

ACTIVITIES PROMOTING THE PEDIATRIC CLINICAL TRIALS AND THE DRUG DEVELOPMENT
IN UNITED STATES AND EUROPEAN UNION

Nao Tsuchida

*National Center for Child Health and Development
Department of General Pediatrics and Interdisciplinary Medicine*

Abstract

Off-label use in children is common throughout the world and a key issue. Activities resolving these issues have been stepped forward in the U.S. and EU since the 1990s. It has been tried various expedients to achieve the purpose. Because of need for facilitating the development of pediatric trials of therapies for HIV, eventually the following laws were legislated. These are referred to children in whole or in part.

U.S.: FDA Modernization Act (1997)

Best Pharmaceuticals for Children Act (2002)

Pediatric Research Equity Act (2003)

FDA Amendments Act of 2007

EU: Paediatric Regulation (2007)

These approaches would promote the pediatric clinical trials and the drug development in the U.S. and EU. We have to work together cooperatively with the U.S. and EU.

Key words : Pediatric Research, Off-Label Use, Therapeutic Orphan, Drug Development, Clinical Trials

成育医療臨床研究基盤整備勉強会

Chapter 9 : Enhancing Causal Interference in Observational Studies

Designing Clinical Research 4th Ed.

16th June, 2008
National Center for Child Health and Development
Clinical Research Center
Takanobu Maekawa

Contents:

- 関連の種類
- 偽の関連を除外する方法
 - 偶然
 - バイアス
- 真の関連から、さらに因果関係を導く
 - 交絡の予防

関連 Associations

- 観察研究 (Chapter 8) の結果、AとBに関連が疑われる
- 知りたいこと
- AとBは本当に関連があるのか (関連の真偽)
- AはBの原因なのか (Cause-Effect; 因果関係)

ある関連が疑われた時 5つの可能性を考える

偶然 Chance (random error)	偽の関連	関連が疑われるが、偶然であり、実はコーヒーと心筋梗塞は無関係である	$A \not\leftrightarrow B$
バイアス Bias (systematic error)	偽の関連	関連が疑われるが、測定方法に問題があった可能性があり、実はコーヒーと心筋梗塞は無関係である	$A \not\leftrightarrow B$
結果-原因 Effect-cause	真の関連	実は逆で、心筋梗塞がコーヒー摂取の原因である	$A \leftarrow B$
交絡 Confounding	真の関連	コーヒー摂取は、心筋梗塞の原因である第3の外的要因と関連している	Factor X $A \leftrightarrow B$
因果関係 Cause-effect	真の関連	コーヒー摂取は、心筋梗塞の原因となる	$A \rightarrow B$

(TABLE 9.1, p128) A: コーヒー摂取 B: 心筋梗塞

ある関連が疑われた時 5つの可能性を考える

偶然 Chance (random error)	偽の関連	関連が疑われるが、偶然であり、実はコーヒーと心筋梗塞は無関係である	$A \not\leftrightarrow B$
バイアス Bias (systematic error)	偽の関連	関連が疑われるが、測定方法に問題があった可能性があり、実はコーヒーと心筋梗塞は無関係である	$A \not\leftrightarrow B$
結果-原因 Effect-cause	真の関連	実は逆で、心筋梗塞がコーヒー摂取の原因である	$A \leftarrow B$
交絡 Confounding	真の関連	コーヒー摂取は、心筋梗塞の原因である第3の外的要因と関連している	Factor X $A \leftrightarrow B$
因果関係 Cause-effect	真の関連	コーヒー摂取は、心筋梗塞の原因となる	$A \rightarrow B$

(TABLE 9.1, p128) A: コーヒー摂取 B: 心筋梗塞

「偶然」による偽の関連

	Control群	60	40	
	MI群 ₁	12	8	
	MI群 ₂	19	1	
				$P=0.6$ $P=0.0054$

コーヒー摂取習慣あり(60%)
習慣なし(40%)

コーヒー摂取と心筋梗塞に関連がない場合を仮定すると...

- 心筋梗塞群におけるコーヒー摂取率は全体集団と同一(60%)であることが期待される(MI群₁)
- だが、偶然この期待通りにサンプリングされないこともある(MI群₂)
- 偶然により生じる期待値とのズレを偶然誤差(Random error)という。
- 偶然誤差が統計的に有意である場合、Type 1 error (α-error)という。

「バイアス」による偽の関連

- Research Question
 - 研究者が知りたいと思う問題
 - 理想
- Study Plan
 - 実施可能性を考えて妥協して作った研究計画
 - 作り出した現実
- バイアス
 - 理想と現実のズレ
 - 研究者の恣意や問題のある測定方法、予想しない原因により、特定の方向に結果が歪められる。

Research Question & Study plan

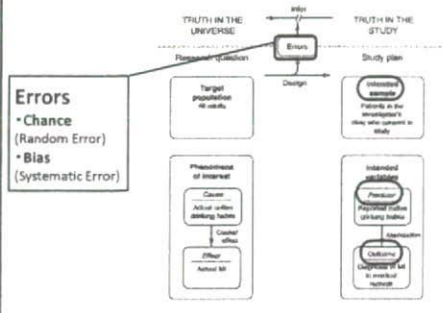


FIGURE 9.1. Abstracting bias by comparing the research question and the study plan.

バイアスの原因

- 主な3つ
 - 対象集団の代表性
 - 測定する予測因子は真に興味あるリスクを捉えているか
 - 測定する結果因子は真に興味ある状態を捉えているか
- コーヒーの例では
 - 対象が心筋梗塞以外の入院患者
 - コーヒー摂取量が一般成人より少ないかも知れない
 - 他の病因による心筋梗塞のリスク
 - 測定方法
 - 時期、回数、尺度、...

偽の関連の除去には

- デザイン段階での対処が重要
- 偶然による場合
 - サンプルサイズを大きくする
 - 測定精度を高める (Chapter4, Table 4.2)
- バイアスによる場合
 - 3点(対象, 予測因子, 説明因子)を検討する
 - 他の研究との比較
 - 感度解析(解析の段階での工夫)

ある関連が疑われた時 5つの可能性を考える

偶然 Error(random error)	偽の関連	関連が疑われるが、偶然であり、実はコーヒーと心筋梗塞は無関係である	A \nrightarrow B
バイアス Bias(systemic error)	偽の関連	関連が疑われるが、測定方法に問題があった可能性があり、実はコーヒーと心筋梗塞は無関係である	A \nrightarrow B
結果-原因 Effect-cause	真の関連	実は逆で、心筋梗塞がコーヒー摂取の原因である	A \leftarrow B
交絡 Cofounding	真の関連	コーヒー摂取は、心筋梗塞の原因である第3の外的要因と関連している	Factor X A \rightarrow B
因果関係 Cause-effect	真の関連	コーヒー摂取は、心筋梗塞の原因となる	A \rightarrow B

TABLE 9.1, p128

A: コーヒー摂取 B: 心筋梗塞

真の関連

- 因果関係を示唆するエビデンス
- 関連の普遍性 (consistency)
 - 関連の強さ (strength)
 - 量-反応関係 (dose-response)
 - 生物学的妥当性 (biologic plausibility)
 - 時間的妥当性 (temporal)
- 注意点)
- 交絡因子と結果の関係もまた、一貫性や量反応関係を示す可能性がある。
 - 一般的には、交絡因子によって生じた関連は、真の因果関係より関連が弱いのが普通である。

結果-原因

原因因子と結果因子が逆

- CRP高値と心筋梗塞
- 携帯電話の使用と交通事故

- 原因と結果の時間的前後関係を考慮する。
- 時間的前後関係が比較的明確なコホート研究では問題となることはすくない。

注意)

- 結果が生じる(発病)までに潜伏期間の長い原因の場合、コホート調査でも前後関係の解釈が難しい場合がある。例えば2型糖尿病と膵臓癌発症の関連。

交絡(Confounding)

- 交絡因子(Confounding variable)

A confounding variable is one that is associated with the predictor variable, and a cause of outcome variable.

- 交絡がおきている

- 交絡因子によって関係がゆがめられている状態



例えば

心筋梗塞のリスクを調べる観察研究

- コーヒー習慣と心筋梗塞発症に関連が疑われた
コーヒーは真の原因か？
- 真の原因がないか？
- コーヒーは真の原因と関連のある事象にすぎない
無実の傍観者(innocent bystander)ではないか？



例えば

心筋梗塞のリスクを調べる観察研究

- コーヒー習慣と心筋梗塞発症に関連が疑われた
コーヒーは真の原因か？
- 喫煙は心筋梗塞のリスクである
- コーヒーと喫煙には高い相関がある
- 喫煙が交絡因子、コーヒーは無実の傍観者

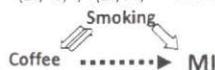


例題(Appendix 9A)

	喫煙 あり		喫煙 なし	
	MIあり	MIなし	MIあり	MIなし
コーヒー習慣 あり	80	40	10	20
コーヒー習慣 なし	20	10	40	80
	100	50	50	100

MIに対するコーヒー摂取のオッズ比

- 喫煙者 = $(80/20) / (40/10) = 80 \times 10 / 20 \times 40 = 1$
- 非喫煙者 = $(10/40) / (20/80) = 10 \times 80 / 40 \times 20 = 1$



例題(Appendix 9A)

	喫煙 あり+喫煙 なし	
	MIあり	MIなし
コーヒー習慣 あり	A	B
コーヒー習慣 なし	C	D

MIに対するコーヒー摂取のオッズ比

- 喫煙者+非喫煙者 = $(A/B) / (C/D)$

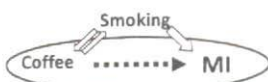


例題(Appendix 9A)

	喫煙 あり+喫煙 なし	
	MIあり	MIなし
コーヒー習慣 あり	90	60
コーヒー習慣 なし	60	90
	150	150

MIに対するコーヒー摂取のオッズ比

• 喫煙者+非喫煙者 = $(90/60) / (60/90) = 90 \times 90 / 60 \times 60 = 2.25$



例題(Appendix 9A)

	喫煙 あり		喫煙 なし	
	MIあり	MIなし	MIあり	MIなし
コーヒー習慣 あり				
コーヒー習慣 なし				
	A	B	C	D

MIに対する喫煙のオッズ比

• $(A/B) / (C/D)$



	MI(+)	MI(-)
喫煙(+)	A	B
喫煙(-)	C	D

例題(Appendix 9A)

	喫煙 あり		喫煙 なし	
	MIあり	MIなし	MIあり	MIなし
コーヒー習慣 あり	80	40	10	20
コーヒー習慣 なし	20	10	40	80
	100	50	50	100

MIに対する喫煙のオッズ比

• $(100/50) / (50/100) = 100 \times 100 / 50 \times 50 = 4$



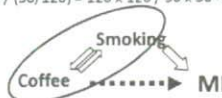
	MI(+)	MI(-)
喫煙(+)	100	50
喫煙(-)	50	100

例題(Appendix 9A)

	喫煙 あり		喫煙 なし	
	MIあり	MIなし	MIあり	MIなし
コーヒー習慣 あり	120		30	
コーヒー習慣 なし	30		120	
	150		150	

コーヒー習慣の喫煙のオッズ比

• $(120/30) / (30/120) = 120 \times 120 / 30 \times 30 = 16$



	喫煙(+)	喫煙(-)
Coffee(+)	120	30
Coffee(-)	30	120

交絡への対処

- 研究計画(デザイン)
 - 対象者の限定
 - マッチング
- 解析段階
 - 層別解析
 - 回帰分析

デザインによる対処

- 対象者の限定
 - 均一な背景の集団にだけ焦点をしぼる
 - わかりやすい
 - 研究テーマに的が絞られる
- 一般性は損なわれる
- 場合によっては十分なサンプルの確保が困難

デザインによる対処

- マッチング
 - ある因子の分布を比較群間でそろえること
予測因子 + 関連が予測される因子
- ペアマッチング(pairwise matching)
 - コーヒー + 非喫煙 → 非喫煙 Control
 - コーヒー + 喫煙10~20本 → 喫煙10~20本 Control
- 頻度マッチング(frequency matching)
 - (喫煙10~20本) 群20人 → 喫煙10~20本 Control 40人

特徴

- マッチング因子による交絡が制御される
- コントロールが選択しやすい
- 未測定因子の影響も制御できる
 - 例: 兄弟マッチングで遺伝要因・環境要因を制御
- マッチング因子の影響を調べられない
- 特殊な解析が必要
- オーバーマッチングによる検出力の低下
 - 群間で、マッチング因子の分布を揃えたために予測因子の分布も似てしまうこと
 - バイアスにはならない

解析による対処

- デザイン段階での制御が重要
 - たくさんの因子での限定やマッチングは不可能
 - 研究前にわからないこともある
 - 交絡の強さの程度がわからない
- 主な二つ
 - 層別解析
 - 回帰分析

層別解析

- 方法
 - 対象者を属性(層別因子)で分類
 - 層ごとに解析
 - 結果を併合する
 - 人数に応じて重み付けすることが多い
- サブグループ解析
 - 一部の対象者についてのみ解析・結果の提示

層別解析の特徴

- わかりやすい
- 層の作り方は任意
- 層別因子はカテゴリ変数として扱う
- 層別因子が多いと層内に対象者が少なくなり、結果が不安定になる
 - 人種・性別・年齢・体格・喫煙・居住環境...
- 測定してある因子しか使えない

対処方法の選択

- 「対象者の限定」
 - 興味ある集団が決まっていたら
- 「マッチング」
 - 年齢・人種・性別など他の因子に影響をうけない属性因子について
- 「層別解析」
 - 予定しているときはサンプルは多めに
- 「回帰分析」
 - 補足的に
 - 色々やってみる

まとめ

- ある事象AとBに関連が疑われた場合、5つの可能性を考える
- 一番知りたいのは因果関係である
- そのために「偶然」「バイアス」「交絡」を除去
 - 偶然: α エラー、精度とサンプルサイズ
 - バイアス: 研究計画の吟味
 - 交絡: 対象選択と評価方法を吟味する

まとめ 真の関連

因果関係を示唆するエビデンス

- 関連の普遍性 (consistency)
- 関連の強さ (strength)
- 量-反応関係 (dose-response)
- 生物学的妥当性 (biologic plausibility)
- 時間的妥当性 (temporal)

注意点)

- 交絡因子と結果の関係もまた、一貫性や量反応関係を示す可能性がある。
- 一般的には、交絡因子によって生じた関連は、真の因果関係より関連が弱いのが普通である。

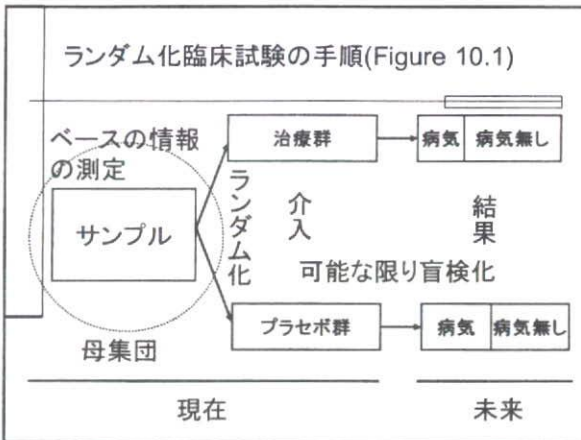
2008年7月7日 臨床研究センター勉強会
 医学研究のデザイン 第10章
 実験的研究をデザインする

国立成育医療センター
 内分泌代謝科 臨床研究フェロー
 磯島 豪

実験的研究では

- 研究者は対象者に介入(治療)を行い、その結果を観察する
- 因果関係の推論が可能である
- 特に、ランダム化と盲検化が大切である
- 費用と時間がかかる
- 一度の研究で証明できる仮説が少ない
- 対象者に、重篤な副作用が生じることがある
- 研究テーマによっては、適切で無いものもある

研究テーマに関する知見が十分成熟した段階になって初めて行う



本日の内容

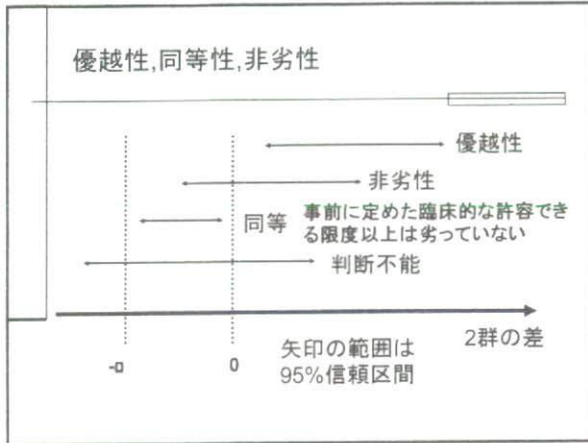
- 介入とコントロール群の選択
- 測定する結果の選択
- 対象者の選択
- ベースラインの変数の測定
- ランダム化と盲検化

介入とコントロール群の選択(1)

- 介入の選択
 - 介入の選択することが、臨床試験計画の第一段階である
 - 効果と安全のバランスは、研究の状況に依存する
介入を考える時、重症の病気に対しては、効果の方が、軽症の病気に対しては、安全の方が大きな問題となる
 - 一つの介入の方が、複雑な介入よりも結論が明確になる。しかし、医療の現場では単純化が難しい場合もある
 - 単純な介入の方が、実施可能性が高まる

介入とコントロール群の選択(2)

- コントロール群の選択
 - 介入を全く受けず、盲検化されているのが最高のコントロール群である。介入と区別出来ないプラセボを用いる事がそれに相当する。
 - 試験治療の結果のリスクを下げる共介入の存在について注意する必要がある。
 - 標準治療と新しい治療を比較する際には、同等性試験が行われる
 - 対象者がランダム化されることが難しいと予想される有望な外科治療などでは、wait-list controlが使用されることもある。これは、介入の結果がすぐに出る場合に限られる



- 測定する結果の選択
- 測定する結果の選択は,試験デザイン,費用など多くの事に影響する
 - 臨床試験には,種々の結果測定が含まれるが,主要評価項目を1つ決めるべきである.主要評価項目を決めることでサンプルサイズが計算できる
 - 試験の結果に影響する代替指標を用いることも有用である.代替指標には,疑問となった結果を良く表しているものもある(例:HIVウイルス量),そうでないものもある(例:骨密度)
 - 介入による予想される副作用は,必ず測定されなければならない.

- 対象者の選択 (第3章参照)
- 選択基準の設定
 - inclusion criteria (取り込み基準)
研究にふさわしく,かつ統計学的に有意な介入効果を得やすい集団を定義する
 - exclusion criteria (除外基準)
むやみに設けすぎない.設定される場合,理由は主に5つに分類される.
1) 試験治療が有害, 2) 試験治療が効果がない可能性が高い, 3) アドヒアランスが悪い可能性が高い, 4) 経過観察が困難, 5) 研究への参加に現実的な問題がある

- ベースラインの変数の測定 (1)
- 対象者の特性を記述する
- 研究結果の一般化可能性を他の人が判断できる材料の測定
 - 予測因子やサブグループ化に役立つ因子を測定する
- 主要評価項目とは別の予測因子と結果の関連を研究できる
- 群間で予測因子の分布が偶然異なった場合に,統計学的に補正して比較できる
- サブグループにより,介入の効果が検討できる

- ベースラインの変数の測定 (2)
- 結果変数のベースラインを測定する
- 2値変数の場合は,研究開始時点での状態と研究後の結果を比較,連続変数の場合は,研究前後での変化を比較することが多い
 - 無駄な測定はしない
- 測定を最小限にして大規模な臨床試験を行う
- ランダム化により交絡因子が除去されるはず
 - 検体バンクを確立する
- 結果に影響する生物学的予測因子が後から予測できる
- 後から別のテーマでの研究ができる

ランダム化によるベースライン因子の交絡の除去と盲検化による共介入の交絡の除去(表10.2)

関連の生じた原因	結論の確実性を高めるための方策
1.偶然	観察研究の場合と同じ
2.バイアス	
3.結果 - 原因関係	実験的研究ではなし
4.交絡	ランダム化前の交絡因子 → ランダム化 ランダム化後の交絡因子 → 盲検化
5.因果関係	

ランダム化とは

- 予見不可能に群を決定すること
 - 選択バイアスを防ぐ
- 群間の比較可能性を確率的に保証する
 - 未知の因子についても保証する
 - 偶然偏ることもある
- 第三者が割り付けるのが理想
- ランダム化の遵守

ランダム化の前に

- ベースラインデータの測定
- 選択・除外基準を満たすことの確認
- インフォームドコンセントの確認

小規模研究におけるランダム化

- コイントス
- 乱数の利用
- 予見不可能性の保証が大切

大規模研究におけるランダム化

- 第三者機関で割付
- 研究者の意図が入らないことが大切

ランダム化の方法

- 割付比
- ランダム化法
 - 通常のランダム化
単純ランダム化, ブロックランダム化, 層別ランダム化
 - 動的割付(割付結果を見ながら次の人を割り付ける)
 - 最小化法, つぼモデル

単純ランダム化

- 割り付ける群をA群とB群とすると
 - ランダムにAとBを発生させる
 - 数が少ないとAとBの数が違う可能性がある
 - 性別, 年齢といった治療に影響する因子が片方に偏る可能性がある

ブロックランダム化

- 一定の人数ごとにブロックを作りその中で割り付ける方法
- 割り付ける群をA群,B群として,1ブロックの数を4とすると,
 - A,Bの並べ方は,6通り(ABAB, ABBA, AABB, BAAB, BABA, BBAA)
 - 4人のブロックに,どの順列が割り付けられるかはランダムに決定される
 - 4人ごとに各群の人数が均等になる
 - エントリー時期による対象者の質の変動に対応

層別ランダム化

- 均等にしたい因子で層を作る
- 一般的に層別ブロックランダム化が多い
- 層別因子は,変動の少ないものを2-3個
- ブロックサイズは,一般的に
 - 治療群の数の倍数
 - 4-6が多い

盲検化

- 試験を実際に行う人が割付結果を知らないこと
- 比較治療同士が見分けがつかないことが前提
- 研究過程でのバイアスの混入を防ぐ
 - 割付の遵守におけるバイアス
重症な患者に新治療を割り付ける
プラセボ群の被験者が他の薬に頼る
 - 結果の評価におけるバイアス
よくなった気がする
- ランダム化は研究開始時点でのバイアスを防ぐ手順

盲検のレベル

- 非盲検
- 単盲検
 - 被験者のみが知らない
 - 併用治療や評価においてバイアスが入る
- 二重盲検
 - 臨床医も被験者も知らない
- 三重盲検
 - データモニタリング委員会もどっちが実薬か知らない
 - 安全性に問題がある

盲検化のコスト

- 盲検化に必要なもの
 - プラセボの製造
 - コード番号の貼付
 - 記録の保持
 - コード番号のみで配送,配薬
 - 薬剤部との連携
 - 緊急時のためのKey

盲検化の維持

- 検査結果を誰が管理するか
 - 治療者,被験者に検査結果を知らせない
 - ダミーデータを用いる
 - 基準値を超えたときはそれだけを伝える
- 被験者に対する推量テスト
 - バイアス混入の可能性について議論する

盲検化出来ないとき

- わかってしまう薬もある
 - におい,色,味,注射跡,痛み,、、、
- PROBE試験
 - Prospective Randomized Open Blinded-Endpoint
 - 結果の評価だけは盲検的に行う
- できるだけ変動の少ない結果指標が良い

医学検査に関する研究をデザインする

2008年8月18日 勉強会
内分泌代謝科臨床研究フェロー
磯島 豪

本章の内容

- 診断検査の有用性を評価する
- 測定の再現性に関する研究
- 測定の正確性に関する研究
- 臨床判断における検査結果の有用性に関する研究
- 診断検査の実施可能性、コスト、リスクに関する研究
- 検査実施がアウトカムに及ぼす影響に関する研究
- 診断検査研究のデザインや分析における留意点

表12.1が本日のまとめ

Question	Possible Designs	Statistics for Results*
How reproducible is the test?	Studies of inter- and intraobserver and laboratory variability	Proportion agreement, kappa, coefficient of variation, mean and distribution of differences (avoid correlation coefficient)
How accurate is the test?	Cross-sectional, case-control, or cohort-type designs in which a test result is compared with a gold standard	Sensitivity, specificity, positive and negative predictive value, receiver operating characteristic curves, and likelihood ratios
How often do test results affect clinical decisions?	Diagnostic yield studies, studies of pre- and posttest clinical decision making	Proportion abnormal, proportion with discordant results, proportion of tests leading to changes in clinical decisions, cost per abnormal result, or per decision change
What are the costs, risks, and acceptability of the test?	Prospective or retrospective studies	Mean costs, proportions experiencing adverse effects, proportions willing to undergo the test
Does doing the test improve clinical outcome or have adverse effects?	Randomized trials, cohort or case-control studies in which the predictor variable is receiving the test and the outcome includes mortality, morbidity, or costs related either to the disease or to its treatment	Risk ratios, odds ratios, hazard ratios, number needed to treat, number and ratios of desirable and undesirable outcomes

* Most statistics in this table should be presented with confidence intervals.

診断検査の有用性を評価する

- How reproducible is the test?
 - 測定の再現性
- How accurate is the test?
 - 測定の正確性
- How often do test results affect clinical decisions?
 - 検査が臨床判断に及ぼす影響
- What are the costs, risks, and acceptability of the test?
 - 検査に伴うコスト、リスク、実施可能性
- Does doing the test improve clinical outcome or have adverse effects?
 - 検査のアウトカムへの影響

医学検査の研究における一般事項

- ゴールドスタンダード
 - 病気の存在を示すもの、定義上のもがある
- 患者のスペクトルと検査結果
 - サンプルの選び方で結果が変わる、
 - スペクトルバイアス: 観察集団と母集団の特性が違う時に起こる(患者の特性や検査結果の特性)
- 変動の原因、一般化可能性、サンプリング法
 - 変動の原因が、患者間の個体差の時もある
 - 多くの検査結果は、実施する人、判定する人、環境にも影響を受ける
 - 技量や判定能力を要する検査には、検査や判定を行う人を適切にサンプリングする
 - 正確性やコストが施設間で異なる場合は、複数の施設で研究を行うことで一般化可能性を高める
- 盲検化の重要性
 - 検査以外の情報により判定が影響される
- 実費と定価
 - 定価は高めに設定してあるので、実費を算出した方が良いが、実際には難しい
 - 平均経費を直接算出したり、定価に各施設の平均的「実費・定価比」を掛ける方法

(1) 測定の再現性に関する研究

- 観察者内変動 intraobserver variability
 - 同じ測定者あるいは施設が測定を行った場合の結果の変動のこと
- 観察者間変動 interobserver variability
 - 複数の測定者間で、検査結果が異なること
- 再現性の研究にはゴールドスタンダードは必要ないが、同じ結果が出て正しいとは限らない
- デザインは横断研究
- 検査の全プロセスを総合した再現性を評価するのが普通
- 検査の開発または改善を評価する際には、一部のステップの再現性を評価することもある

カテゴリー変数の分析

全一致率(concordance rate)

- 測定者間で完全に測定が一致した割合のこと

カッパ係数(kappa statistics)

- 偶然によって期待される一致を除外して評価するために開発された指標
- 合計の数字から期待される一致からどれほど実際の一致がかけ離れているかを示す
- 1(完全不一致)から+1(完全一致)の範囲をとり,0は一致度が偶然によって期待しうる程度と全く等しい
- (>0.8: very good, 0.6-0.8: good)

カッパ係数(kappa statistics)

$$= \frac{\text{Observed agreement (\%)} - \text{Expected agreement (\%)}}{100\% - \text{Expected agreement (\%)}}$$

Interobserver Agreement on Presence of an S4 Gallop

	Gallop Heard by Observer 1	No Gallop Heard by Observer 1	Total, Observer 2
Gallop heard by observer 2	15 $20 \times (15/100)=3$	5	15
No gallop heard by observer 2	10	75 $80 \times (85/100)=68$	85
Total, observer 1	20	80	100

Note: The concordance rate is the percentage of the time two observers agree with one another. In this example, both observers either heard or did not hear the gallop in $(15 + 75)/100 = 85\%$ of cases.

全一致率85%,期待される一致71%,カッパ係数 0.48

High Agreement When Both Observers Know Gallops are Uncommon

	Gallop Heard by Observer 1	No Gallop Heard by Observer 1	Total, Observer 2
Gallop heard by observer 2	0 $5 \times (5/100)=0.25$	5	5
No gallop heard by observer 2	5 $90 \times (95/100)=90.25$	95	95
Total, observer 1	5	95	100

Note: When both observers know that an abnormality is uncommon, they will have a high concordance rate, even if they do not agree on which subjects are abnormal. In this case the observers agree 90% of the time, although they do not agree at all on who has a gallop.

全一致率90%,期待される一致90.5%,カッパ係数 -0.053

連続変数の分析

- 研究デザインによって異なる
 - 例えば,2種類の温度計の比較では,対になった測定値間の差の平均値と分布
- 連続変数の一致度には相関係数は用いない方が無難である(外れ値に大きな影響を受けるため)
- 多数の測定者,施設,機器の間での変動を評価するには一般には,変動係数(coefficient of variation)を用いる
 - 変動係数(CV)= 測定結果の標準偏差/平均値

(2) 測定の正確性に関する研究

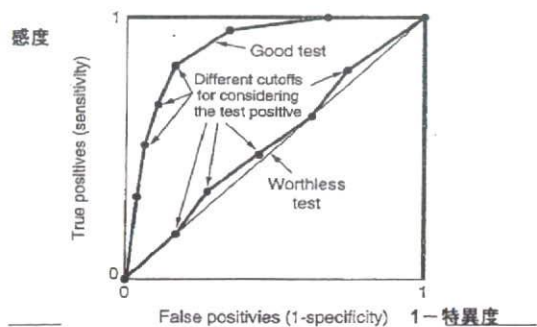
- 検査がどの程度正しい結果を与えるかについての研究
- 診断検査研究: ケースコントロール研究や横断研究に類似した研究デザインが用いられる
- 予後診断研究: コホート研究に似た研究デザインが用いられる
- アウトカムの判定を盲検化することが重要
- 倫理的にゴールドスタンダードが施行出来ない場合の工夫
 - 例えば,乳房撮影での陰性者に対する病理生検

感度/特異度, 陽性的中率/陰性的中率

		Gold Standard		
		Disease	No Disease	Total
Test:	Positive	a True-positive	b False-positive	a + b Positive predictive value* $a/(a+b)$
	Negative	c False-negative	d True-negative	c + d Negative predictive value* $d/(c+d)$
Total		e + c Sensitivity $a/(a+c)$	b + d Specificity $d/(b+d)$	

これらの指標は存在率 prevalence に関する指標で,発生率 incidence とは関係のない指標なので予後判定の評価指標には用いない

Receiver Operating Characteristic curve



尤度比

$$\text{Likelihood ratio} = \frac{P(\text{Result}|\text{Disease})}{P(\text{Result}|\text{No Disease})}$$

得られた診断結果が、どの程度まで標的疾患の検査前確率を上げるか、または下げるかを示す

尤度比が1より大きいと、検査前確率を増加させ、1より小さいと検査前確率を減少させる

JAMAのUser's guideによると、

尤度比により、検査前確率から検査後確率の変化は

・ >10 or <0.1 ⇒ 大きな変化、決定的な変化

・ 5-10 or 0.1-0.1 ⇒ 中等度の変化

・ 2-5 or 0.2-0.5 ⇒ 小さな変化

・ 1-2 or 0.5-1 ⇒ わずかな変化

Example of Calculation of Likelihood Ratios from a Study of Predictors of Bacterial Meningitis or Bacteremia Among Young Febrile Infants

White Blood Cell Count (per mm ³)	Meningitis or Bacteremia		Likelihood Ratio
	Yes	No	
<5,000	5 6%	96 4%	2.0
5,000-9,999	18 29%	854 39%	0.7
10,000-14,999	8 12%	790 36%	0.3
15,000-19,999	17 27%	266 13%	2.1
≥20,000	15 24%	151 7%	3.4
Total	83 100%	2,177 100%	

(3) 臨床判断における検査結果の有用性に関する研究

- 検査が正確でも、疾患が稀な場合は有用性は場合による
- 陽性になることの多い検査でも、理学的所見を超える情報を提供しない場合は意味がない
- 診断実益研究 diagnostic yield studyでは、
 - 検査の指示が出された時に、その検査結果が異常となる割合
 - その検査結果の他の情報による予測可能性
 - 異常所見が出た場合の患者の利益と不利益
- 前後比較研究 before/after studyでは、
 - 検査結果が臨床判断に及ぼす効果を直接評価する
 - 臨床判断の良否を評価するには、疾患の予後や治療効果が明確である必要がある
- 分析は検査結果の割合を、95%信頼区間で提示する

(4) 診断検査の実施可能性、コスト、リスクに関する研究

- 記述的研究である
- サンプルングが重要
- 分析は、一般に平均、標準偏差、中央値、範囲、度数分布といった単純な形で表現される
- 実施可能性や、有害影響のリスクは明確に区分できないことが多いので、許容範囲の基準を設定する必要がある
- 定価=コストではない、
 - 直接に算出する方法や、定価料金に各施設の平均「実費-定価比cost-to-charge ratio」を掛けて間接的に算出する方法がある

(5) 検査結果がアウトカムに及ぼす影響に関する研究

- 検査を受けた人のアウトカムが、受けなかった人よりもよいことを示す
- ランダム化比較試験が理想的だが、実際には観察研究のスタイルをとる
- 予測因子となっているのは、検査実施だけではなく、その後それに伴って生じるすべての事象である
- アウトカムは有害事象も含めて把握する

観察研究

- 適応の交絡に注意する
 - 検査を受けた人と受けない人の、背景の違い
- 治療や経過観察に標準的な様式が存在しないことに注意する
 - 検査後のフォローアップに問題がある可能性もある

スクリーニング検査に関する優れた観察研究

大腸癌で死亡した患者とそうでない患者で、S字結腸内視鏡検査の受検頻度を比較して、オッズ比0.41(95% CI 0.25-0.69)を得た。さらに交絡を考慮して、近位結腸癌(S字結腸内視鏡では検査出来ない)とS字結腸内視鏡検査の受検頻度の比較では、オッズ比0.96(95% CI 0.61-1.50)という結果を出した。

臨床研究

- ランダム化比較試験が行えると良いが、現実的には困難なことが多い(有症患者の医療にすでに使われている検査の場合は、特に困難)
- スクリーニング検査の場合には、ランダム化比較試験が行えることが多い
- 分析は研究デザインに依存する

臨床検査研究のデザインや分析における留意点(1)

- Verification bias 1: Selective application of a single gold standard

例) 足関節の外傷の人の骨折の予測因子としての関節腫脹の感度/特異度の研究で、レントゲンを撮影した人だけについて調査すると、腫脹の無かった人は、レントゲンを撮影されない傾向があるので、2×2表でc、dの数値が減少するため、実際よりも感度が上昇し、特異度が減少する

対処法:

ゴールドスタンダードが適用されないための研究に含まれない人についても、厳格にゴールドスタンダードの適応基準を決める
一般的なゴールドスタンダードが適用されない人については、別のゴールドスタンダードを定める(次に示す問題が生じるが、)

	Fracture	No Fracture
Swelling	a	b
No swelling	ca	da

臨床検査研究のデザインや分析における留意点(2)

- Verification bias 2: Different gold standard for those testing positive and negative

例) 腸重積の診断で、超音波所見陽性の人は、注腸造影(ゴールドスタンダード)を受けるため、2×2表の数値は変化しないが、超音波所見陰性の人は、経過観察にて腸重積が否定される(もう一つのゴールドスタンダード)。しかし、超音波所見陰性で、経過観察中に自然寛解する人が存在するため、感度も特異度も増加させることになる。

	Intussusception	No Intussusception
Ultrasound +	a	b
Ultrasound -	c	d

臨床検査研究のデザインや分析における留意点(3)

- サンプルサイズの不足
疾患やアウトカムが稀であれば、きわめて多くのサンプルが必要
- 除外基準の不適切な適用
除外基準を分子だけに当てはめ、分母に当てはめない
- 結果の施設固有性
- 境界例や解釈不能例の除外
検査に伴うコストや不都合について研究する場合、検査実施失敗は重要な情報。患者自身に原因があるのかそうでないのかをはっきりさせておく。
境界領域については、それをカテゴリーの1つとして、尤度比を計算したり、境界域所見についてROC曲線を作成することが出来る

まとめ(1)

- 診断検査の有用性はさまざまな観点から評価可能。一般的には観察研究が用いられ、記述統計値と信頼区間が計算される
- 診断検査研究の対象者は、実地臨床を反映するものでなければならない
- 盲検化が重要
- 観察者内変動と観察者間変動などの再現性の検討は、検査評価の最初のステップとして行われる

まとめ(2)

- 検査の正確性の検討には、ゴールドスタンダードが必要
- 正確性を評価する指標として、感度、特異度、ROC曲線、尤度比などがある。予後判定の指標には、リスク比やハザード比がある。
- 診断検査の価値を最も厳密に評価するには、臨床試験を実施する必要がある

Chapter 13 : Utilizing Existing Databases

Designing Clinical Research 4th Ed.

25th August, 2008
National Center for Child Health and Development
Clinical Research Center
Takanobu Maekawa

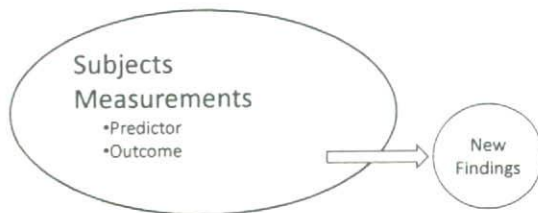
Introduction

What do you do first, to answer your
“Research Question”?

- Background Research
 - What is already known
 - What kind of databases are available

Database What is already known...?

- Databases medical records, cohort, registry...



Summary

- Many Research Questions can be answered quickly and efficiently using data that have already been collected.
- Three general approach to using existing data
 - Secondary Data Analysis
 - Ancillary Studies
 - Systemic Reviews

Secondary Data Analysis

- Data sets from Existing Databases

Databases?

- Medical records
- Previous studies
- Other databases

Databases: examples

- Tumor Registries
 - Surveillance, Epidemiology, and End Results (SEER) Program
 - Covering about ¼ population
- National Death Index
 - National registry
 - All death in US since 1978
 - Name, Birth date, social security number

Databases: examples

- Databases developed by
 - Medicare
 - Department of Veterans Affairs
 - Kaiser Permanente Medical Group
 - Duke Cardiovascular Disease Databank
 - San Francisco Mammography Registry
 - National Registry of Myocardial Infarction (NRFMI)

Secondary Data Analysis

- Advantage
 - Time saving
 - Cost saving
 - Reflects “real world” than specific trial setting
- Disadvantage
 - Cannot control study design
 - Population
 - measurements

Secondary Data Analysis

- One good source of data for secondary analysis is...
 - Completed research project @ your institution
 - Public databases

Secondary Data Analysis

- Large community-based data sets
- Useful for studying
 - Effectiveness (\Leftrightarrow efficacy) of intervention (Tx)
 - Utilization of intervention
- Discovering
 - Information on rare adverse events

Ancillary Studies

- Advantage
 - Time saving
 - Cost saving
 - To some extent, you can specify a few key additional measurements for well controlled subjects, with little cost and effort
- Disadvantage
 - Cannot control fundamental study design

Ancillary Studies

- Good opportunity for you...
 - Cohort study
 - Clinical trials
- that include either the predictor or outcome variable for your research question

Systematic Reviews

- Opportunity for a new investigator
- Can be familiar with the literature
- Can be “expert” on the research question

- Meta-analysis
- Statistical aspects of a systemic review

Elements of a Good Systematic Review

1. Clear Research Question
2. Comprehensive identification
3. Clear Definition (inclusion and exclusion)
4. Uniform and Unbiased abstract of findings
5. Clear and Uniform presentation of data
6. Summary Effect Estimate and Confidence Interval
7. Assessment of the heterogeneity
8. Assessment of publication bias
9. Subgroup and sensitivity analysis

Designing Clinical Research 4th Ed.
Chapter 13, Table 13-1

Systematic Reviews

- Presentation of the Data

- Three types of information
 - Characteristics of each study included
 - Sample size, outcomes, length of follow-up, characteristics of the population, methods...
 - Results of the individual studies
 - Risk estimates, confidence interval, p-values
 - Summary Effect Estimate and Confidence Interval

Systematic Reviews

- Summary Effect Estimates
 - Essentially average effect weighted by (1/variance) of the outcome of each study
 - Small studies result in a large variance, large studies result in a small variance
- Variance of Summary Effect
 - Various formula
 - $(1/\sum \text{weight})$
- Summary Estimates $\pm 1.96 \times \text{Variance}^{1/2}$

- Different approach, different result

Systematic Reviews

- Heterogeneity of the studies
- Approach to heterogeneity
 - Review the individual studies
 - Study design, study populations, intervention, or outcome...
 - Test of homogeneity
 - Null hypothesis (studies are homogeneous)
 - Usually use Chi-square test, P-value 0.10
 - Subgroup and Sensitivity Analyses

Systematic Reviews

- Publication Bias
- Two way to deal with Publication bias:
 - Identify unpublished studies
 - Sensitivity analysis : does adding these unpublished results change the summary estimate?
 - Difficult to identify, difficult to abstract the required data
 - Estimate publication bias
 - Kendall's Tau, a coefficient of correlation
 - Biased Funnel Plot: smaller the sample size, higher RR