9	18	1.00	3.5
	あり	21	6	27	3.50	

(例) オッズ比(ECG異常/ECG正常) = 2.2
→ ECG異常があると、イベントの起こるリスクが正常時の2.2倍

2009

28/41

ロジスティック回帰モデル

○ 対数オッズ(ロジット)に関するモデル

$$\log(\text{オッズ}) = b_0 + b_1x_1 + b_2x_2$$

イベントの起こり易さ(オッズ) ← 切片
性別 (x₁)
0: 女性
1: 男性
ECG所見 (x₂)
0: 正常
1: 異常

ロジスティック回帰モデル → 対数オッズをパラメータの足し算で表現 (一般化線形モデル)

29/41

回帰パラメータの解釈

$$\log(\text{オッズ}) = b_0 + b_1x_1 + b_2x_2$$

(例) b₁ = 1.35 の場合

$$\log(\text{男性のオッズ}) = b_0 + b_1 \times 1 + b_2$$

$$\log(\text{女性のオッズ}) = b_0 + b_1 \times 0 + b_2$$

$$\log(\text{男性のオッズ}) - \log(\text{女性のオッズ}) = b_1$$

$$\log(\text{オッズ比(男性/女性)}) = b_1$$

$$\text{オッズ比(男性/女性)} = \exp(b_1)$$

慶應大・医 CCR教育研修プログラム 2009

性別:
0: 女性
1: 男性

30/41

ロジスティック回帰分析の結果

(例) 冠動脈疾患イベントのリスク因子

表: パラメータ推定値

パラメータ	自由度	推定値	SE	Wald 統計量	P値
b_0 → 切片	1	-1.20	0.47	6.64	0.010
b_1 → 性別	1	1.15	0.50	5.25	0.022
b_2 → ECG異常	1	0.90	0.36	6.11	0.014

$$\exp(b_1) = \exp(1.15)$$

$$\exp(b_1 \pm 1.96 \times SE) = \exp(1.15 \pm 1.96 \times 0.50)$$

表: オッズ比の推定値

	オッズ比	95%信頼区間
性別	3.1	(1.2, 8.4)
ECG異常	2.5	(1.2, 5.0)

慶應大・医 CCR教育研修プログラム
2009

31 / 41

ロジスティック回帰分析の出力2

○ モデルの当てはまりの確認

表: 当てはまりの良さ(Goodness-of-Fit)の検定

	Value	DF	Value/DF	P値
Deviance	1.60	3	0.53	0.660
Pearson	1.60	3	0.53	0.660

帰無仮説: モデルの当てはまりは良い

対立仮説: モデルの当てはまりは良くない

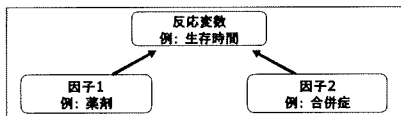
帰無仮説は棄却されず → モデルの当てはまりに問題はない

慶應大・医 CCR教育研修プログラム
2009

32 / 41

(3) 評価項目が生存時間の場合

- 生存時間
 - 死亡までの時間、発作の起こるまでの時間など
- 統計手法
 - ログランク検定 → デザインベース(ノンパラ)
 - Cox回帰モデル → モデルベース(セミパラ)



33 / 41

例題3: 骨髄腫患者を対象とした無作為化比較試験

- 研究目的
 - 生存時間を評価指標として、薬剤Aと薬剤Bを比較する

生存時間データの例

OBS	時間(日)	打ち切り	投与群	合併症
1	8	1	0	1
2	180	1	1	0
3	632	1	1	0
4	852	0	0	0
5	52	1	0	1

ここでは、0を打ち切りとする

慶應大・医 CCR教育研修プログラム
2009

0: 薬剤A
1: 薬剤B

0: なし
1: あり

34 / 41

生存時間解析でよく見る図 ～ 生存関数のグラフ ～

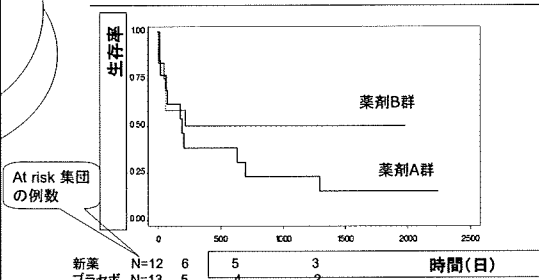


図: 骨髄腫患者における生存関数のKaplan-Meier プロット (Peto et al (1977) から引用)

35 / 41

生存時間解析で使う統計モデル

- Cox 回帰モデル (D.Cox, 1972)

$$\log(\text{ハザード}) = \lambda_0(t) + b_1x_1 + b_2x_2$$

イベントリスクの大きさ

ベースラインハザード

投与群 (x_1)

0: 薬剤A
1: 薬剤B

合併症 (x_2)

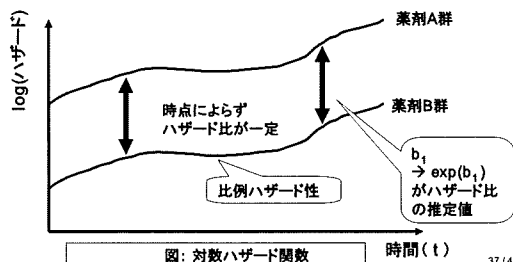
0: なし
1: あり

Cox回帰モデル → 対数ハザードをパラメータの足し算で表現 (一般化線形モデル)

36 / 41

比例ハザードモデルとは・・・

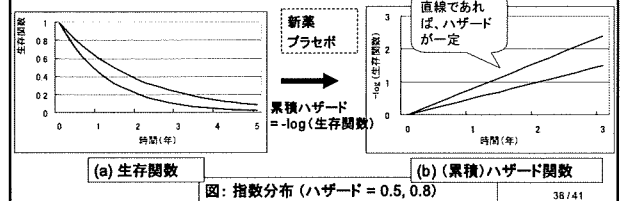
→ Cox 回帰モデルの別名



37 / 41

生存関数とハザード関数の関係

- 生存関数 : 時点 t の生存確率
- ハザード関数: 時点 t の生存集団での生存率の変化率 → 瞬間死亡率 (単位: /時間)



38 / 41

Cox 回帰の結果

○ 骨髄腫の臨床試験

イベントと打ち切り数が正しいことを確認!

表: イベント数と打ち切り数の要約

全体	イベント	打ち切り	打ち切りの割合(%)
25	17	8	32.0

表: パラメータ推定値 (最尤法)

パラメータ	自由度	推定値	SE	カイ2乗統計量	P値	ハザード比
b_1 → 投与群	1	-1.22	0.60	4.18	0.041	0.30
b_2 → 腎機能	1	4.28	1.19	12.90	<0.001	72.09

慶應大・医 CCR教育研修プログラム
2009

(注) タイは正確法により調整

39 / 41

まとめ

- 統計の基本知識 (推定と検定)
- 共分散分析の結果の解釈
 - Lsmean (調整平均) の意味
- ロジスティック回帰分析の結果の解釈
 - パラメータ推定値とオッズ比の関係
- Cox回帰の結果の解釈
 - 比例ハザードモデルの意味
 - パラメータ推定値とハザード比の関係

慶應大・医 CCR教育研修プログラム
2009

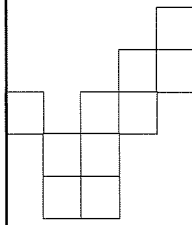
40 / 41

参考文献 (統計学の入門書)

- 岩崎学 (2000) 統計的データ解析のレシピ. 日本評論社
- 大橋靖雄, 浜田知久馬 (1995) 生存時間解析. 東京大学出版会
- 折笠秀樹 (1995) 臨床研究デザイン. 真興交易医書出版部
- 竹内正弘 (2000) 生物統計学 入門. 丸善
- 丹後俊郎 (1993) 新版 医学への統計学. 朝倉書店
- 浜田知久馬 (1999) 学会・論文発表のための統計学. 真興交易医書出版部
- 吉村功 (1987) 毒性・薬効データの統計解析. サイエントリスト社

慶應大・医 CCR教育研修プログラム
2009

41 / 41



慶應義塾大学医学部 臨床研究講習会

臨床研究・治験の実際

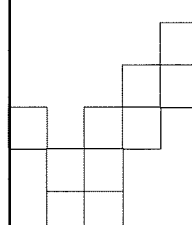
クリニカルリサーチセンター
佐藤裕史

2009年12月19日

いくつかの疑問

- 日本の基礎医学の国際水準の業績と、臨床の業績に差があるのはなぜか？
In 2002, Japan contributed to around 8% of papers in Nature but was responsible for less than 4% of contributions to The Lancet
(Lancet 369:1333, 2007)
- なぜ海外で使えて日本で使えない医薬品が多いのか？
- 時代遅れの医薬品や日本にしかない薬がどうしてあるのか？
- なぜ臨床試験の多くは日本の外(欧米・アジア諸国)で行なわれているのか？
- 全世界の製薬会社の上位10社になぜ日本企業が入っていないのか？

- 臨床研究全体の動向
- 臨床研究・治験の異同
- 臨床研究・治験の国際化と日本の現状



臨床研究全体の動向

主要国際誌における基礎・臨床論文数の国別割合

TABLE 1. SUMMARIES OF TOP-RANKING 20 COUNTRIES FOR BASIC SCIENCE ARTICLES*

Country	1991 (N=1781)	2000 (N=3200)	1991-2000 percent
United States	59.7	55.3	66.4
United Kingdom	8.6	9.9	8.4
Germany	5.4	4.9	14.6
Japan	7.6	3.7	3.3
France	7.3	7.2	3.7
Canada	2.4	2.2	2.7
Switzerland	2.1	1.9	1.9
Netherlands	0.9	1.5	1.1
Australia	0.7	1.5	0.0
Sweden	0.8	0.9	0.8
Italy	0.4	1.1	0.8
Israel	0.8	0.9	0.8
Austria	0.6	0.8	0.4
Nigeria	0.3	0.4	0.4
Spain	0.3	0.4	0.3
Finland	0.1	0.1	0.1
Denmark	0.1	0.1	0.1
China	0.1	0.1	0.1
Norway	0.1	0.1	0.1
New Zealand	0.1	0.2	0.1

TABLE 2. SUMMARIES OF TOP-RANKING 20 COUNTRIES FOR CLINICAL ARTICLES*

Country	1991 (N=2857)	2000 (N=2898)	1991-2000 percent
United States	60.0	52.1	53.7
United Kingdom	16.2	17.3	18.3
Canada	3.1	3.3	7.8
France	1.1	1.1	1.4
Italy	1.0	1.4	1.2
Australia	0.9	1.1	1.1
Germany	0.8	1.1	0.9
Japan	0.5	0.9	0.8
Switzerland	0.6	1.0	0.8
Denmark	0.5	0.7	0.7
Sweden	0.6	0.6	0.7
Ireland	0.4	0.8	0.6
Japan	0.5	0.6	0.6
Spain	0.5	0.3	0.5
New Zealand	0.2	0.8	0.5
Israel	0.1	0.4	0.4
Belgium	0.5	0.2	0.2
Norway	0.4	0.2	0.3
South Africa	0.3	0.3	0.3

*Rankings based on average values for 1991 to 2000. Data for New Zealand and Norway are based on articles published in 1992, 2000 and 2005 to 2000, respectively. Data for Israel published in 1991 and 1995. All articles in Nature and Lancet have been included. Data for all countries related to basic science. Percentages do not sum to 100 because many of other countries are not included. Data for China include Hong Kong.

New E J Med 347:1211-1212, 2002

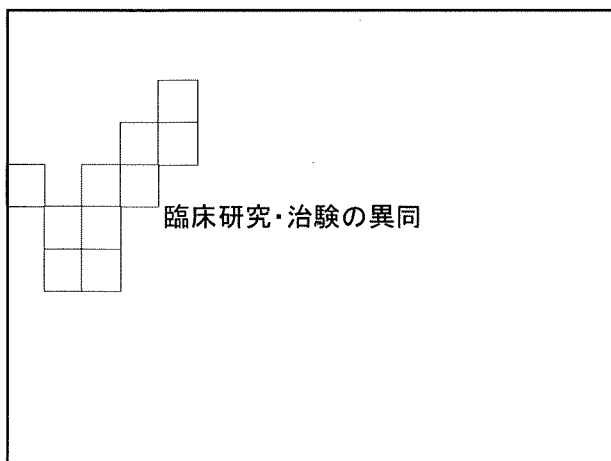
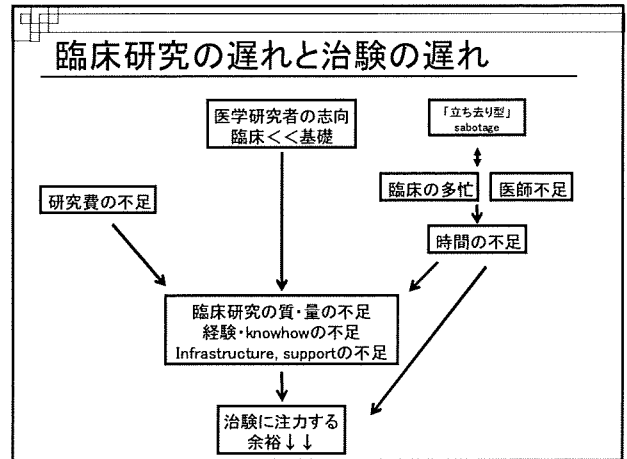
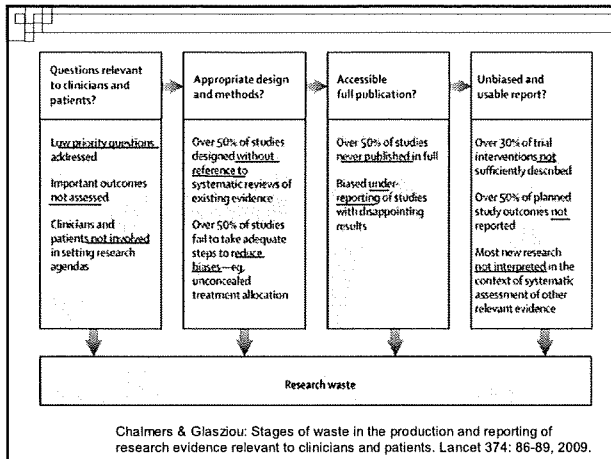
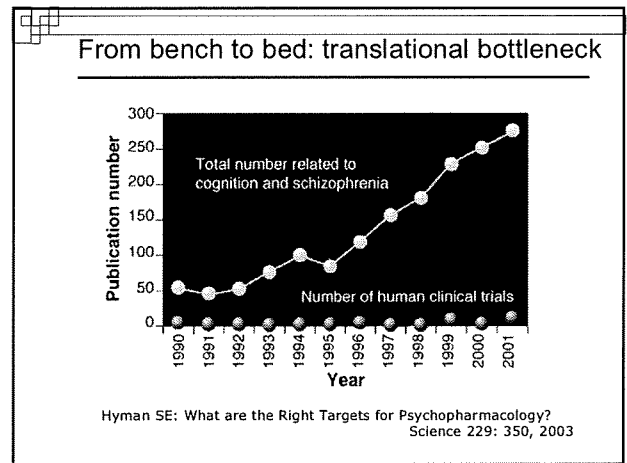
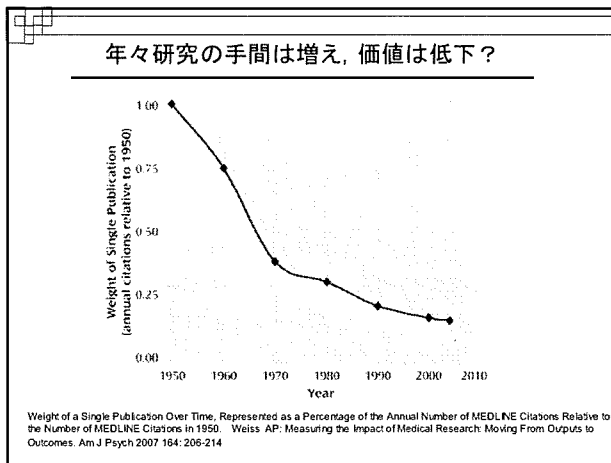
Basic articles (Nature Med, Cell, J Exp Med)

Rank	1993-1997		1998-2002		2003-2007	
	USA	N of articles	USA	N of articles	USA	N of articles
1	USA	5097	USA	2769	USA	2674
2	England	365	Germany	404	Germany	442
3	Germany	221	Japan	371	Japan	369
4	Switzerland	244	England	352	England	314
5	France	239	France	256	France	269
6	Japan	236	Canada	209	Canada	204
7	Canada	227	Switzerland	209	Switzerland	166
8	Italy	122	Italy	132	Italy	155
9	The Netherlands	109	The Netherlands	114	The Netherlands	127
10	Australia	97	Australia	106	Australia	120

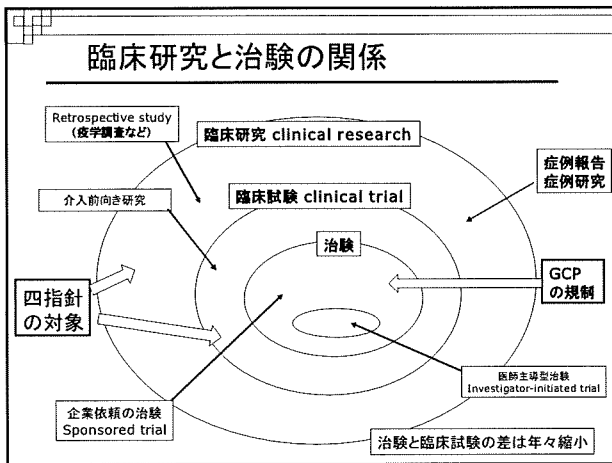
Clinical Articles (N Eng J Med, Lancet, JAMA)

Rank	1993-1997		1998-2002		2003-2007	
	USA	N of articles	USA	N of articles	USA	N of articles
1	USA	3314	USA	2955	USA	2677
2	England	590	England	644	England	673
3	Canada	377	Canada	511	Canada	462
4	The Netherlands	277	The Netherlands	502	Germany	343
5	France	274	France	432	France	300
6	Germany	253	Germany	410	The Netherlands	294
7	Italy	236	Italy	374	Italy	279
8	Switzerland	166	Switzerland	282	Australia	260
9	Australia	155	Australia	261	Switzerland	252
10	Sweden	155	Sweden	224	Belgium	177
11	Scotland	151	Scotland	216	Sweden	166
12	Japan	122	Japan	183	Scotland	145
13	Belgium	110	Denmark	152	Spain	141
14	Denmark	98	Belgium	152	Denmark	135
15	Spain	90	Spain	126	China	102
16	Finland	88	Finland	121	Norway	86
17	Israel	53	Austria	88	Finland	79
18	Austria	50	Norway	83	Japan	74
19	Wales	46	Wales	75	Israel	67
20	Norway	46	New Zealand	46	New Zealand	45

Takayama, 2008



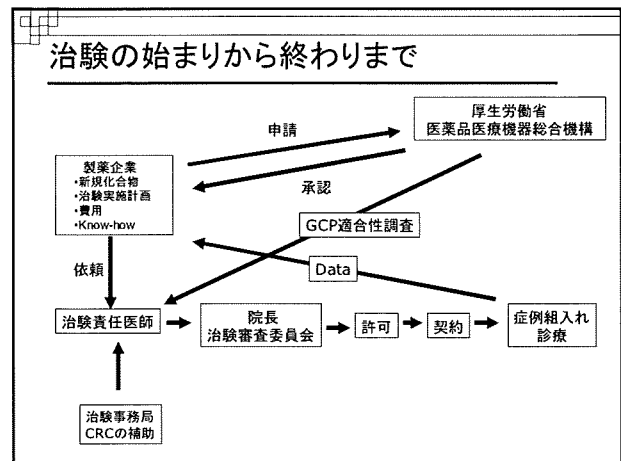
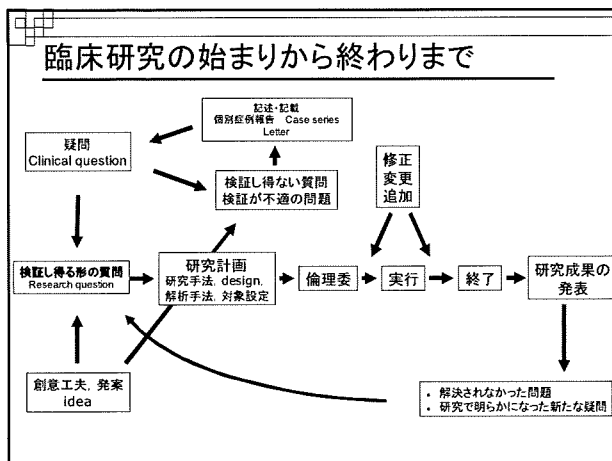
- ### 臨床研究 と治験の区別 (日本)
- 臨床研究 clinical research: ヒトを対象とする研究の総称
 - 治験: sponsored clinical trial
 - 新たな治療手段(薬剤, 医療機器)の有効性・安全性を実証するために行い, 結果を規制当局に提出し, 製造販売承認を受けて, 保険適用下で使用するための臨床試験
 - 日常診療や一般臨床研究と異なる水準の実施(異なる法的規制(GCP: Good Clinical Practice))が要求される



臨床研究と関連する規範

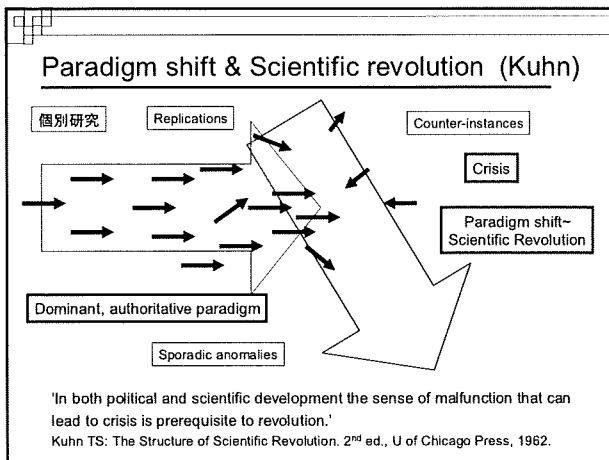
	法	行政指針
治験 (企業主体) (医師主導型)	<ul style="list-style-type: none"> ●薬事法(14, 80, 82条) ●薬事法施行令 ●医薬品の臨床試験の実施の基準に関する省令(平成9年厚生省令28号) 	
臨床研究	なし	四指針 <ul style="list-style-type: none"> ●臨床研究に関する倫理指針 ●疫学研究指針 ●ヒトゲノム指針 ●ES細胞指針

米国: 臨床研究(治験を含む)は全てnational research act, clinical research actの下に



- ### 臨床研究全体に共通する困難
- 方法論: 教育, 研修, 指導, 補助が不十分
 - 意識づけ: 基礎研究と比較した場合の, 臨床研究の科学性に対する懐疑
 - 動機づけ: incentiveの少なさ
 - Financial
 - Promotional
 - Academic
 - Vocational
 - 日常臨床の多忙
 - 設備, 施設, 資金の上の困難

- ### いかに発表するか: 日本の盲点
- 学会での口頭発表+討論
 - 論文投稿+査読者との交渉~掲載~批判letterとの応酬
 - 何れも日本人研究者は他国(他アジア諸国を含む)より不得手
 - 理由
 - 英語の問題: 主ではない(日本語でもうまくない)→発表技術一般の未熟
 - 態度の問題: 発表技術の軽視
 - 訓練の問題: 高等教育で一切なし
 - 対策
 - 泥縄式訓練, 補習
 - Medical writing ~ supervision, editorial assistance
 - Presentation skill training
 - How to survive debate & how to chair a meeting



- ### Paradigm shiftを扱うのは良い総説の責務である
- Review article自体も大事なoriginal contributionである
 - 単なる概説、要約ではない
 - 個別研究の流れ、方向、anomalies, counter-instancesを俯瞰するのがreview articleの役目
 - Ideal review should include a verbal and statistical evaluation of ideas, methods, and results in primary research...Reviews should treat individual papers as 'subjects'.
 - Methodological review
 - Conceptual review
 - > To evaluate the quality and origin of the information on which medical knowledge is based
 - Integrative review
 - > Metaanalysis
- (Berrios GE: The review paper in psychiatry. In: Freeman C, Tyrer P(eds): Research Methods in Psychiatry, 2nd ed., Gaskell,1992)

Hierarchies of evidence?

(Rawlins, Lancet 372, 2152-61, 2008)

Evidence, in the present context, has only one purpose: it forms the basis for informing decision makers about the appropriate use of therapeutic interventions in routine clinical practice...

The notion that evidence can be reliably placed in hierarchies is illusory. Hierarchies place RCTs on an uncomfortable pedestal for, although the technique has advantages it also has disadvantages. Observational studies have defects but they also have merit. judgments are an essential ingredient of most aspects of the decision-making process.

Criteria
1++ High quality meta-analyses, systematic reviews of RCTs, or RCTs with a very low risk of bias
1+ Well conducted meta-analyses, systematic reviews of RCTs, or RCTs with a low risk of bias
1- Meta-analyses, systematic reviews of RCTs, or RCTs with a high risk of bias
2++ High quality systematic reviews of case-control studies or cohort studies or high quality case-control or cohort studies with a very low risk of confounding, bias, or chance
2+ Well conducted case-control or cohort studies with a low risk of confounding, bias, or chance
2- Case-control or cohort studies with a high risk of confounding, bias, or chance
3 Non-analytic studies (eg. case reports, case studies)
4 Expert opinion

Levels of evidence

- ### Important limitations of RCTs
- (Rawlins, Lancet 372, 2152-61, 2008)
- The null hypothesis
 - 'No difference' hypothesis
 - Inappropriate when benefit previously shown
 - 73% of RCTs fail to refer to past evidence of benefits of the Tx in question
 - Probability
 - Difficulties in interpreting p values
 - Stopping rules of trials ~ serious pitfalls; no consensus
 - Subgroup analysis
 - The problem of multiplicity
 - Generalisability
 - Explanatory trials: to estimate the biological effect of a Tx
 - Pragmatic trials: to estimate the effectiveness of a Tx
 - Resources

- ### EBMに関する根本的誤解
- 日本で横行している見解 (evidenceに患者をあてはめる)
 - 「診療ガイドラインに則った医療をすることがEBMである」
 - 「EBMに基づく標準的診療」「EBMのガイドラインに当てはまらない患者はどうしたらいいのか」という本末顛倒の誤解に
 - (李路亮. 市場原理が医療を亡ぼす—アメリカの失敗. 医学書院. 2004)
 - EBMの原義 (患者にevidenceを適用できるどうか)
 - 「眼前のこの患者にこの知見はあてはまるのか」という視点から、文献などのevidence(の適用可能性)を吟味すること
 - 「この患者にはこのevidenceは不適として棄却することもあり得る」

EBM is not “cookbook” medicine. External clinical evidence can inform, but can never replace, individual clinical expertise, and it is this expertise that decides whether the external evidence applies to the individual patient at all and, if so, how it should be integrated into a clinical decision.

Sackett DL et al.: Evidence based medicine: what it is and what it is not. BMJ 312: 71-72, 1996.