

臨床研究 デザイン

学習ノート

長谷川明弘

東洋英和女学院大学

www.hasegawa-akihiro.com

1

「臨床研究デザイン」とは？

- 臨床活動の中から生まれた疑問から研究成果を導き出すために研究の計画を立て運営しながら実行すること
- 大切なのはユーザーの為になる研究であること
 - ユーザーとは
 - クライアント、患者、利用者など専門職がサービス提供する対象
 - 実践者が研究成果を臨床現場で役立てられること

2

臨床現場の使命 AIMES

実践者が研究の成果を

- 適切に(**Appropriately**),
- 素早く(**Immediately**),
- 安価に(**at Minimal cost**),
- 効率よく(**Efficiently**),
- 安全に(**Safely**)

ユーザーに届けること

3

ガイドラインとは？

- 適切な方法を促進し、構築されたエビデンスを実践に浸透させる役割がある
- ガイドラインに繋がるエビデンスの構築が臨床研究には求められる

4

臨床研究の意義

● **Efficacy**

- 実験など統制した(理想的な)条件下での有効性

● **Effectiveness**

- より現実的な(実践)場面での有効性

臨床研究は、この2つのEの有効性を埋める意義がある。構築された研究成果(evidence:主張を検証するために提出された証拠)が実践方法を改善し、ユーザーのためになることが最終目標となる

5

リサーチクエスチョン 研究したいことを宣言すること

- 日常の実践の中で生じた疑問や、日頃、観察する「人の心(広い意味で行動)」に関する疑問から生じる「研究テーマ」のこと
 - 心理学、医学、哲学、宗教学、生理学、脳科学など、どのような学問領域から疑問を解消するか
- 疑問をどのように研究テーマにするのか
- リサーチクエスチョンの質をチェックする基準がある
 - FINER基準やFIRM²NESS基準と呼ばれる

6

リサーチクエスト 究極的な影響力

- 研究成果が臨床活動を変えうること
- 研究成果がユーザーに生じる良い結果を生み出すこと
- 研究成果が制度や政策を変えること

7

FIRM²NESS基準

- Feasible(実施可能性があり)
- Novel(新奇性があり)
- Interesting(真に興味・関心のあるテーマで)
- Relevant(切実な問題で)
- Modifiable(要因やアウトカムが改善可能な)
- Measurable(測定可能な・可視化できている)
- Ethical(倫理的で)
- Specific(具体的で・絞り込まれている)
- Structured(構造化された)

(福原,2008)を一部変更

8

FINER基準

- Feasible 実施可能性
- Interesting 真の興味
- Novel 新奇性
- Ethical 倫理性
- Relevant 必要性

(Hulley and Cummings, 1988)

背景情報

Background Information

- 日常業務の中から漠然と出てきた疑問(クリニカルクエスト)からリサーチクエストへと構造化する
- 過去の先行研究(Background Information)を把握する事が重要となる
 - 何が分かっているのか
 - 何が分かっていないのか
 - 研究によって何を明らかにすべきか

10

文献検索の切り口

- 類似した研究テーマの出現頻度と動向
- どのような専門家が実施してきたのか
- 年齢層や対象はどうなっているのか
- 測定しておくべき項目を把握すること
- 測定バイアスはどうなのか
- 実施上の課題はどうなのか
- 批判的な目で文献を吟味する

11

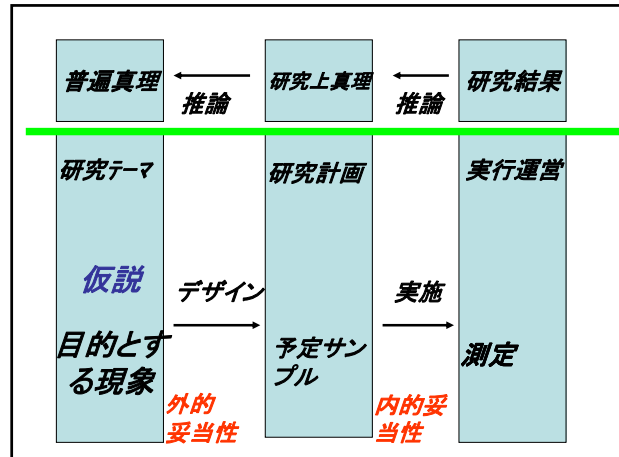
文献検索

- 類似した研究テーマの出現頻度
- どのような対象や領域で実施されてきたのか

12

Feasible(実施可能性)

13



母集団とサンプル

- 母集団(population, universe)
- サンプル・標本(sample)

15

サンプリング

- 母集団から標本を抽出すること
 - 簡便法
 - 確率法
- バラツキ
 - 測定をするたびに一致した値が得られないこと
 - 測定回数を増やすことで解消する
- バイアス
 - 測定を繰り返して誤った値を信じてしまうこと
 - 手続きの不統一、測定機器の変更から生じる

16

サンプリング 簡便法

- 参加施設を選択し、そこを利用した対象者を原則として全数調査する
- 選択基準を厳密にする必要がある
 - 厳密にしないとバイアスの入る可能性が高い
 - 選択基準
 - 調査時期や対象年齢など

17

サンプリング 確率法

- 母集団から確率的に標本を選ぶ方法のこと
- 全国の中から参加施設を無作為(ランダム)にいくつか選択する。各施設内の対象者もある比率で無作為に選択する。
 - 実際は、ターゲットをきめて選択基準を満たした対象者を連続的に抽出する方策がとられている。
 - ランダム・サンプリングは外部妥当性(一般化可能性)について一定の保証をしてくれる
 - なお無作為割り付けとは別の概念

18

サンプルサイズ

- 研究から得られる統計的に有意な差として検出されるために必要なサンプルサイズを確保する必要がある
- 対象者数が適切であること
 - 対象者の範囲
 - 研究目的
- 算出できない場合は、1群10–15人程度を想定することが多い
- 予備調査を行うことが好ましい

19

選択基準

対象者の範囲を具体的にする

- 取り込み基準 Inclusion Criteria
 - 年齢、性別、研究目的に沿った条件を備えていること
 - 募集範囲と期間
- 除外基準 Exclusion Criteria
 - 必要最低限は除く必要が出てくる。追跡調査が困難、データにバラツキが生じる、倫理的な判断

20

外部妥当性と内部妥当性

- 外部妥当性(external validity)
 - バイアスがないようにサンプルを集めることで高められる
 - 一般化可能性(generalizability)
- 内部妥当性(internal validity)
 - サンプル内部の結果保証の程度
 - 研究目的に沿って不完全な標本であるとサンプルから除外すること(バイアスによる除外)で保証できなくなる。

21

サンプリングのまとめ

- サンプルから得た結果を母集団に一般化するためには、確率的サンプリングがベスト
- 層化抽出法をお勧めする
- (人種)、経済的背景、年齢、性別の各群から等しい数の対象者を設定する。とりわけ、自分の研究テーマにふさわしい性別年齢を設定して、均等になるように集める。
- リクルート法を確立する

サンプリングに伴う誤差

- 偶然誤差 random error 精度・尺度に関連する
精神科にうつ病患者が最初に訪れる真の確率が20%だった場合に、100人のうつ病患者からサンプリングした場合、20例が精神科医に訪れているはずだが、実際には18や23などの数値が得られる。この差分は誤差である。すなわち、原理的にいって、誤差を含んだ数値しか測定することができない。誤差を減らすには対症患者数を増やすしかない。ある程度の偶然誤差は避けられない
- 系統誤差 systematic error 正確度・真度に関連する
何らかのバイアス(歪み)によって、測定結果が特定の方向にずれること。サンプルサイズを増やしても改善されない。これは、研究デザインの改善、特に測定法の改善によって系統誤差を小さくできる

実験のための3原則 Fisher

実験を考える際

- まず局所管理の原理により系統的な誤差をできるだけ排除し(local control)
- それでも残るどうにもならない系統誤差をランダム化の原理により偶然誤差に転化させ(randomization)
- さらに反復の原理によりそれらの誤差の大きさを評価する(replication)

24

精度・定度(precision)

- 再現性 一致度 信頼性 が 同義語
- 測定者による誤差
同じ測定者が繰り返し測定する際
複数の測定者が測定する際
- 測定手段による誤差
同じ測定装置で繰り返し測定する際
別の測定装置で繰り返し測定する際
- 対象者による誤差
自己記入式質問票の回答の仕方
面接の際にどのように回答するか

精度・定度の向上法

- 測定方法の標準化
実施マニュアル、測定法のコツ、間違いの有無の確認法、間違いの修正法について記載
SCID 質問紙 開かれた質問の回答整理
- 測定者の技能の向上
勉強会やセミナー 練習
- 測定手段の改善
予備調査の過程で改善する
本実験では変更せず一貫して同じ方法で行う
- 測定の反復

正確度・真度(accuracy)

- 真の値を測定できているか？
- 測定者バイアス 血圧を低めに読む傾向、構成的面接法で誘導質問する傾向など
- 対象者バイアス 脂肪を取るのが悪いと考えているとあまりとってないように回答するなど
- 測定手段バイアス キャリブレーションをしていないなど
- ゴールデンスタンダードと感度と特異度を比較することで真度がわかる ない場合は妥当性を検討する

正確度・真度の向上法

- 精度・定度の向上法に加えて
- 気づかれない方法で測定する
- マスク化
- 機器のキャリブレーション

妥当性

- 内容妥当性
表現妥当性: 質問が聞きたいことを聞くために適切な表現になっているか？
包括妥当性: 調べたい概念の全側面を包括的にあますことなく聞いているか？
- 構成概念妥当性
健常者とケースとの間で絶対異なる特性を持たなければならない場合、調べ方がその二つを区別しているか
- 基準妥当性
関連する他の測定値と正しく相関をもつか？
- 予測妥当性
関連するはずの事象を予測している関係を持つか？

サンプルサイズ

- 研究デザインにふさわしいサンプルサイズ
- 記述的研究の場合仮説はいらないが予備調査や過去の文献から分布率を把握しておく
- 仮説はその他の研究では必要
- 簡潔で的確で始める前に文章化する
- 差なし仮説と差あり仮説
差なし仮説の有意性が統計学的に否定されれば差があることが検定される

研究デザインにおける基本概念

- 実験ユニット
 - 観察や測定を行う際の実体
 - 各ユニットは互いに独立していることが前提
- ブロック
 - 一様な実験ユニットから構成されるグループ
 - 測定する箇所に分けて実験することをブロッキングと呼ぶ
 - 各グループはブロック(層)と呼ばれる
 - 性別に分けた場合、男と女はブロックになる

31

研究デザインにおける基本概念

- レスポンス
 - 実験目的のアウトカム変数のこと
 - エンドポイントとも呼ばれる
 - 最終的な判断基準など、評価指標を定義しておく必要がある
 - 臨床的に意味があり、分かりやすいものを指標に選ぶべき
- 要因(factor)
 - レスポンスに影響する制御可能な変数のこと
 - 要因のある値のことを水準(level)と呼ぶ
 - 性別が要因の場合は、男性・女性がそれぞれ水準になる

32

片側と両側の仮説

- 予測因子とアウトカムが一方向の関連がある場合、片側で、両方向なら、両側 という
- クロスセクショナル研究の場合、両方向しか仮定できない
- 差がないということを証明したい場合は少ない
- したがって、「両側の差あり仮説」に基づいて、サンプルサイズを決める

多くの研究では

- $\alpha=0.05$ $\beta=0.20$ Power=0.80で設定される
- 多くの物質や尺度を測定する場合
仮説の主従をきちんときめておいて、主たる仮説が検証される条件でサンプルサイズを決めること

サンプルサイズの決め方

- 臨床研究の場合、最終的に統計を行う際に、t検定 χ^2 乗 相関 の何れを使うで、計算方法が異なる
- 配布した表を用いてサンプルサイズを選んでください
- T test の場合
効果量 E (アウトカムの平均値の差)とアウトカムの標準偏差 S の比 E/S と α β の値を用いて決める

サンプルサイズが固定している場合

- 限られた患者数しか得られないだろうなあと
いうとき
- **パワーを固定して、効果量を推定**
効果量が大きなものだけ研究対象になる
- 効果量を固定して、パワーを見積もる

サンプルサイズを小さくする

- 連続変数を使用する
- ペアー検定を用いる
- より定度の高い測定法をもちいる ばらつきが減るので。
- ケースコントロール比を 1:5 などにする
- より頻度の高いアウトカムを用いる

α エラー、 β エラー、効果量

- 第一種の過誤 α エラー
疑陽性 ないのにあるというまちがい
- 第二種の過誤 β エラー
疑陰性 あるのにないというまちがい
- 効果量
自分の見たい関連の強さ、差の大きさ
A B 各群が、a b の値をとるとき
 $|a-b|$ が 両者の差で 効果量

検出力 パワー

- 始める前に α エラーと β エラーのレベルを決めてから研究すること
- $\alpha=0.05$ とすると、ないのにあると判定しまう確率を5%許容するという意味
- $\beta=0.10$ とすると、効果量を10%の確率で見逃してしまうことを意味する
- $(1-\beta)$ は、この時、90%であるから、効果量を90%の確率で検出できることを意味する
- したがって、 $1-\beta$ を統計学的パワー(検出力)と呼ぶ

統計学的パワー

- サンプルサイズに依存する
- アウトカムの発生頻度に依存する

40

α エラーと β エラー

41

effect size

42

SN比

- シグナル
 - 効果の指標に対する2群の差
- ノイズ
 - 調査対象個々の違い
 - デザインによっては個人内のバラツキを設定することで少ない標本数で検証を試みることがある
- シグナルがノイズに比べて大きいかどうかで真の差かどうかを検証する
 - 研究ではノイズを少なくすることが望ましい

43

研究デザインの質

1. 系統的レビュー (systematic reviews)やメタ・アナリシス(meta-analysis)-*2次的研究*
2. ランダム化比較対照試験(randomized controlled trials)-*介入研究*
3. 非ランダム化比較試験(controlled trials w/o randomization)-*観察研究*
 - A) コホート研究またはケースコントロール研究
 - B) 事例集積、比較群のない画期的実験
4. 事例研究(case study)-*記述研究*
5. 著名研究者や権威の意見

44

事例研究 (case study)

- 新たな発見や考察を提供したり、介入法の議論が目的となることが多い。
- 得られた結論に関して、一般化が難しい。
- **仮説生成的研究**
- 最近では、Narrative Study という研究方法がある。 **質的内容分析**
 - 来談者の語る言葉をそのまま用いて分析する
- **関連研究の裏打ちにもなる。**
- 研究者の、着想ならびにアイデアが問われる

記述的研究

- どのような患者が、人口中に何人いるか
- どのような心理的特性が発がん後に生まれるか
- 仮説は不要で、難しい統計はいらない
- しかし、測定の真度、定度をいかに高めるか、また、サンプルサイズをどの程度にするかをとことん追求して行うこと

研究デザインの分類

- **観察研究**(observational study)
 - 事例の抽出を工夫するが、原則的には現在行っている実践活動を継続しながら仮説検証を試みる
 - 仮説形成、探索的な研究に適している
- **介入(実験)研究**(intervention (experimental) study)
 - 積極的な実践活動を行うことで仮説を検証しようとし、現在の実践活動を変更することが生じうる
 - 仮説検証、確認的な研究に適している

47

観察研究 observational study

- 要因の間の相関、関連を発見するための研究
- 事例集積研究(case series study)
 - 横断研究(cross-sectional study)
 - 縦断研究(longitudinal study)
 1. ケースコントロール研究(case control study)
 2. コホート研究(cohort study)
 - a. 前向き(prospective)
 - b. 後ろ向き(retrospective, historical)
 3. ネステッド・ケースコントロール研究(nested case-control study)

48

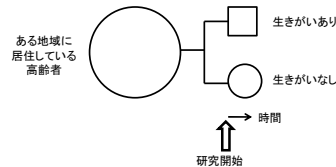
事例集積研究 case series study

- クリニカルクエスチョンに基づいた実践活動を行った事例研究をデータとして蓄積する研究
 - 比較的短期間に観察された事例の報告
 - コントロール群は設定していない
 - 探索的な研究デザイン
 - 仮説形成を目標とする

49

横断研究 cross-sectional study

- 現時点においてランダムに標本抽出を行ってデータを集める
 - 時間軸を考慮していない
 - 因果関係を立証するには弱い研究デザイン
 - 統計解析が複雑になる傾向

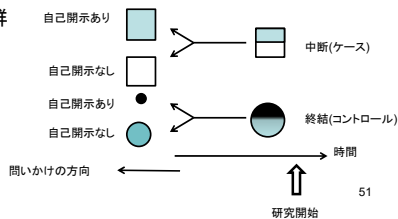


50

ケースコントロール研究 case control study

— 縦断研究 —

- 時間軸を考慮している(後ろ向き retrospective)
- 研究開始時点から過去に遡ってデータを集める
- ケース毎に要因以外の条件を一致させる
 - 各ケースに対してコントロールは数例選択する
- コントロール群の選択が難点
- 短期間で容易

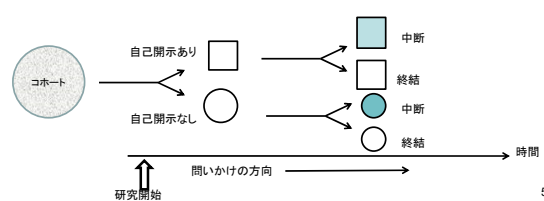


51

前向きコホート研究 prospective cohort study

— 縦断研究 —

- 時間軸を考慮している(前向き prospective)
- 長期間追跡が必要である(追跡率の高さが必要)
 - イベントの設定を追加して多くの仮説検証が可能となる
 - 脱落ケースが生じる可能性もある

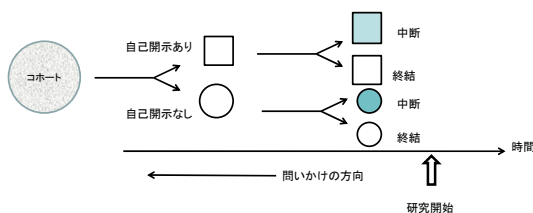


52

後ろ向き(既存)コホート研究 retrospective(historical) cohort study

— 縦断研究 —

- 時間軸を考慮している(後ろ向き retrospective)
- 既存のコホートの一部分を用いて因果関係を検証しようとする研究デザイン

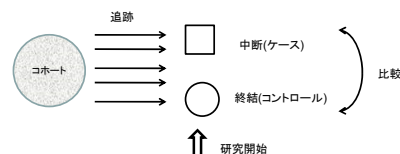


53

ネステッド・ケースコントロール研究 nested case-control study

— 縦断研究 —

- 時間軸を考慮している(後ろ向き retrospective)
- 既存のコホート研究かランダム化臨床試験の一部を用いて因果関係を検証しようとする研究デザイン。
 - 研究仮説を持った時点でコホート内からケースを同定し、要因を一致させたコントロール群を選択



54

介入(実験)研究

intervention (experimental) study

相関に関する仮説を実証するための研究

- 比較対照試験(controlled trial(concurrent control))
 1. パラレル(parallel)
 - a. ランダム化(randomization)
 - b. 非ランダム化(non-randomization)
 2. 逐次(sequential)
 - a. 自己対照(self-controlled)
 - b. クロスオーバー(cross-over)
 3. 外部対照(既存対照含む)(external (historical control))
- 対照なしの研究(uncontrolled study)

55

個人間比較と個人内比較

- 個人間比較
 - 実験群と対照群を設定
 - 個人間の変動(個人差)を誤差(ノイズ)と考える
 - 2群間の平均的な違い(シグナル)を比較する
- 個人内比較
 - 2種類の介入を提示して介入前後を比較する

56

介入の割当方法

- パラレル比較法:parallel comparison(平行比較法、群間比較法、同時対照試験)
- 個人対照法:self-controlled(前後比較、左右比較)
- クロスオーバー法

57

パラレル比較法

介入の割当方法

- パラレル比較法:parallel comparison(平行比較法、群間比較法、同時対照試験)
 - 個人間比較を採用する。
 - 介入開始という出発点から比較する群を設定して平行して追跡する。

58

個人対照法 介入の割当方法

- 個人対照法:self-controlled(前後比較、左右比較)
 - 個人内比較を採用する。
 - 各個人が対照として割り当てられ、介入の前後で比較する。
 - 個人への介入をランダムに割り当てるN-of-1デザインも提唱されている

59

クロスオーバー法 介入の割当方法

- 比較したい介入を交差させるデザイン
 - 介入時期と非介入時期を設定可能
 - 2つの介入を比較設定可能
 - サンプルサイズを減らせる
 - 時期効果、持越効果を考慮する必要がある
 - 結果が出るまで時間を要するため脱落サンプル増加

	群	時期		
		I期	II期	III期
(a)	群1	A => B		
	群2	B => A		
(b)	群1	A => B	=> B	
	群2	B => A	=> A	
(c)	群1	A => B		
	群2	B => A		
	群3	A => A		
	群4	B => B		

60

系統的レビュー (systematic reviews)

- 論文をデータベース検索や参考文献リストなどによって収集し、そのデータを総括して評価すること。
- 1. 研究テーマ、リサーチ・クエスチョンを適切に設定する。
- 2. 研究を漏れなく収集する。出版されていないトライアルを含めて収集する。
- 3. 各研究の妥当性の評価をする。
- 4. 評価の結果を要約書式にまとめる。
- 5. メタアナリシスによる統計解析を行う。
- 6. 解析結果を解釈する。
- 7. 総括して、定期的に更新してユーザー届ける。

61

曖昧な疑問を構造化する PECOとPICO

- Participants
 - 誰に
- Exposure / Intervention
 - 何の要因によって / 何をすると
- Comparison
 - ~の要因がない(または 他の要因がある)
 - 何をしない (または他のことをする) のと比較して
- Outcomes
 - どうなる

62

説明変数とアウトカム変数を区別すること

- 説明変数(独立変数・目的変数、予測変数)
 - 原因と見なす項目のこと
 - 要因や介入となる要素のこと
 - アウトカム変数(従属変数・結果変数、目的変数、基準変数)
 - 研究における結果と見なす項目のこと
 - 測定可能であること(信頼性と妥当性が検証済)
 - 客観的指標: ドロップアウト数、面接回数、生理指標
 - 主観的指標: 心理尺度
- 主要なアウトカム変数を1つに絞り込むこと

63

構造化抄録 structured abstract

- 背景・目的
 - Context, Objectives
- 研究デザイン
 - Design
- データ収集状況
 - Setting
- 標的集団、参加者
 - Subjects, Participants
- 介入または要因
 - Intervention, Exposure
- 測定手段・主要なアウトカム変数と統計学的手法
 - Main Outcome measure & analysis
- 結果
 - Result
- 結論
 - Conclusion

64

プロトコール＝研究デザイン

- プロトコール(研究計画書)を予め用意する
- 背景
- 研究課題(目的、仮説)
- デザイン(比較の方式)
- 評価変数(エンドポイント)
- 統計的考察(事例数と期間の根拠、解析法)

65

予備調査

- 特定のテーマを選んで、催眠をした場合にある効果が想定される場合
- 催眠をしたときとしない時の その効果の大きさの差を見極める たとえば、NK活性が40%が、50%に上がったり、20%が30%にあがるなら、効果量は10%となる
- この数字と、SD値(適当に決める)から、ケースコントロールスタディーのサイズが決まる

本研究

- マニュアルの記載を始める
- 対象者 数 期間 催眠の方法 手順
- データ採取の流れ
- 統計解析の手法
- どういう結果になれば成功か はっきり認識
- 実施

効果判定に用いる測度

- 論文では明確に記述する必要がある
 - 海外で広く使用されている尺度を用いることが望ましい。
- 統計量を記述し、追試やメタ分析に活用できるように示すことが必要であろう
 - 事例研究では、得点の素点を示す

68

Dismounting study

- dismount
 - 〈物を〉(台座・支えなどから)降ろす,取り外す
- 臨床現場では、さまざまな要因が複合的に入り交じっている。
- 介入も要素が複雑に構成されている。
- 介入法のどんな側面が有効性を発揮しているのかを明確にするために、介入法を構成している要素を一つずつ取り除いて、同様の効果が得られるかを検討していく研究方法

69

「操作の妥当性」を保証するには

- 心理学的介入にバラツキがあつては結果が歪む可能性がある
- 心理学的介入について手順を統一するためのマニュアルや実施者の経験年数、研究協力者の除外基準を設定するなどの工夫が必要となる

70

参考・引用文献

- 福原俊一 2008 リサーチクエスチョンの作り方 健康医療評価研究機構
- 渡部一宏・福原俊一 2009 日常業務からResearch Questionへ Parma Tribune No.1から12
- 折笠秀樹 1995 臨床研究デザイン 医学研究における統計入門 真興交易医書出版部