

日本医学教育学会 医学教育研究開発小委員会主催  
第2回医学教育研究技法ワークショップ 報告

概要

日 時：平成18年10月14日（土）13:00～10月15日（日）12:00  
場 所：東京大学医学部総合中央館 3F 333号室

プログラム

【一般目標】

- ・医学教育研究を実施するために必要な研究手法を習得する

【個別目標】

- ・医学教育における自らの研究課題に対し、具体的な研究計画を立案できる
- ・医学教育研究に特徴的なデザイン、測定について列挙できる
- ・与えられたデータに対し、基本的な統計解析方法が実施できる

10月14日（土）

- 13:00-13:15 開会挨拶，概要説明
- 13:15-14:00 自己紹介，アイスブレイキング
- 14:00-14:30 講演：森本先生—リサーチクエスチョンから研究計画へ
- 14:30-16:00 グループ討論1：各自のテーマによる研究計画—デザイン，サンプリング
- 16:00-16:45 全体発表と討論
- 16:45-17:00 休憩
- 17:00-17:20 講演：大西先生—データの種類，統計解析
- 17:20-18:20 グループ討論2：同じテーマによる研究計画—データと解析の種類
- 18:20-19:00 全体発表と討論

10月15日（日）

- 8:30-9:00 福島統先生講演「医学教育誌編集委員会が目指す研究論文のあり方」
- 9:00-9:30 質問と全体討論
- 9:30-9:45 休憩
- 9:45-11:00 グループ討論3：仮想データに対する統計解析方法
- 11:00-11:45 全体発表と討論
- 11:45-12:00 閉会

## 開会挨拶、概要説明

東京大学医学教育国際協力研究センター 大西弘高

皆様こんにちは。医学教育学会医学教育研究開発委員会委員長の大西弘高です。本日は第二回医学教育研究技法ワークショップにお越しいただき誠にありがとうございます。

私は2000年から2002年までイリノイ大学に留学しておりました。その際診断推論に関する研究にタッチし、教育研究の方法論が通常の生物学的な研究と違うなと感じていました。また疫学的研究とは多くの要因を同時に検証する点で共通しますが、用いる統計手法が少し違っても気付きました。そういうことを医学教育学会ではどの程度考えて議論してきたのかと思って理事会で意見を出したところ、「では委員会を作って議論して下さい」ということになりました。

前期の委員会では、なぜ医学教育研究は難しいのかというような議論をしたりしてきましたが、現場で様々な問題を感じておられ、でも研究をするには少し困難があるような先生方と、理論的な面から教育研究に問題を感じる我々との間でもっとディスカッションが必要だろうと結論づけました。そういうふうなことを考えたときに、やはり皆さんといろんな意見、あるいは知恵を出し合っているようなことを考えていければ一番いいのではないかと考えて、こういうふうなワークショップを企画することになりました。実際のところ皆さんのいろんな意見を聴かせて頂きますと、その医学研究というか臨床研究にしる、基礎医学的な研究にしる、ある程度研究というもののアイデアというのは、むしろ教育研究をやり始めたような僕らみたいな人間よりも、皆さんの方がお持ちなのだろうと思います。ところが教育研究をいざやろうとすると、医学教育の用語とか作法がよく分からないと、上手くいかないのかなあと感じています。医学教育の作法については、医学教育のいろんなワークショップもありますので、そういうふうなところでおいおい身につけて頂かないと仕方がないんだと思うんですけども、そういうもののワークショップにある程度出て、じゃ研究やるとなるとなるとどこがその障壁とかですかね次の何かbreak throughになるんだろう、というところがなかなか皆さん見出せないというか、いい知恵がないというかですねそういうふう感じの声をよく聞きました。そういうところってどうすればいいのか、僕もまだまだ悩んでいるところではあるんですけども、今年のワークショップでは、とりあえず研究の全く基本的な話をして、皆さんに仮想データをもとにいろんな分析を試みて頂きました。講演は東大の教育学研究科の市川先生に、心理学・教育学等の研究の現況と難しいところということでお話し頂きましたが、この講演に対しての評価は随分高かったので、やはり皆さんそれなりに高い見識を持って教育研究とは何かということについて色々と学び、それを高めていきたいというふうな気持ちが非常に強いんだなという思いを持ちました。

今年は皆さんの“日頃いろいろ感じておられるような疑問点”とか“自分は実はこういうふうなテーマについて調べたかったんだ”という意見を出し合ってもらって、各グループでそういうふうな一つ、各自のテーマに従って研究デザインをしていくことを一日目にやろうと考えています。

それから明日は、福島先生、医学教育学会の『医学教育』という雑誌がありますけれども、そちらの編集委員会代表という形で福島先生にご講演頂くということになっています。『医学教育』という雑誌も基本的には研究した成果を論文に纏めて出すということになるはずなんですけれども、“研

究とは何ぞや”というところに関していろんな議論をなさっていて、結局投稿規程が変わるなどの動きが去年ありました。ただその動きがはたして我々の目指している方向とどのくらい合致しているのかというの少し分からない面もあり、ですね、そのあたりは意識の高い参加者の皆さんからのいろんな意見をむしろ頂きたいというふうなことを編集委員会の方からも聞いています。ですので、そういう一つのいい機会としてですね明日の講演というのを楽しみにして頂ければと思います。それから最後、明日の午前の後半はこちらから仮想データを提供して、それをいろんな形で纏めていくというふうな作業をやらしてもらおうかな、というふうに考えています。そういうことで、今日・明日の2日間にわたってですけれども、皆さんの熱心な参加を宜しくお願いします。

第2回医学教育研究技法ワークショップ  
2006年10月14-15日 東京大学

## リサーチクエスチョンから 研究計画へ

日本医学教育学会医学教育研究開発委員会  
京都大学医学教育推進センター  
森本 剛

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

1

## アウトライン

- 医学教育研究とは
- 医学教育研究を始める前の準備
- 研究デザイン
- 対象集団設定～入口
- 評価項目(介入・因子・結果)設定～出口
- 評価項目の信頼性と妥当性
- プロトコルとインフォームド・コンセント、IRB
- (変数の種類)

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

2

## 医学教育研究？

- 何でもいい
- いろんな職種・レベル ～ 対象集団subjects
  - 医師、看護師、薬剤師、技師、コメディカル、教員
  - 学生、研修中、スタッフ
  - システム？病院？患者？
- いろんな内容 ～ 評価項目measurements
  - 学業成績、知識、技術、態度 ⇔ 臨床研究
  - 教育的介入、因子との関連、時間的变化

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

3

## お言葉

- 評価項目 measurement
  - 研究者が介入や因子の与える影響をみたい指標
  - 成績、国試合否、態度、技能、、、
- 介入 intervention
  - 研究者が評価を目的として、導入・制御する因子
  - 特別な講義、小テスト、ローテーション、研修先
- 因子 factor
  - 集団の中に偶然にばらついて存在し、評価項目に影響を与える要素
  - 出身校、男女、出席率

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

4

## 医学教育研究

- 学会
  - 日本医学教育学会
  - Association of American Medical Colleges
  - Association of Medical Education in Europe
  - Society of General Internal Medicine
- 雑誌
  - 医学教育
  - Medical Education
  - Academic Medicine
  - Medical Teacher
  - Journal of General Internal Medicine

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

5

## 準備

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

6

## 研究を始める前に考えること

- テーマを単純・明快に
  - セッティング・対象集団・介入(因子)・評価項目を明確にする
- 意味がある？既に答えがない？
  - 同僚などに聞いてみる
  - 教科書・レビュー論文・MEDLINE
  - “似たような”研究がある？ OK！

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

7

## テーマを明確に

- 「小試験によって医学生の学習意欲は高くなるか？」
- 「毎講義の最後に行われる小試験は医学部4年生の臨床系授業において質問紙法による講義への学習意欲を高くするのか？」
- 現実≫理想
- 自分(自校)の教育環境に合わせることも重要

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

8

## 研究の意義

- 同僚・若手(学生?)や上司などに聞いてみる
  - 同僚・若手～データ集積に協力してくれるかも
  - 上司～研究の便宜を図ってくれるかも
- 過去に似た研究あり
  - 全く同じ**介入**の研究があれば、確かに魅力↓
  - 対象集団・評価項目etc. みーんな一緒ですか？
  - 科目や教育環境の違いが影響しそうですか？
  - 何か小さなことでも**付け加える**ことはありそう？

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

9

## 研究に必要なもの

- 意欲
  - 研究の準備・実行・発表には数年かかります
- 時間
  - 大学教員は忙しい、臨床医はもっと忙しい
  - 時間を作る努力
  - 日常業務に研究作業の一部でも組み入れ？
- 経験者
  - 特に初めての研究では必須
- (お金)

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

10

# 研究デザイン

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

11

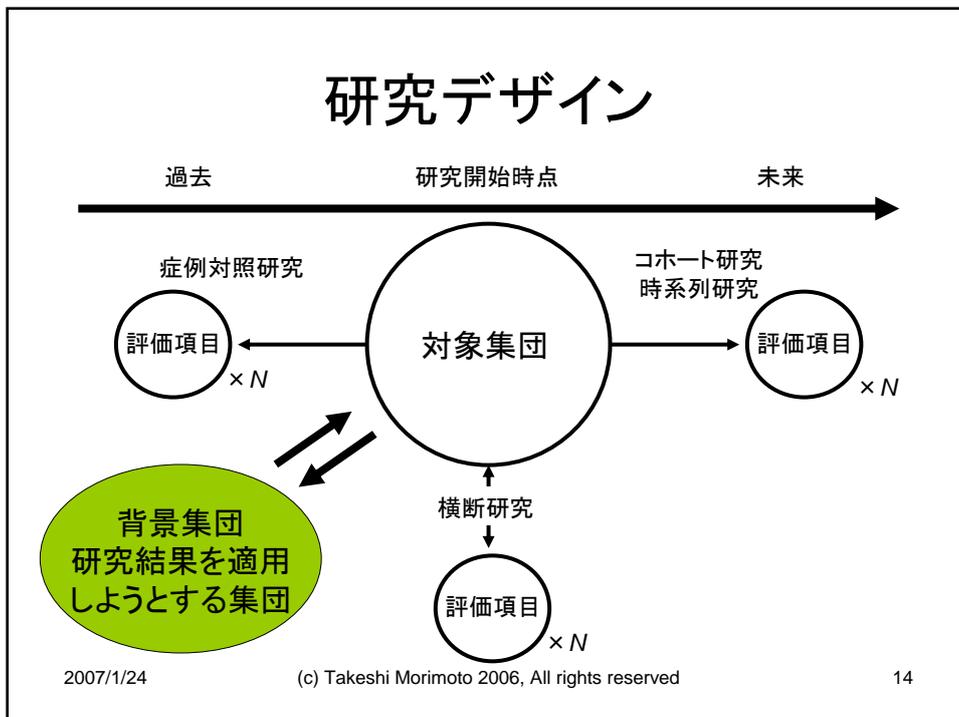
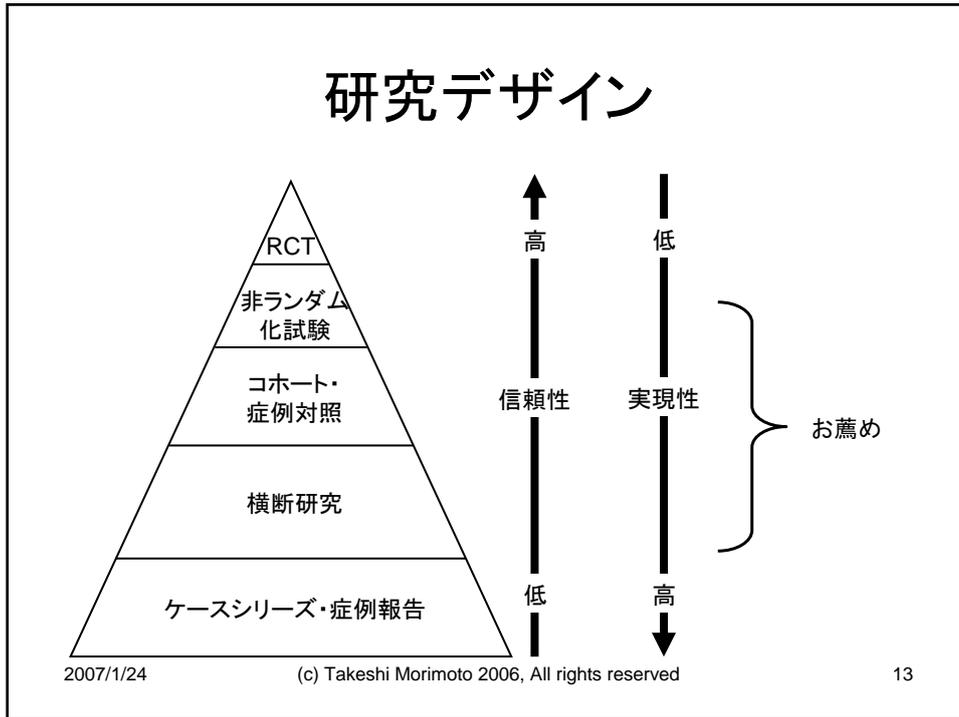
## 介入・因子の評価

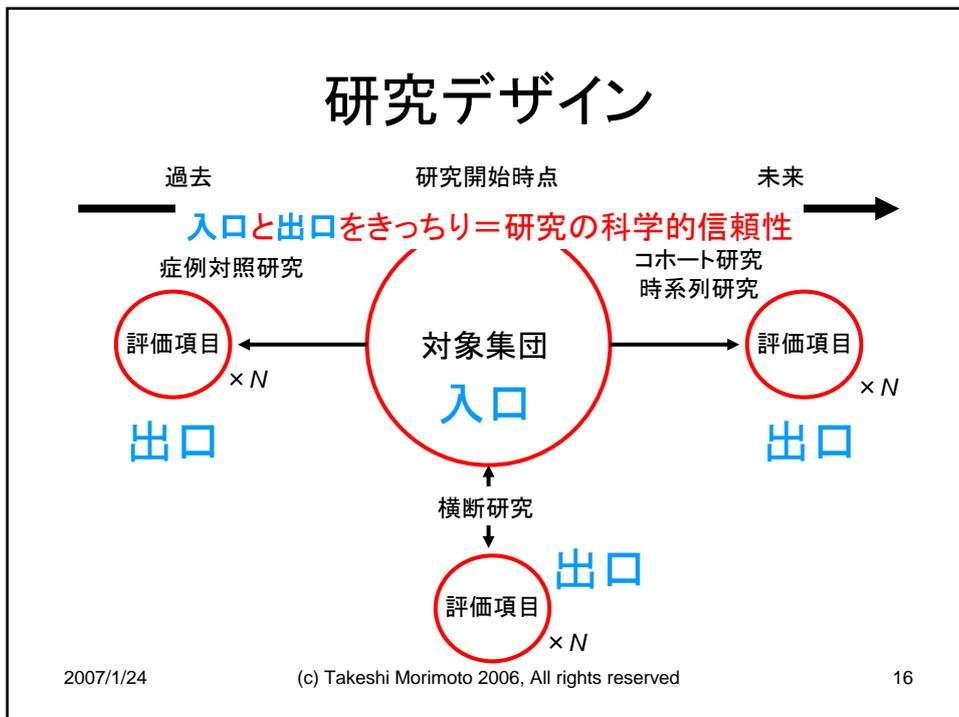
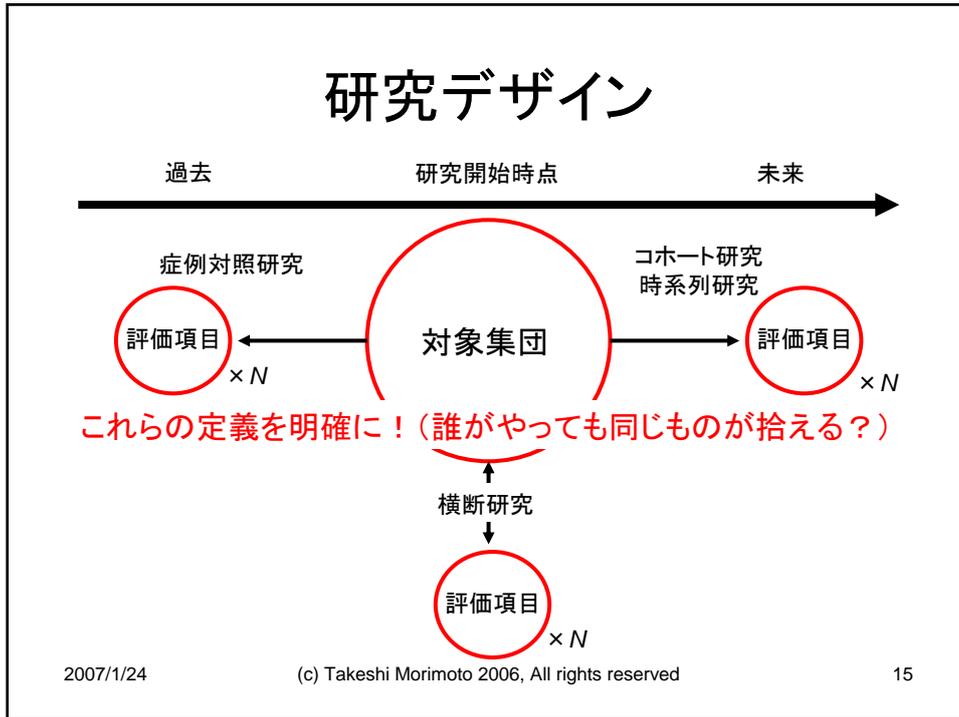
- 一発測定(横断研究)
  - 因子の評価は可能、因果関係は不明、介入不可
- 前後測定(コホート、時系列)
  - 変化が測定可能、因子の評価も可能
  - 介入の効果や因果関係は不明
- 対照群設定(RCT、非ランダム化)
  - 介入の効果の評価可能、因果関係も評価可能
  - 教育的介入をやらない群の設定が可能か？

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

12





## キークエスチョン

- 1) 誰が研究しても、同じ対象集団が採用されるか？
- 2) 誰が研究しても、同じ評価項目が同じように判定されるか

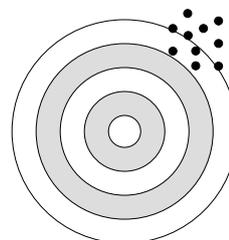
2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

17

## 判定の重要性

- 評価項目の信頼性、妥当性
- 評価者の信頼性、妥当性
- 信頼性: 評価項目の安定性
- 妥当性: 真の評価項目を  
きちんと反映



2007/1/24

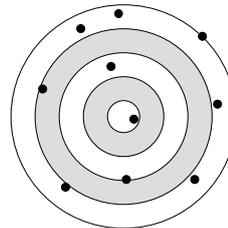
(c) Takeshi Morimoto 2006, All rights reserved

18

## 判定の重要性

- 評価項目の信頼性、妥当性
- 評価者の信頼性、妥当性

- 信頼性: 評価項目の安定性
- 妥当性: 真の評価項目を  
きちんと反映



2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

19

## 信頼性 reliability

- システミックなエラーであるバイアスとは違う
- ランダムなエラーの大きさを評価
- 信頼性の評価
  - テスト–再テスト→Correlation
  - テスト–他のテスト→Correlation
  - 内的統一性 internal consistency
    - クローンバッハ  $\alpha \geq 0.85$  (理想)、0.70 (最低)

(Nunnally JC 1967)

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

20

## 妥当性 validity

- 真に測定したいもの～抽象的、困難
- ゴールドスタンダードとの比較
  - 感度・特異度
- 内容から評価
  - エクスパートパネル
- 内的構造を評価
  - 因子分析

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

21

## 評価者の信頼性

- Correlation
  - 一致率
  - $\kappa$  値
- これらのお皿50枚は芸術的か?

大西助三郎

	優良	不可	計
優良	23	22	45
不可	2	3	5
計	25	25	50

森本  
格之  
進

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

22

## 一致率・κ 値

		Reviewer A			大西助三郎				
		○	×	計					
Reviewer B	○	A	B	W1	森 本 格 之 進	優良	23	22	45
	×	C	D	W2		不可	2	3	5
	計	Y1	Y2	N		計	25	25	50

一致率  $P_o = (A + D) / N$   
 偶然により、 $(W1 \times Y1 + W2 \times Y2) / N$ 人で判定が一致しうる  
 偶然による一致率  $P_e = (W1 \times Y1 + W2 \times Y2) / N^2$   
 κ 値: 偶然を超えての一致率  $(P_o - P_e)$ を、  
 得ることの可能な最大の一致  $(1 - P_e)$ で割ったもの

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

23

## 一致率・κ 値

		大西助三郎		
		優良	不可	計
森 本 格 之 進	優良	23	22	45
	不可	2	3	5
	計	25	25	50

一致率  
 $P_o = (A + D) / N$   
 $= (23 + 3) / 50 = 0.52$

偶然による一致率  
 $P_e = (W1 \times Y1 + W2 \times Y2) / N^2$   
 $= (45 \times 25 + 5 \times 25) / 50^2$   
 $= (1125 + 125) / 50^2$   
 $= 0.5$

κ 値  $= (0.52 - 0.5) / (1 - 0.5)$   
 $= 0.04$

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

24

## κ 値

- 3段階以上の順序変数でも可能
  - Weighted  $\kappa \hat{=}$  correlation
- Rule of thumb (Landis JR and Koch GG 1977)
  - 0.00 poor agreement
  - 0.01 – 0.20 slight agreement
  - 0.21 – 0.40 fair agreement
  - 0.41 – 0.60 moderate agreement
  - 0.61 – 0.80 substantial agreement
  - 0.81 – 1.00 almost perfect agreement

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

25

## 評価法・評価者の信頼性

- 同一者1回目 7/10 vs 2回目 5/10
- 評価者A 52点 vs 評価者B 60点
- Pearson correlation coefficient
- Intra-class correlation coefficient (ICC)
- Rule of thumb (Fleiss JL 1986)
  - 0.00 – 0.40 poor reliability
  - 0.40 – 0.75 fair to good reliability
  - 0.75 – 1.00 excellent reliability

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

26

## 現実的なこと

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

27

## プロトコル

- 臨床試験じゃないのなら、気楽に書こう
- 目的・仮説
- 背景
- 対象集団選択基準・除外基準(入口)
- 評価項目と評価法(出口)
- 標本数と研究期間～現実的
- 解析方法～記述統計(平均/頻度+信頼区間)
- 倫理委員会

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

28

## 解析方法の記載(例)

### 5. 解析の概要

#### 1) 主なエンドポイント

一般市民における学生や研修医による診療の頻度、内容、教員・上級医の立ち会いの頻度、満足度、阻害要因、希望

#### 2) 主な解析

一般市民における学生や研修医による診療の頻度、内容、教員・上級医の立ち会いの頻度、満足度、阻害要因、希望などについて記述統計(頻度、95%信頼区間)を行う。

学生や研修医による診療の経験に関するデータを潜在的説明変数としたロジスティックモデルを用いて、学生や研修医の診療を受け入れる態度に関連する因子を抽出する。

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

29

## インフォームド・コンセント

- 教育・診療上の必要性から**集積された**データ(入試成績・教科成績・共用試験結果・資格試験合否、問診票)を利用する場合には免除\*
- 教育・診療上、評価・改善のために**明確な理由**を持って**行われた**調査(授業評価、患者満足度調査)を利用する場合には免除\*
- 研究のために、新たなアンケートや介入を**行う**場合は必要
- 各施設の倫理委員会の判断に準拠
  - \* 個人情報 that 特定できる場合には議論あり

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

30

## 倫理委員会・IRB

- 脆弱集団vulnerable populationの保護
  - 教員が学生に行うアンケートには無言の強制力
  - 指導医が研修医に行う介入・アンケートにも圧力
  - ⇔ 関係ない組織から頼まれたアンケート～無視
- 論文化の際のバリアー・保険
- 研究デザインの確認、問題点抽出
- 例外なく倫理委員会の承認を得る

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

31

## データセットの作り方

- エクセルで十分(勿論、アクセスはベター)
- できるだけ、業務の合間にreal timeで
- アンケート・調査票=>データセット～手伝い
- 独立研究IDを～同じ標本が2-3度登録もあり
- エクセルの縦(行)に患者を、横(列)に変数
- テキスト記載は最小限に～別に列を作る
- できるだけ観察値もしくはあり・なしで登録
- 全角の数字や記号・スペースはトラブルの元

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

32

## 世界3大宗教

- 「連続変数」教
  - 神経ステーション 22/37 (59%)
  - 自己評価 8/10
  - 参加日数 21日
- 「平均値±SE」教
  - 平均自己評価点数  $3.7 \pm 0.8$  (5点満点) SDは2.8
- 「多変量解析」教
  - 多変量オッズ比 1.0002 ( $p < 0.05$ )

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

33

## 変数

- 連続変数
  - 正規分布
  - 非正規分布
  - イベントまでの時間 (censored)
- 順序変数
- 名義変数 ( $\geq 3$ 群)
- 2元変数

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

34

## 連続変数

- “測定”のデータの多くはこの形
- Pros
  - 適切に処理すれば、事象を鋭敏に表現
  - 統計学的検出力も高い
- Cons
  - 欠損値が発生しやすい→多変量に不利
- コツ
  - 分布と欠損値の頻度を常にチェック
  - 数学的処理、順序化、2元化を考慮

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

35

## 順序変数

- 連続変数と2元変数の中間
- 一見連続変数
- Pros
  - 事象を2元変数よりも鋭敏に検出
- Cons
  - 欠損値が発生しやすい
- コツ
  - 理由をつけて閾値を設定して、2元化

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

36

## 名義変数(≧3群)

- 順序化できない対象集団の特徴
  - 出身校
- Cons
  - 基本的に数字変換は不可能
  - 統計処理が困難
- コツ
  - 医学的標準、最大標本数の群を標準としてダミーコード化

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

37

## ダミーコード化

標本ID	出身校	Private	National	Oversea
1	公立高校	0	0	0
2	私立高校	1	0	0
3	国立付属	0	1	0
4	帰国子女	0	0	1

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

38

## 2元変数

- イベントの評価
- 他のタイプの変数の変換型
- Pros
  - 欠損値が少ない ~ 欠損値=“なし”
  - 解析や結果の解釈が容易
- Cons
  - 鋭敏さに欠ける
- コツ
  - 2元変数の組み合わせ~元々ダミー

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

39

## メッセージ

- 医学教育研究はアンケートを集めて、t検定して終わりじゃないのよ
- データをとる前に十分検討を
- 新しい知見は不要、something addで十分
- コホート・横断研究から
- 入口(対象集団)と出口(評価項目)をきっちり
- 評価項目・評価者の信頼性・妥当性にも配慮
- 標本数・プロトコルは現実的に、IRBは必須

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

40

## グループ討論

- 各グループで持ち寄った研究テーマについて討論し、一つの研究テーマについてPPTを使って発表してください
  - 研究テーマ
  - デザイン
  - 研究対象、施設、人数
  - 評価項目、評価法
  - 解析の方向性

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

41

## お時間

- グループ討論
  - 30分 ブレインストーミング
  - 45分 研究計画
  - 15分 発表の準備
- 発表(5分)
  - 研究テーマ
  - デザイン
  - 研究対象、施設、人数
  - 評価項目、評価法
  - 解析の方向性

2007/1/24

(c) Takeshi Morimoto 2006, All rights reserved

42

## ブレインストーミングのネタ

- 医療者として不適当な学生の入試における因子
- OSCE結果のクリニカル・クラークシップへの応用法
- Early exposureの導入と進路との関係
- 細胞生物学講義で、学生の理解を上げる介入
- 在学時の成績と入職後のバーンアウトの関連
- 臨床英語能力を上げる介入
- CAI (computer-assisted instruction) の受容性
- 研修医の症例提示能力の客観的評価法

講演: 森本先生ーリサーチクエスチョンから研究計画へ

それでは、今日は夜までやりますし、お疲れもありますので、ジャケットは脱ぐようにしませんか、寒くなければネクタイも外しましょうか。京都大学では指導者講習会をするときにはネクタイは厳禁となっております。着用してきたら、皆さん外してもらうようにしております。実は、ここで見ても分かるように、タスクサイドも皆さんネクタイをしていません。ワークショップであり、皆さんで知恵出しながらわいわいと作業するので、できるだけラフな格好でお願いします。後ろにドリンクとかお菓子とかは、タスクの杉本先生が用意して下さっています。今日と明日ではお菓子の内容が違うということですので、今日の分は食べて頂いた方がいいということです。

なかなか秋は学会の抄録とか発表のシーズンで、私みたいな統計屋さんはあちこちで解析をしている関係もあり、昨日もひたすらいろんな解析や準備などをしていたこともあって、今日のスライドができたのは、品川あたりでした。明日皆さんに使うって頂く仮想データは今、作っている最中で、先程から机でちょこちょこ作業しているのは仮想データの作製作業でして・・・(画面を指しながら)今まだ仮想データを一生懸命作っているのですが、こちらは消しておいて・・・こっちですね・・・はい。これから30分ですごく駆け足で、本当は1週間分くらいかかるような事を話しますので、前だけ向いていただけたらいいです。配付資料で同じものもありますし、本当は終わってから配ろうと思っていくくらいです。あまりにも薄く、冊子にしても仕方がない気もしますが、気を利かして挟んでくれたと思うんですけど、まあ前だけ向いて下さい。後で、皆さんでグループワークをするときにでも参考にさせていただいたら結構です。

SLIDEー2 アウトラインを話します。医学教育研究は、この中でもたぶんされている方も多いと思いますが、私自身は新参者です。まず、始める前の準備としてはデザイン、まず対象集団の設定そして評価項目、すなわち介入・因子・結果の設定です。僕はこれらを入口と出口と言っているんですけど、入口と出口だけでも少なくともきっちりしておけば、相当信頼性の高い研究ができるんですね。入口・出口が大事。それから評価項目の信頼性があって、これはこの後の大西先生の統計のところ少し出てくると思います。そして、最近流行のプロトコルとインフォームド・コンセント、IRBの話です。時間がもし余れば、変数の種類の話はしますが、多分そこまでは行かないと思います。

SLIDEー3 医学教育研究をすでにされている方も、多分多いと思います。基本的には、何でもいいです。いろんな職種やレベルを対象集団としますね。結構『医学教育』の雑誌を見ると学生の評価あるいは研修医の評価とかが多く、基本的に対象集団は医師、看護師、薬剤師、技師、コメディカル、教員の評価がありますが、他にはもちろん学生や医師のなかでも医学生であったり、研修医であったり、スタッフであったり、といろいろなレベルがありますね、種類には人間だけじゃなくてですね、システムであるとか、病院であるとか、患者さんもですね、医学研究の対象になります。それから内容ですが、内容というのは具体的な評価をするものですね。評価

項目 measurements と言いますが、簡単なものは、学業成績とか知識ですが、技術や態度といった少し掴みどころのないものを対象とすることもあります。また、教育的な介入であるとか、いろんなファクター、すなわち因子との関連、時間的変化を見ることもあります。結構、技術とか態度って、あまり掴み所がないものを研究する点が、少し臨床研究とは違いますね。臨床研究では心筋梗塞とか死亡などを評価していけばいいのですが、そこあたりのところは、少し信頼性のところで問題になりますね。

SLIDE-4 これから先しばらく、同じ用語を使いますが、このスライドで説明します。Measurements, 評価項目ですね。評価する内容です。これは研究者が介入や因子の与える影響をみたい指標、例えば、成績や国試の可否、態度、技能というところになります。介入とは研究者が評価を目的として、導入したり制御する因子であり、特別な講義をしたりとか、小テストをしたりとか、ローテーションの内容とか、研修先とかですね。因子ですが、これは集団の中に偶然にばらついて存在し、評価項目に影響を与える要素、例えば出身校や男女、出席率というものが入ってきます。これを使いながら、説明していきます。

SLIDE-5 医学教育研究の論文や発表がどういうところに出ているかと、いうことですが、まず学会について日本の場合は医学教育学会、この前の7月に奈良でありましたけど、それ以外では、私はあまり詳しくなく、大西先生が知っておられるけど、アメリカでは Association of American Medical Colleges とか、ヨーロッパでは Association of Medical Education in Europe です。私がよく行くのは Society of General internal Medicine, 一般内科の学会なんですけれども、一般内科は日本での総合診療科というところに相当し、アメリカでも、ほとんどの場合教育担当者と同じ場合が多く、医学教育についてもよく扱われています。次に雑誌ですが、Medical Education, Academic Medicine, Medical Teacher がありますね。一般内科のところでは Journal of General Internal Medicine は時々 Education の特集も組んだりして、比較的熱心です。まあ、私もそんなに目を通してはいるわけではないのですが、こういう雑誌をばらばらと目を通しておくと、どのようなものが医学教育に関する研究か、というのが分かってきます。勿論それぞれの subspecialty, 例えば neurology とか, surgery とか, そういった subspecialty の雑誌でもいろんな医学教育に関するトピックスも扱われますし、BMJ とかでも扱われていますね。

SLIDE-6 さて準備ですけれども、

SLIDE-7 研究を始める前に考えることですが、まずテーマは単純に、セッティングや対象集団、介入因子、評価項目を明確にします。また、意味があるか、既に答えがあるか、などを検討します。皆さん基礎研究とかで、他の研究をされた経験は多分あるんだろうと思いますが、同じことです。同僚などに聞いてみる、教科書やレビュー、MEDLINE を確認するのも、他の基礎研究でも臨床研究でもあまり変わりません。似たような研究があったら？別にいいんです。OKで

す。

SLIDE-8 テーマを明確にするのですが、たとえば小試験によって医学生の学習意欲が高くなるか？よりも、毎講義の最後に行われる小試験は医学部4年生の臨床系授業において質問紙法による講義への学習意欲を高くするのか、というふう出来れば、具体的な研究デザインが組めます。ただポイントは、あまり理想的なことを追求してもなかなか仕事できません。場合によっては現実を優先したらいいと思います。勿論、臨床試験の際にはかなりデザインを詳細に組むんですけど、ま、我々の現状からすると自校の教育環境に合わせることも重要だと思います。

SLIDE-9 研究の意義についてはやっぱり、同僚や若手、学生、上司などに聞いてみることです。同僚や若手はデータの集積に協力してくれるかもしれません。上司なんかは研究の便宜を図ってくれるかもしれません。また、お金を持ってきてくれるかもしれません。過去に似た研究があるという場合、内容が全く同じような研究では確かに魅力は下がりますけれど、例えば対象集団や評価項目等、全部が一緒かどうかですね。授業科目や教育環境の違いが影響しそうか、何か小さいことでも付け加える価値はありそうか、something add といいますけれど、もしそれであればリサーチとしては十分だと思いますね。別に全く innovate な事、新しいことを目指す必要はないと思います。

SLIDE-10 研究に必要なものですが、まず意欲です。ここにいらっしゃる方は殆ど皆さん極めて selected な集団ですけれども、研究の準備から発表までは数年かかることを最初に覚悟しなければならぬので、なかなか1年後に全部結果まで出してしまうと、動物実験実験になってしまいます。時間ですが、大学教員は忙しい、臨床医はもっと忙しいですね。できるだけ時間を作る努力は当たり前ですけれども、できるだけ日常業務に研究内容の一部でも組み入れよう、と私はいつも思っています。それから経験者は募った方がいいですね。特に初めての研究では経験者は必須だと思います。一人でやると大体失敗します。それから、場合によりますが、お金が必要な場合もあります。この辺はその環境や研究内容との関係で決めて頂いて結構だと思います。

SLIDE-11 研究デザインに入ります。あの時間が非常に押しているので、ちょっと駆け足でいきますので、質問はまたあとでグループワークの時に回りますので、もし質問があれば、手元に書き残しておいて頂ければ、個別に対応しますし、後でまとめて答えたりします。

SLIDE-12 介入や因子の評価法です。いろんな評価法がありますが、一発測定、すなわち横断研究ですが、一回だけ学生100人とか研修医50人を対象にばあっと調査してしまうんですね。基本的に因子の評価項目同士、これとこれとかの関係性は見られますけれど、因果関係といいますが、何が卵で何が鳥かというのは不明です。勿論介入することはできません。因子との

関連性をみるということです。しかし、初心者にとって研究のスタートポイントになります。次は、前後測定ですね。いわゆるコホート研究や、時系列研究、これは1回2回3回と測っていくやり方です。変化の測定は可能です。それから因子の評価も可能です。しかし、介入の効果や因果関係は不明です。具体的に言いますと、今日ある評価をします。そして次の日から2ヶ月間ある介入をして、終わった後にまた同じ評価をします。さあ、このときの変化はこの介入の影響でしょうか？なかなか難しいですね。全く違うファクターが間に入ってくることもあります。それから単に時間だけで変化することもあります。それから、Hawthorn 効果といって、人から観察されている、研究対象となっているだけで、パフォーマンスがよくなることがあります。最後に、対照群を設定する研究です。RCTはよく聞きますね。これはランダム化を行うものですね。また、非ランダム化で行うこともできます。介入の効果は評価可能ですし、因果関係も評価可能です。ただ、実際の医学教育研究で何が大変かという、教育的介入をやらない群の設定が可能かというところがいつもいつも問題になります。もちろん、いくつかを組み合わせで、前後測定に介入と非介入を加えたりとか、もう少しひねってクロスオーバーしてみるとか色々ありますが、基本的にはこの3つの枠です、いずれも介入や因子を評価するという作業になります。

SLIDE-13 これは臨床研究にたくさん出てくる研究デザインですね。一番上がRCTで、その下に非ランダム化試験、コホート、症例対照、横断研究、ケースシリーズ・症例報告と続きます。信頼性は一番上が高いんですけど、実現性は非常に低いです。サンプリングも大変です。実現性から考えると、僕はいつも若い人に研究を指導するときには、ここからスタートしてくださいと言っています。このあたりを「お薦め」と書いていますが、このあたりで研究を考えておられれば現実的だと思いますね。

SLIDE-14 研究デザインですが、これは研究デザインについて全く初めての方を対象に少し書いていますけど、まず背景集団ですが、研究結果を適用しようとする集団である背景集団から我々は研究対象集団をサンプルしています。ここがですね、随分違っていると、違っても仕方がないんですけど、違っているということを認識しなきゃいけないのです。もし違っている場合は、後でその研究の結果を適用するときに、どういう差があるのかということをちゃんと明らかにしなきゃいけないのです。従って、基本的に背景集団 background population との自分の研究対象集団 sample との差、違いというのを常に意識します。そして対象集団について、未来に向かって時間を追って見ながら、いろんな評価項目を集めると、いろいろな結果がたくさん出るんですね。これをコホートと言います。また、その時点でいろんな別の評価項目を同時にたくさん測れば横断研究になります。それに対して、過去のデータを遡って行けば、症例対照研究になります。基本的には、いろんな評価項目について未来や過去または同じ時点での評価項目との関連を見ていくというのが殆どの研究デザインになります。

SLIDE-15 これらの研究で、特に大事なのはこれですね、対象集団。それから評価項目です。こ

れらを常にきちんと明確にして下さい。誰がやっても同じものが拾える、たとえば誰かがプロトコル書いて、私が今日研究したものと、次の日に大西先生が研究したものが大体同じになるような、すなわち誰がやっても、対象集団も評価項目も同じものが拾えないとそれは極めて不安定な研究になります。

SLIDE-16 対象集団は入口ですね。とにかく入口と出口をですね、きちっと抑えておけば、かなり科学的になり、信頼性は高くなる。これが誰がやっても同じになる、もちろん、調査の回答の中で学生がふざけたりとか、いい加減なことを書くとか、いろんなことが出てくるんですけども、少なくとも入口・出口をきちっと抑えておけば、相当高い信頼性が得られます。

SLIDE-17 キークエスチョンですけど。誰が研究しても、同じ対象集団が採用されるか？というのが一つと、もう一つは、誰が研究しても、同じ評価項目が同じように判定されるか、というこの2つについて、常に気を配って下さい。

SLIDE-18~19 さて判定の重要性です。評価項目、たとえば知識を見よう、例えば態度を見よう、技能を見よう、という時に、評価項目の信頼性や妥当性を考慮します。これは出口の考え方で、非常に大事です。それから評価者も同様に、ある評価項目があつて、例えば質問紙とかアンケート、OSCE のスコアなどの評価項目を評価する評価者は、OSCE でも既にやっておられますけど、必ず複数の評価者でやっていますよね。この2つのレベルにおける評価の信頼性・妥当性は、とくに医学教育研究では、臨床研究や動物実験、基礎研究よりも重要になってきます。信頼性というのは評価項目の安定性であり、妥当性というのは真の評価項目、例えば本当に能力の高い学生をちゃんと高いと評価できているのかどうかということです。少し言葉だけでは分かりにくいので、こういう的に矢を当てることを考えます。的のど真ん中が真の評価内容と思って下さい。ここを狙いたいと大西先生が鉄砲を撃ちます。ばあん、ばんばんですね・・・おっなかなか、いかがですか？これは信頼性が高いと思われる人？では妥当性が高いと思う人？もう一回いきます。次は杉本先生が鉄砲を撃ちました。ばあん、ばんばんと鉄砲を撃ちました。こんな形ですね。大分ばらけています。これはバラけてはいるんだけども、どうもこれは真ん中を狙っているような感じですね。だから信頼性は高くないですが、妥当なものだろうということになります。逆に、大西先生の鉄砲はですね、非常に狭いところに集中して当たっていますよね。バラつきは少ないです。こういうのは信頼性は高いのですが、実は本来の的とは全然違うところを狙っているんですね。こういうのはあまり妥当性は高くないんだけども信頼性は高い、という話になります。

SLIDE-20 信頼性についてですが、信頼性というのはシステミックなエラーであるバイアスとは違います。システミックなエラーとは、例えば全て右に片寄ってしまっている、こういうのはバイアスと言います。信頼性とは先ほどの的で言うと、ランダムなエラーの大きさを評価します。信頼

性の評価法にはいくつかあって、全く同じものを再テストします。勿論 response shift といって測定物が増えることがないという前提ですが、それで correlation 相関をみるんですね。また、他の関連のあるテストとの相関を見る方法があります。それから、機械的に内的統一性を見る internal consistency という方法があって、評価項目の内部で、お互いに項目同士の相関を見ます。そのとき、クローンバッハ  $\alpha$  で判定しますが、例えば 0.85 は理想、0.70 はギリギリと判断したりしますが、この基準は大分古い概念ではあります。

SLIDE-21 次に妥当性についてですが、本当に狙いたいものを見ようとしているかということが妥当性です。真に測定したいものはほとんどの場合、抽象的で測定困難であり、実際に測定したものは、その代用であることが多いのです。もしゴールドスタンダードがあれば比較すればよく、感度や特異度で評価できます。ゴールドスタンダードと比較するくらいなら、最初からゴールドスタンダードを測ればいいじゃんという話になりますが、たとえばゴールドスタンダードを測定するのに、大変お金がかかって、大変時間がかかって、大変複雑でという場合、片方は質問紙で簡単に評価できるんじゃないかというものがある場合は、ゴールドスタンダードを使って、少しパイロットのデータで感度と特異度を測って、ある程度妥当だと思えば、そっち、簡単な方法を日常的に使えばいいんじゃないかなということになります。それから、質問紙なんかで、内容から評価することがあります。エキスパートパネルで、きちんと見たいものが見られているかというのを評価します。また、統計学的に内的な構造を評価する方法で、因子分析というのがありますが、時間の関係で、これらの話はこのワークショップの範疇からは外れます。

SLIDE-22 評価者の信頼性についてですが、実技試験なんかで2人3人の評価者がいる場合の信頼性の評価の方法にはいくつかあります。Correlation や一致率、 $\kappa$  値という方法があって、Correlation の話はあとで話しますが、比較的初めての方が分かり易いのは  $\kappa$  値だと思いますので、ここで少しやってみます。ここにお皿があります。これが芸術的だと思う人？あまりそうではないと思う人？まあ、そんなものですかね。たとえば、このようなお皿が50枚あって、それぞれについて芸術的かどうかということが問題になっているとします。それを大西助三郎と森本格之進という2人が評価することにします。大西助三郎は、この皿50枚のうち、25枚が優良で、25枚は不可。森本格之進は45枚はOKと評価した、ということにします。こういうデータがあるときに、この2人の評価法は信頼性が高いでしょうか？信頼性が高いと思われる人？これではあかんという人？皆さん、いいセンスをしていますね。ちなみにですね、これは私が造ったお皿です(笑)。つい最近ですが、 $\kappa$  値を説明するためにだけに作ったんですよ(笑)。

SLIDE-23 一致率についてですが、テキストを読んで頂いたら分かりますけど、reviewer が2人で、こういう形で分かっているとします。一致率というのは単に2人が合ったところを足して全体で割ったものが一般的な一致率です。ただ、偶然にN人中の  $W1 \times Y1$  人と  $W2 \times Y2$  人は一致します。これはそれぞれの期待値ですよ。期待値は端の変数と変数をかけたものを全体で

割ったものです。期待値を足したものをもう 1 回全体で割れば偶然の一致率が出ます。一方、 $\kappa$  値というのは観察された一致率  $P_o$  から偶然に起こった一致率  $P_e$  を引いたものを、得ることができる最大の一致で割ったものです。実はこの例で計算してもらおうと思っていたのですが、基本的に時間が押しているのと、既にもうスライドを配ってしまったので、ここでは計算はしません。

SLIDE-24 この例での一致率は、50%くらいしか一致していませんね。半分は当たっているんですけど。偶然の一致を計算すると、偶然が結構多いですね 0.5 です。半分は偶然に一致するという結果ですね。半々ですね。で、これで  $\kappa$  を計算すると 0.04 となります。この 2 人の一致はほとんど偶然の結果という話になります。お皿が悪いのか、または 2 人の仲が悪いのは分かりませんが、このお皿は 2 人では評価しにくいという話になります。

SLIDE-25  $\kappa$  値についてですが、一般的には 2 変数について評価します。これは臨床試験、臨床研究に多いスタイルですが、2 つ以上の、例えば 3 段階以上の順序変数についても可能です。順序変数で離れたものにウェイトをかけることで、Weighted  $\kappa$  値が出ますが、これは correlation 相関と非常に近いものになってきます。おおざっぱに考えると、0.04 というのは、ちょっとだけ agreement がありますけど、ほとんど agreement がないと判断されます。ちょっと古典的になりますが、大体 0.6 とか、0.8 とかのあたりで評価して下さいということになりますが、お手元の資料を参考にしてください。

SLIDE-26 他に評価法について重要なことですが、評価者の信頼性があります。例えば同じ評価者が全く同じものを評価して 1 回目はこうで、2 回目はこうだった、という場合。それから、ある人についての評価が、評価者 A は 52 点、別の評価者 B が評価したら 60 点とか言う場合があります。OSCE などでありますよね。一つの方法は Pearson の correlation ですが、これは皆さんも動物実験なんかでやられたたことがあるのではないのでしょうか。これは X 軸と Y 軸の座標上に点を打って、その上に線を引くやつですね。Pearson の correlation を使うことが多いですね。それから、こういう人の判定の領域では Intra-class correlation という概念があります。これは ANOVA と言って、大西先生この後話しますが、その表を使って計算できます。この Pearson と Intra-class correlation はものすごく似ているので、まああまり気にせず、こっち、下がですね、苦手であれば上で評価されても、ま、最低レベルのところはクリアーしています。少なくとも信頼性の評価はしないよりした方が明らかにいいので、簡単なものでもいいから、した方がいいというのが私のメッセージです。これはですね、Rule of thumb で 1986 年のもので、ちょっと古いですけど、相関係数についてのガイドです。Pearson が 0.4 以下ではちょっと待つてよ、という話になります。0.75 から上はよい、というふうに言われています。

SLIDE-27 あと残り 5 分ですね。少し現実的な話をします。ちょっと変則でいきます。

SLIDE-28 プロトコルについてです。基本的に臨床試験ではないので、気楽に考えてもらって結構です。薬とかの臨床試験は人に有害を及ぼしますので、非常に慎重にプロトコルを作成しますが、ま、教育研究は少々ミスっても学生は死ぬわけではないので、気楽に書いていいと思います。まず目的、仮説を書きます。それから背景を書きますね。次に対象集団の選択基準を書きます。それに除外基準を明記します。先程説明した、入口を明確にします。それから評価項目、評価法を書きます。出口に当たりますね。何を測るのか、どうやって測るのかを決めて、出口を明確にします。結構、標本数とか研究期間も問題になることも多いですが、臨床試験だとかなり綿密にやりますが、教育研究ではサンプル数については基本的に現実に即して書いてもらって結構です。これこれ、こういう理由で、うちは100人にしました、とか書けばいいです。サンプル数を計算して、一学年500人のデータが必要となっても、それだけのデータが取れないのですから話になりません。基本的には現実的にことを書いてもらって結構です。解析方法もシンプルでいいです。いわゆる記述統計ですね。平均であるとか、頻度とか、信頼区間とかです。せいぜいそれに一行くらい多変量解析を足す程度でいいと思います。それから後でもう一回言いますが、倫理委員会を意識して書いて下さい。

SLIDE-29 これは解析法の記載例ですけども、私が最近倫理委員会に出したものです。一般市民における学生や研修医による診療の頻度、内容、教員、上級医の立ち会いの頻度、満足度、阻害要因、希望がエンドポイントになります。主な解析については、これこれこれについて記述統計、頻度、95%信頼区間を出します。また、これらのデータを潜在的説明変数としたロジスティックモデルを用いて、学生や研修医の診療を受け入れる態度に関連する因子を抽出する。たったこれだけです。これで京都大学の倫理委員会はすぐに通りました。極めて楽でした。

SLIDE-30 インフォームド・コンセントについてです。一般的に、教育上、診療上の必要性から普段集積されたデータ、例えば入試成績や教科成績、共用試験の結果、資格試験の可否、問診票についてはインフォームド・コンセントをとることは免除ですね。確かに共用試験をするために、学生から一々インフォームド・コンセントをとっていませんよね。ですが、アスタリスクをつけているのは少し条件があるということで下に書いています。それから教育上、診療上の評価や改善のために明確な理由を持って行われた調査、たとえば授業評価ですが、授業が終わった後に評価してもらっていますが、いちいちインフォームド・コンセントをとっていません。患者満足度調査でも免除されていますね。しかし、研究を目的とした新たなアンケートや介入を行う場合は必要です。これこれの研究でこういうことをしたいので、という場合には必要です。これは対象が学生であれ何であれ必要ですね、基本的には各施設の倫理委員会に準拠してやってください。あと、個人情報ですね。個人情報についてはやっぱり、それが特定できるようなデータを扱う場合には、もう一度インフォームド・コンセントを取り直す、という議論もあります。

SLIDE-31 次に、倫理委員会についてです。日本では倫理委員会と言いますが、外国では IRB Institutional Review Board と言って、全然やっぱり訳が違うんですよ。その辺はあんまり気にせず論文で書くときは Institutional Review Board とか書いて、日本語では倫理委員会と書くんですね。基本的に倫理委員会の目的は脆弱集団 vulnerable population の保護ということです。基本的に人を対象にする全ての研究には必要です。例えば、教員が行う学生のアンケートには基本的に無言の強制力がかかっています。指導医や研修医に行う介入のアンケートにも圧力がかかっています。彼らを保護するために、倫理委員会を通るのが基本路線になります。例えば、その逆を考えればいいですね。関係のない組織から頼まれたアンケートには研修医はほとんど無視しますよね。配られても回答0とかね。なぜかという、基本的に上司が行うアンケートは圧力がかかっているということです。倫理委員会の目的は、こういう集団を保護すること、それからもう一つは論文が審査される場合のバリアー、保険ですね。最近では、こういう vulnerable population に対してなんらかの介入があった場合は、IRB 審査の有無が問われます。IRB に出すもう一つの目的は、研究デザインの確認ですね。少なくとも、問題点の抽出にはなりません。結論ですが、医学教育研究であろうと何であろうが、例外なく倫理委員会の承認を得てください。というのがメッセージです。

SLIDE-32 データセットの作り方という話しですが、これはもう、ぱっぱぱっと手元の資料を見てください。

SLIDE-33 わが国の医学研究3大宗教は、キリスト教イスラム教仏教でなくてこういう宗教があつてですね、手元の資料を見ていただいたら結構です。(ちよん、ちよん、ちよん、ちよん、ちよん、ちよん、ほい・・・)

SLIDE-40 最後ですが、医学教育研究はアンケートをとってt検定で終わりじゃないのよという話でした。データをとる前に十分検討を、それから something add ですね。また入口、出口、信頼性、妥当性を十分に検討し、かつ現実的に、最後に IRB ですね。

では、グループ討論を今から始めます。今から1時間あります。まずグループの中で、普段、自分の日常の教育業務の中から、テーマを持ち寄って来て頂いていると思いますので、それをこれから討論して下さい。そして、そのいくつかのテーマの中の一つをパワーポイントを使って発表して頂く作業をします。発表ですが、研究テーマ、これまでどんなことをしてきた、またこれからどんな研究をしたい。デザインに加えて、研究対象や、施設、人数ですね。それから、どういう評価項目を作りたい、また評価法はどうするのか。また、解析の方向ですね。どんな解析をしようとするのか、非常に primitive なもので結構です。それらをたたき台として皆さんで勉強というか、学習して下さい。それから過去にやられてないかどうかというのは MEDLINE とかを見てもらったらいいんですけど、

ここでは、そこまでは気にせず、頭の中にあるものだけでやってください。お時間ですが、今からグループ討論 30 分で、ブレインストーミングをしてみてください。それから 45 分間くらいで研究計画の打ち合わせで、1つのテーマについて研究計画を作成して下さい。それから、15 分で発表の準備をして下さい。発表は5分で、今言った研究テーマ、デザイン、研究対象、施設、人数、評価項目、評価法、解析の方向性について発表して下さい。ええと、各委員が、うろうろと皆さんのところに行くので、適当に捉まえてですね、どうするんや、というふうにしめつけていただいたら結構でございます。私はその辺で一生懸命明日のデータを作っておりますが、必要でしたら遠慮なくお声かけしていただければ結構です。よろしいでしょうか。ちょっとブレインストーミングのネタですけど、なにせ、これはさっきですね、12 時 50 分くらいに、大西先生が「どうしようこれ」とか言い出してですね、一生懸命、今私が 15 分か 30 分くらい考えたネタですので、基本的にあまり valid じゃないです。勿論、妥当性かどうか分かりません。基本的に、医療職になるのに不適當な学生の存在に我々のところも困っております。OSCE の結果のクリクラへの応用法、early exposure 導入と進路の関係、細胞生物学の学生の理解を上げる介入、いろいろ介入とかですね、私には無理ですが。あと、在学時の成績と入職後のバーンアウトの関係とか、臨床で使える英語能力を上げる介入とか、今流行の computer-assisted instruction の受容性とか、研修医の症例提示能力の客観的評価法とか。ま、適当にこれらはネタというか、こんな感じで皆さん、わいわいと仲良くやって頂いて結構です。さて housekeeping な質問ございますか？あの内容についてはまた適当に捉まえて頂いて、私そこにありますから、直接本とかを持ってきて、これどういうこっちゃ、と聞いて頂いて結構です。何かスタッフからありますか？(発表順番とか・・・)ああ発表順番、あとで乱数表を作っておきます。(笑)常に乱数表ですよ、人生は。いかがでしょうか？発表時間は5分？ディスカッションは1時間？45分？まあいろいろあります。はい、では皆さんやってください、ええとそこにプロジェクターがありますので、コンピュータ持って来て下さいと言っていますよね？(コンピュータがないグループとかありますか？)各グループ1台もないところはないですね？大丈夫？じゃそれをお願いします。

# 医学教育誌編集委員会が 目指す研究論文のあり方

日本医学教育学会

編集委員

福島 統

2006年10月15日第2回医学教育研究技法ワークショップ

## 第13期 日本医学教育学会編集委員会

委員長 : 鈴木淳一

副委員長: 畑尾正彦、齋藤宣彦

委員 : 相澤好治、福島 統、森田孝夫

松岡 健(32巻5号から)、庄司進一(33巻2

号から)、高木 康、吉岡俊正(33巻5号から)

顧問 : 牛場大蔵、堀 原一、尾島昭次

- 第13期以前は、運営委員は編集委員を兼ねていた。→  
運営委員会と編集委員会の分離

## 31巻2号(2000年4月発行)

編集後記から(鈴木淳一編集委員長)

「筆者は図らずも編集委員長に返り咲くことになった。尾島昭次新学会長のたってのご希望に沿ったのであるが、2人の副委員長、そして幸いにも、3人の新鋭編集委員の皆様にご協力いただくことになった。・・・(中略)・・・新組織による編集委員と協力して新しい編集を実行していきたい。レフェリーは、今後、学会会員のすべての方々にお願いすることになった。・・・」

## 特集

1. モデル・コア・カリキュラムと共用試験(33巻2号、2002年4月)
2. 大学外卒前医学教育:地域での医学教育  
Community-Based Medical Education(34巻3号、2003年6月)

## 34巻4号(2003年8月)

### ■ 編集後記から(鈴木淳一編集委員長)

私どもが編集を担当して早3年半がたった。..われわれの編集委員会のスタートには、2つの決断があった。今なら書いてもよいと思うが、篠原出版の困難があった。...編集委員は学会発足以来ずっと運営委員が兼任していたのを止め、編集担当にふさわしい方々を選任した。...熱心な討論、厳しい査読で投稿会員には苦情を貰うこともあったが、多くの熱心な査読者には御礼の申し上げようもない。

## 第14期 日本医学教育学会 編集・出版・HP委員会

委員長 : 庄司進一

副委員長: 堀内三郎

委員 : 相澤好治、福島 統、森田孝夫、  
松岡 健、高木 康、吉岡俊正、椎橋  
実智男(35巻2号から)

顧問 : 堀 原一、尾島昭次、齋藤宣彦

## 第14期の編集

- 34巻から、学会抄録集は補冊となり、年7冊の発刊となった(34巻補冊、2003年7月)。
- 牛場大蔵氏追悼号(2003年12月)
- 特集「卒後臨床研修必修化後の専門医制度を考える」(35巻3号、2004年6月)
- 特集「臨床研修後の進路:研修必修化2年を経て」(36巻5号、2005年10月)
- 新投稿規程の発表(36巻5号、2005年10月)
- 特集「各国の医学教育」(36巻6号、2005年12月)

## 第14期での編集後記

- 35巻1号(2004年2月:庄司進一編集委員長):  
本号に掲載した論文「『心周期:初めの一步』において学習者と評価者とを盲検化した randomized controlled trial」はランダム化比較デザインを用いた教材評価の研究で、医学教育研究の原著論文の基準を十分に満たしていると考えています。

- 35巻4号(2004年8月:庄司進一編集委員長):  
教育学の研究では、比較の測定方法、変動要因の分析、などを厳密に行い、統計学的分析と論理的考察からいいすぎでない結論を引き出す研究が期待されます。本誌はこれらの視点から原著と報告を分けて考えています。
- 35巻5号(2004年10月:庄司進一編集委員長):  
本誌への投稿論文が増え、学位論文として投稿される方もおられます。投稿論文の査読の迅速化に努力しております。できるだけ医学教育学の高いレベルの論文を掲載したいと考え、どの投稿論文も何度も修正し、できるだけ載せると今までの方針から少しずつ転換を図っております。

- 35巻6号(2004年12月:庄司進一編集委員長):  
本号の文献紹介は、『医学教育の研究』についての論文はどこを探せばよいか?』の題で、本誌の「これまでの原著のカテゴリーにされた論文のうち、調査や試行を含めても原著とされる論文以外の記事の方が多く、さらにそのなかで『研究』論文と言えるものは全体の3%にも満たないのではないかと見なされる」と書かれています。医学教育学の国際誌の中でトップランクはAcademic Medicine で「医学教育の研究」論文は11.23%とのデータも引用されています。近々本誌の投稿規程を改訂し、投稿者や読者に原著論文の定義などが明確になるようにしたいと考えております。そしてわが国の文化に根ざした、わが国の医学教育の研究者が主体的に行った医学教育の研究が数多く発信されるようになることを願っています。

## 編集委員会の悩み

- 原著論文とは何か？原著と報告をどのように区分するのか？
- 特集を通じ、広く読者に伝えるべきことは何か？
- どうやったら、「掲載不可」にできるのか？

## 投稿規程の改訂

- 医学教育研究を進めるために、「医学教育研究」の原著とはなにか、を定義していかなければならない。
- 「原著」にしては根拠が弱かったり、研究のデザインがされてなかったものを、安易に「報告」としてきたことに対する反省があった。
- 機関誌が持つ「誘導作用」についての自覚が必要である。  
→ 投稿規程改訂グループ(グループ長:吉岡)

## 第15期 日本医学教育学会編集委員会

委員長 : 吉岡俊正

副委員長: 堀内三郎、伴 信太郎

委員 : 相澤好治、福島 統、森田孝夫、  
高木 康、椎橋実智男、天野隆弘、  
武田裕子

顧問 : 齋藤宣彦、庄司進一

---

## 第15期の編集

- 特集「平成17年度『特色ある大学教育支援プログラム』、『現代的教育ニーズ取組支援プログラム』」(37巻1号、2006年2月)
  - 特集「卒前医学教育専従部署の紹介」(37巻1号、2006年2月～37巻3号、2006年6月)
  - 特集「卒前医学教育責任委員会の紹介」(37巻5号、2006年10月～)
-

## 第15期の編集後記

- 37巻1号(2006年2月:吉岡俊正編集委員長):『医学教育』は日本医学教育学会の機関誌であり、また高等教育に関する日本の数少ない専門誌でもあります。本誌は医学教育学会の発展とともに読者が増え掲載される論文も高質になってきました。……投稿論文が増え、査読システムが充実し、さらに新たな投稿規定が制定されました。……今期の委員会は、質の向上とともに範囲の拡大にも努力したいと思います。すなわち、新投稿規程にも書かれているように、本誌は医学を中心として医療系教育全般を対象とする教育研究の公開の場でありたいと思います。また、卒前教育だけでなく、卒後研修、大学院教育、生涯学習などを中心とした情報公開の場・

## 旧投稿規程

- 目的: 広く医学教育に関連する研究成果の発表ならびに当学会の活動を含め情報交換をおもな目的とし、
  - 原著: 目的、対象、方法、結果および考察の明快なもので、独創的な学術論文(6,000字)
  - 総説: 解説、紹介、翻訳など(6,000字)
  - 論壇: 意見、主張、提案など(4,000字)
  - 報告: 学会、研究会、委員会、視察などの報告(6,000字)
  - 資料論文: 資料として意義のあるもの(6,000字)
  - 学生のページ: 学生の投稿欄(4,000字)
  - アナウンスメント、ニュース: 学会、機関委員会などの活動の予告、速報(600字)
  - 文献紹介、書評: (600字)
  - てがみ: (600字)

## 新投稿規程(2006年1月1日施行)

- 「医学教育の目的」: 本誌は医学教育だけでなく、歯学教育、看護学教育、薬学教育など広く医療人の育成に関する研究ならびに党学会の活動を含めた情報交換に資することを目的とする。
- 「論文」の査読: 2001年から2005年受付の投稿論文採択率は83.2%である。
- 「てがみ」、「アナウンスメント」、「ニュース」、「文献紹介」の区分を「論文」とは分けた。

## 論文の区分:原著-総合的研究

- 総合的研究は、根拠に基づく新たな知見を提供する論文である。教育研究では、定量的研究だけでなく定性的研究も重要であると本誌は考える。定量的研究には測定方法と分析方法とが明確な、横断的および縦断的研究が含まれ、定性的研究は新たな仮説・治験を示唆する根拠となる非定量的なエビデンスの集積が含まれる。単一の非定量的観察結果、教育方略の実施結果などは、エビデンスの集積ではないのでこの区分から除かれる。

## 原著-探索的研究

- 観察に基づく記述的研究論文である。単一の独創的教育法の実践、従来の見解の実証、国外における教育方法の導入事例などが相当する。

→ 原著を総合的研究と探索的研究に分ける。

- ① 研究目的、研究方法、結果の吟味、新見
- ② 作業仮説、設計された観察、観察結果の検証、考察

## その他の論文

- **総説**: 著者自身およびその他の複数の論文、エビデンスを根拠にする事象の解説、紹介、新たな概念の提示の論文である。
- **主張**: 限られた見解、少数のエビデンスに基づく、提案、議論、意見、主張を述べた論文である。
- **報告**: 教員あるいは学生の経験、体験、国外のレポートなどで、資料として広く活用できる論文である。
- **資料**: 教育マニュアル、行動目標、ガイドライン、あるいは方法・考察のない統計データなどで、医学教育に活用できる資料となるもの。
- **特集・招待論文**
- **英文論文**

## 査読要項

- 対象となる論文： 原著、総説、主張、報告、その他
- 査読過程：
  - ①編集委員長による検討
  - ②担当編集委員による2名査読(学会員:6週以内)
  - ③問題のある場合は3人目の査読者を選定する
  - ④再査読は1回までで、その後は編集委員会で採否を決める
  - ⑤査読者2名が不採用としたものは編集委員会に報告する
  - ⑥編集委員会からの調整・修正依頼も1回のみ
  - ⑦不採用の場合は疑義照会できる
  - ⑧5ヶ月以内に採否を決定

## 第13期から第15期の編集委員会

医学教育研究→医療者教育研究、高等教育研究

- 研究目的
- 研究のデザイン
- データ収集方法の再現性と妥当性
- データ解析の方法論(方法論的限界、妥当性)
- 解析結果からの結論

## 福島統先生 講演「医学教育誌編集委員会が目指す研究論文のあり方」

おはようございます。慈恵医大の福島です。実は編集委員会の委員長はアメリカに行っちゃって、・・の調査に行っていて、それから副委員長は・・でそれで仕方なく私に回ってきました。『医学教育』の雑誌のことで振り返ってみたいと思います。

SLIDE—2 実は私が医学教育の編集委員会に入ったのは第13期で、今15期なんですね。その2つ前の13期から入ったんです。医学教育学会のことをご存知だと思いますけれど、第14期という前の期がですね、ドラスティックに改革された時でありまして、14期の時に例の評議員理事制になったのです。それ以前はですね運営委員会があって、僕も全体が分からないですけど、割と年取った方が集まって牛耳っていたと伺っておりまして、その運営委員会という体制の最後の時に実はそれが第13期です。第13期の時にそれ以前はですね編集委員会と運営委員会が一緒だったんですね。運営委員会は20人くらいいたらしいんですけど、編集委員会と運営委員会同じで運営委員会の中で編集作業をしていたと聞いています。それじゃいけないというんで第13期のときに当時の会長だったのが尾島先生ですが、尾島先生が編集委員会を独立させるということをしたようであります。それが第13期医学教育学会編集委員会でそのときの会長は鈴木淳一先生で当時の学会の副会長が副委員長をやるということだったので畑尾先生と斎藤先生がそれをやったということになりました。そのときに実は運営委員でない人間で編集委員をする人ということで三人最初追加されたのが北里の相澤先生、公衆衛生ですね、私が基礎ということで私、あと臨床系ということで当時埼玉医大の森田孝夫先生が入ったというのが実は始まりであります。ですからこの時初めて運営委員会が独立したんですけど、その時にはですね2000年の4月に発行されたのがこの新しい委員会の最初の方でありまして、

SLIDE—3 その時の編集後記に鈴木淳一編集委員長が書いたのが「筆者は図らずも編集委員長に返り咲くというか、昔長くやっていたようですが、尾島昭次新会長のたつての希望に沿ったものである。2人の副委員長と、新鋭だそうですが、三人の編集委員が全く新規に入ることになって、それでやりました」とそれで組織をつくって、新しい編集方針にするんだということを宣言しています。新しい編集方針の一つがですね実はレフェリーにありまして、どうも(僕も詳しいことは知らないですけど)昔は運営委員の間で査読を回していた節があるようでありまして、それをですね、これからは査読というのは学会員のすべての人に頼んでもいいんだ、というふうにしてですねオープン化したということを鈴木先生はされました。それと同時にですね編集委員会主導で特集を組むってということもやりました。

SLIDE—4 丁度この時期はですね、激動の時期でありまして、モデル・コア・カリキュラムと共用試験という特集を組んだくらいで、丁度、モデル・コア・カリキュラムが出てきて共用試験のトライアルが始まるという時期に丁度当たっていますし、それからもう一つ組んだのが

Community-Based Medical Education という特集を組みました。モデル・コア・カリキュラムと共用試験の特集に関しては時期、そういう時期だったから組んだということだけでありまして、こちらの Community-Based Medical Education の方は明らかに編集委員会が作為を持って特集を組んだという一つの例であります。というのは臨床実習の改善という中で大学病院の中だけでやるというのはおかしいよねと問題提起をちゃんとしましょうと、それからいわゆる診療所実習だとか地域での病院実習だとかということを特に考えていた地域病院の実習だけでお茶を濁したら駄目だよよねというようなことを提起したいという意志があって、こういう特集を組みました。そのことをして3年半、第 13 期の編集委員会が続きます。3 年半の半というのは、例の評議員理事制というのを医学教育学会に導入するときちょっともたついたということで、3 年半になっちゃったわけですが、

SLIDE—5 そのときの最後の鈴木先生の編集後記はこれです。「私どもが編集を担当して早 3 年半が経った。われわれの編集委員会のスタートには 2 つの決断があった。今なら書いてもいいのだろうが」と書いていますが、ま、たいしたことはないんですが、一つは篠原出版がつぶれちゃったということで篠原出版新社に移行するという時期だったということでちょっとこずりました。それはそれでいいんですけどもその次が問題で、「編集委員は学会発足以来ずっと運営委員が兼任していたことを止め」、これ止めるに当たってはかなりその長老たちから批判を受けたようであります。だけど尾島先生と鈴木先生はそれをやったわけで、それで、運営委員会が兼任していたのを止めて、編集担当に相応しい人達を選んだと、そして実は鈴木先生はこの編集委員会を非常に大事にされていて、誰も欠席しないで、ずっと連続して無欠席の編集委員会だったので、それを非常に誇りにされていました。(編集委員会では時々ビール飲まされていたのはちょっと辛かったですね(笑)・鈴木先生はビールが好きだったので、それはいいとして、)そのそれで査読を厳しくするというを一生懸命された、そのために鈴木先生が何回か犠牲になって、reject した人から散々なことを言われたということがありました。それ以前まではあまり reject されていなかったようでありまして、医学教育で reject するとは何事だということを鈴木先生は随分いわれたみたいですが、そういうことを始めたのがこの第 13 期であります。そういう意味で reject するというのをやり出したということとそれからもう一つはその査読者の幅を広げたということが第 13 期の特徴です。

SLIDE—6 それと同時に実はホームページも立ち上げるということを第 13 期はやりました。そうするとですね、第 14 期評議員理事制になった時の委員長が庄司進一先生で、このときはですね、編集委員会とは言わない、編集・出版・HP 委員会というふうに名前が変わっております。副委員長は副会長の堀内先生がやって、そして委員としては追加されている人たちがいて、このメンバーこの中にもいますが、こういう形になりました。庄司先生はですね・鈴木先生はなかなかその原著論文というのをどうするかとうことに関してあまりそのつっこんで議論はされなかったんですが、庄司先生はどちらかというと何が原著であるかということを編集委員会の度に話題

にされる方でありました。第 14 期の庄司先生の時代はですねちょっと拡大路線があつて、そのホームページをもっと頑張りましょう、みたいなことをやっていたのと

SLIDE—7 それから年間 6 冊発行から 7 冊にしたというのが庄司先生の時代です。要するに学会の抄録集を補冊にしたということです。残念なことに牛場先生の追悼号を出さなくてはならなくなってそれもやった。それであと特集ですけれども、これも編集委員会は作爲的にこのテーマをわざわざ2つ連続して取り上げるということをしました。つまり卒後臨床研修がどうだったのかということと卒後研修が2年経って、さあその次にどうするのかということと連続して35巻、36巻というふうに卒後研修の方にフォーカスを当てることを致しました。あと、大西先生(関係していますね)の「各国の医学教育」もありまして、これもめずらしい話でありまして、編集委員会以外の人から、このときは吉岡先生だよ(そうです)。「各国の医学教育」というのがありますが、今までは特集というのは編集委員が担当するということだった、編集委員会の中だけで担当するということでしたけど、そういう特集の企画も広げて、このときは編集委員としては吉岡先生がやって、編集委員会外として大西先生に助力を頂いて2人の共同企画という形をとって、間口を広げましょうみたいなことをいたしておりました。

SLIDE—8 庄司先生は作爲的にどうも編集後記の中で原著論文は何かということとをちょこちょこ小出しにしたようでありまして、今から思うと。これは35巻6号の編集後記でありますけども、庄司先生が編集委員長になってすぐのときかな、半年位経ってからのやつだと思いますけど、「本号に掲載した論文・これは渋谷先生が書いたものですが、心周期:初めの一步にて学習者と評価者とを盲検化した randomized controlled trial はランダム化比較デザインを用いて教材評価の研究で、医学教育研究の原著論文の基準を十分に満たしていると考えています」、というようなことをわざわざ編集後記に書いています。これはあとで賞をもらうんですよ。(なんでしたっけ...) わざわざこれを編集後記に書いていて、庄司先生は非常に、何て言いましょか...ただアンケートしましたとか、授業実践をしましたとかいうのは原著ではない、もっと高いレベルのものを用いなきゃいけない、そのためには大きくは研究の目的と方法論が合致しているかどうか、方法論的なテクニカルエラーをどうやって排除するのかということが考えられているかどうかということもいつもおっしゃっていました。もし、観察研究するにしてもですね、やはり予備研究があつて、ある作業仮説があつてそして本調査がある。というようなものでなければ原著ではないんだということとを一生懸命議論をしていました。

SLIDE—9 そういうことがあるので実は編集後記の中にもそのようなことを実は庄司先生は書き残していらっしゃるわけで、「教育学の研究では比較の測定方法、変動要因の分析を厳密に行い、統計学的分析と理論的考察から言い過ぎでない結論を引き出す研究が原著論文で」、嫌味が入っているようですが、「言い過ぎない結論を引き出す」ということをわざわざ書いています。本誌はこれらの視点から原著と報告を別けて考えるというようなことを宣言しています。実はこ

れ、原著にするか報告にするかというのか凄い大変なんですよ。というのは学位論文で出してくる方がいてですね、それを報告とこちらがやるとですね、怒鳴り込んでこられて、それじゃ学位がとれないじゃないかと、という話しになる、すごくもめたということもこのときありました。それが書いてありますね下に。「本誌への学位論文としての投稿論文が増え、学位論文として投稿される方もおられる。投稿論文の査読の迅速化には努力しています。できるだけ医学論文の医学教育学の高いレベルの論文を掲載したいと考え、どの投稿論文も何度も修正し」、この“何度も修正し”というのが大事でありまして、4回とかやったよね。で、修正してできるだけ載せると今までの方針から少しずつ転換を図っております。実はですね、かつては査読は1回、という…。査読は1回でそれで思いどおりにならなくても reject なかったもので、そうなると載せちゃうと、そうするとかなりその悪いものが平気で載っていたであろうと思われま。それを鈴木先生の時代からそれを変えていきました。たしか記憶だと、4回か5回かやりとりした記憶がある論文があります。それで最後に reject したというものもかつてあって、そういう意味ではかなりしつこく再査読を繰り返して、査読者との間でやりとりをするというようなことを奨励はしなかったですけど、やらざるをえないというか、やってきました。

SLIDE—10 ところがですねこの庄司先生の終わりの時代になるとですね、堀原一先生が、実はこの医学研究についての論文はどこを探せばよいのかという記事を載せられます。それに対する編集後記がこれです。「本誌のこれまでの原著のカテゴリーにされた論文で、これは堀先生が言っていることでありますが、調査や試行を含めても原著とされる論文以外の記事の方が多く、さらにそのなかで『研究』論文といえるものが全体の中で3%にも満たないのではないかと見なされる」と堀原一先生がお書きになられています。医学教育の国際誌の中で Academic Medicine ですが、医学教育の研究論文は Academic Medicine では 11.23%である。「近々本誌の投稿規程を改訂し、投稿者や読者に原著論文の定義などが明確になるようにしたいと考えております。」これは庄司先生のお言葉です。「そしてわが国の文化に根ざした、わが国の医学教育の研究者が主体的に行った医学教育の研究が数多く発信されるようになることを願っております。」というようなことを書いていらっしやいまして、明らかにその堀先生は実は国際比較をして、Academic Medicine であるとか、Medical Education とかと比較されてですね、それで、医学教育の雑誌もつとががんばらなきゃだめだね、ということおっしゃっています。

SLIDE—11 結局、編集委員会いろいろ鈴木委員長の時でも庄司委員長時代も非常に悩みましたよね。理由は先程からお話しているように、“何が原著か”という定義づけでかなり悩みました。

SLIDE—12 それでその原著でないものは、安易に報告としていたという深い反省がありました。その議論の中ですね、やっぱり先程、庄司先生も書いていましたけど、投稿規程が法律になりますから、投稿規程をどうにかしなきゃいけないという話しになってまいりました。これ1年間近くや

りましたかね、1年近くやりました。それで医学教育を進めるためには、医学教育研究の原著とは何かという定義づけをしなければならない。そのためには投稿規程を改訂しなければならない。そして今までそういうことをしていなかったから、原著だ報告だ、報告っていても訳分かんない定義で創ってきたではないか、で、むしろ雑誌がですね、原著はこういうものですよ、だからお願いします、書いて下さいねという姿勢をとらなきゃいけないということで半年以上やったかな。

SLIDE—13 投稿規程の改訂グループというのが編集委員会の中でできて、そのグループ長が吉岡先生であったということでもあります。(でちょっと飛ばして..)実は編集後記、後で投稿規程の話は後でしますけれど、そして今年1月1日から第15期になって吉岡先生が委員長になってこのメンバーで今活動しているということになります。

SLIDE—14 15期は始まったばかりなので、あまり特集というのはやっていないです。特集としてはGP特集をやって、ただその下に2つあるのは作弄的なものがありまして、実は編集委員会でまだ決定はしていませんけど考えていることは、卒前医学教育、要するに医学教育ユニットの特集を組みました。医学教育ユニットの特集を組んで、その次にですね、要するに政治的にやったのですが、所謂教務委員会の特集あえてやりました。これには理屈があります。というのは実は議論の最中ですからあまりあれですけど、実はこの次にですね、教育事務組織の特集をしたいということを今考えているところです。もし、そういうことができると、今特集というのは誘導作用があると申しましたけれど、実は我々が考えています誘導作用の一つはですね、教育管理ですね。Administrationを取り上げていく時期にきていると思っているわけで、皆様よくご存知のように高等教育の世界ではそのAdministrationのことを管理に関して行政管理ということに関しての問題点というのは高等教育の方で、今凄く大事にされていて、いわゆるスタッフ development という話と繋がっているわけです。ところが医学教育というのはまだそこまで十分いってなくてただユニットができてきて、教員組織なのか職員組織なのかはわかんない..という組織ができてきてお茶を濁しているという段階かもしれません。さしあたり..そういうこと..それでは医学教育のマネジメントはできないだろうと、そのためにはやはり教員組織と教務組織と事務組織というものが日本の中でどうなっているのかということは調べていかなければいけないのかなんてことをやって実はそんな作為をこめてちょっと特集を..やることになりました。

SLIDE—15 さて、15期になってですね、吉岡編集委員長になって吉岡先生のこれが所信表明ですが、「医学教育は日本医学教育学会の機関誌であり、また、高等教育に関する日本の数少ない専門誌でもあります。」これを吉岡先生は強調したいわけですね。高等教育の分野であるということを彼は強調したいわけです。「本誌は医学教育学会の発展とともに読者が増え掲載される論文も高質になってきました。投稿論文が増え、読システムが充実し、そして新たな投稿規程が制定されました。今期の委員会は質の向上とともに範囲の拡大にも努力したいと思

います。」これも重要なことでありまして、医学教育ではなくて医療者教育にしていこうということ  
を鮮明に出しております。範囲の拡大にも努力したいと思います。すなわち新投稿規程にも書  
かれているように本誌は医学を中心として医療系教育全般を対象とする教育研究の公開の場  
でありたいと思います。また卒前教育だけでなく、卒後研修、大学院教育、生涯学習などを中  
心とした情報公開の場にしたい、というそういう方針でやっています。

SLIDE—16 さて、実はその一番大事な投稿規程でありまして、実は昔の投稿規程は凄いもので  
ありまして、見て頂ければ分かりますけど、まず目的ですが、広く医学教育に関連する研究成  
果の発表ならびに問う・・・(ごめんなさいミス・・・)・・・本学会ですね、活動を含め情報交換をおも  
な目的としている。これよく分からないですが、それよりもっと問題なのは原著の定義でありま  
す。原著の定義はですね、目的、対象、方法、結果および考察の明快なんて書いてありますが  
何が明快かというのが・・・これしか定義がないですね、これではどうにもなりません。しかもで  
すね、総説、論壇(意見や主張ですけど)、報告というのを項目がこういう定義だったんですね。  
学会、研究会、委員会、視察などの報告、つまり原著にならないから報告にするというのは全く  
投稿規程上誤った使い方をしていたわけでありまして、分かっているんです。原著にたくない、  
さあどうするか、総説じゃないよね、論壇っていったら怒るよね。そういう世界でやっておりまし  
て、で・・・ま・・・こんなものだったんです。これじゃどうもならないわけでありまして、それで、その  
新投稿規程というものを吉岡先生が中心に行って作られたわけでありまして、ちょっとかなり外国  
の雑誌の投稿規程をもじっているところがあるというのは否めないのですがそれでも結構  
議論して作りました。

SLIDE—17 その中でまず医学教育の目的の中にですね本誌は医学教育だけでなく歯学教育、  
看護学教育、薬学教育など広く医療人の育成に関する研究(・・・すみません・・・党→当)学会の  
活動を含めた情報交換に質することを目的としているということで、医療者教育という方向にい  
くよ、ということを宣言しました。それから論文の査読の状態を明記しまして、2001年から2005  
年の間の採択率が83.2%と書いてありますがこれは実は甘い時代を入れてあります。現実の  
問題として75%です。でして、それなりに編集委員会としては傷を負いながらも頑張っているこ  
とをご理解いただければと思いますが。あと「てがみ」とか「アナウンス」、「ニュース」「文献紹  
介」の区分を並列にしないで明らかに「論文」とは違うということで分けたということになりました。

SLIDE—18 さて、原著の扱いでありまして、原著の扱いではかなりの苦労がありました。というの  
はこれが原著だよねというのと、ちょっと原著にはならないよね、というのを原著、報告というふう  
にやって差別化していた。でもこの差別化はどうか・・・しなきゃいけない形を考えるそれで苦  
肉の策が原著—総合的研究の原著と探索的研究というのにあえて分けるということをしたわけ  
です。まずその私たちが本質的に原著だと思っていたのはこっちのほうでありまして、原著—  
総合的研究というのはですね、根拠に基づく新たな知見を提供する、根拠に基づくということ

重要視するというわけで、教育研究では、定量的研究だけでなく定性的研究も重要であるとあえて書いているわけでこれは、Medical Education に書いているわけで、というわけで量的な研究だけでなく質的な研究も勿論歓迎する。あえて定量的な測定方法には明確な横断的・縦断的研究、定性的研究は新たな知見（知見は知って見るほうですから臨床治験とは違う、すみません）、を示唆する根拠となる非定性的なエビデンスが含まれる。ここが重要で単一、非定量的観察結果、教育方略の実施結果などはエビデンスの集積ではない、あえてここで書いています。つまりこれを報告としてお茶を濁していたんですがこれを原著—総合的研究ではないということをも明言して、定量的研究に関してはデータの集め方、それから解析の仕方、それが目的に合っているか、そしてそれが言い過ぎない結論に論理的に導き出されているか、ということを中心にみる、質的な方に関しては仮説があって、仮説を検証するための調査を行って、そしてそれを証明していくというエビデンスを論理的な研究のデザインがあって、そして行われる結果ということを大事にしようそれを総合的研究にするんだと、従って数値だけ並んでいたらいいという考え方はしないように、但し、その質的なものをやるときには一番重要なことは研究のデザインであるということをお話したいわけでありまして。一方、妥協産物がまだ残っておりますが、

SLIDE—19 原著—探索研究であります、観察にもとづく記述的研究論文である。単一の独創的教育法の実践、従来の知見の実証、国外における教育方法の導入事例などが相当する。ということでケースレポートというともた怒られますが、それに近いものをここに残したのは、ちょっとごめんなさいやむをえない、そこまでちょっとできなかつたんで、きっと吉岡先生ももう一、二期やれば少し明確なものを創るんじゃないかと思っておりますが、今はここでちょっと…ごめんなさい。というわけで原著を総合的研究に差別化して、総合的研究の促したいというのが一応編集委員会の考え方でありまして。もう一回書きちゃいましたけど、やはり研究のデザインというものをちゃんと見たいと思っております。目的は何ですか、目的に合ったデザインが組まれているかどうか、そしてその解析方法、テクニカルエラーの部分がちゃんと検証されていて、そして、言い過ぎない結論が導き出されているという条件が大事だろうというふうに思っているわけでありまして。やはり作業仮説を立てていくというデザインが非常に重要だと思います。

SLIDE—20 まあその他の論文として、原著はこういうふうに分けて、その他の論文として総説、主張、報告はちゃんと定義してですね。最初、報告は経験にしようかとしていたんですが、経験じゃわかんないだろうということでまた報告という名前を残しちゃいましたけど、国外のレポートを資料としてそれを広く利用できる活用できる論文として生かしております。それと同時に英文の投稿規程もつけております。それから倫理に関する事項の書き込んだということになります。

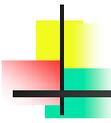
SLIDE—21 さて、査読であります、実はこの査読の基本的なことは 13 期の鈴木編集委員長が英断をもって決めました。つまり担当編集委員制であります。つまり投稿されてきます。投稿さ

れると一番先に委員長がさっと見ます。ざっと見てですね、やばいと思ったときにはもう一人編集委員に送ってですね、これreject ですよというreject。できるだけ早くrejectしないと大変なことになると散々味わったので、今そういう形になっていて、一番最初に編集委員長が見ます。ただどうしても編集委員長が見てもちょっと甘くなっちゃうと後でreject するということが勿論起ってきますが、そして編集委員長はそれを見ると、じゃこれは編集委員の誰に担当させるかということ指名します、それが担当編集委員になります。ですから差読は担当編集委員の名前で送り返されていくということですが、それは担当編集委員は、その論文に関してはほぼ編集委員長ですね。査読者を担当編集委員が自分で勝手に決めるのです。学会員だったら誰でもいいということになっていますから決めます。そしてその査読の結果を担当編集委員が見てそれでどうするかを決めて、たとえばこのまま編集委員会で審議するのか、もしか再査読にまわすのか、若しくはどうしようと編集委員長に相談するのか、そういうことは全部担当編集委員が任されています。しかも、査読期間は6週以内で返ってこなかったら査読者は切っちゃう、切っちゃってブラックリストがあります。あまりお願いしないようにしているのですが、期間がきて来なかったら、第三者を立てる、そういうことになりまして実際やっているんです。これはあまり問題ないんですけど、査読者2人の意見が違う、というときが一番問題になります。査読者の意見がこっちはいいね、こっちは絶対駄目という、そうすると第三の人を立てるわけです。それでその人に査読してもらったものを担当編集委員が最後ジャッジをして編集委員会で報告する。あとはまあ査読者2人が不採用報告してreject するしかないです。それですね実は6番目が投稿規程に書いてあって、これは隠したほうがいいといわれたのですが、出ているから仕様がないうんです。この投稿規程でも編集委員会からの調整修正は一回のみということで本来再査読はないです。再査読は投稿規程上許されていないのです。ですけど、私たち自分たちで作った法律を自分たちで犯しておいて実際はやっています。というのはいうと怒られちゃうかもしれませんが、私の立場でいうとですね、学生のレポートじゃないんだから、というのが来るわけです。しかもそれがそれなりの名前がそれにくっついている論文で、学生のレポートじゃないんだって、本当に怒りたくなるのがきます。さらにですね、人の意見を聞かない人がいる、査読者がこれ統計的なここ間違っているからもう一回検討してください、というんですね検討しないのです。ますます学生じゃないんだよと思うのです。これが私の学生だったら赤を入れる・・・いや実際にそういう意味ではですね、こういうことをいうと怒られますけど・・・その研究体制を疑いたくなります。そこが本当に研究組織になっているのかよと思うのがあってですね、それをreject できない場合は、この形では載せられないと思うわけで、そしたらしょうがないね、掟破りをやっています。私も4回再査読の論文を抱えていますけどそれはあえてやっていますということ白状します。あとはその不採択になったときに疑義照会ができるようなシステムになりますし、頂いたものは5ヶ月以内に決定するという形になっています。

SLIDE—22 というわけでまとめますと、随分13期から医学教育の編集が変わりました。変わった一番言いたいことは医療者教育に視点を移していくということが明確化されたこと、それとあと

う一つはいわゆる研究デザインと解析方法というものについての吟味をもっと厳しくしていきますよ、ということを確認化してきたということお示したいと思います。これが実際私たちが、やってきたことですし、むしろ私達が目指す研究論文のあり方、総合的研究をぜひお願いいたします。というのが編集委員会の考えです。ということです。

どうも、ご静聴ありがとうございました。

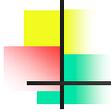


# 統計学的解析

---

大西弘高

東京大学医学教育国際協力研究センター  
日本医学教育学会医学教育研究開発委員会



## 研究デザイン

---

- 事後テスト単独 vs 事前事後テスト
  - 事前事後テストにより改善度が測定可能
- 対照群の有無(対照群有:準実験)
- ランダム化の有無(ランダム化:真の実験)

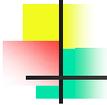
## 一般的な研究デザイン

前実験的	事後テスト単独	X -- O
	事前事後テスト	O <sub>1</sub> -- X -- O <sub>2</sub>
準実験的	非ランダム化比較 事前事後テスト	E O <sub>1</sub> -- X -- O <sub>2</sub> C O <sub>1</sub> ----- O <sub>2</sub>
	ランダム化比較 事後テスト単独	R E X --- O <sub>1</sub> C --- O <sub>1</sub>
真の 実験的	ランダム化比較 事前事後テスト	R E O <sub>1</sub> -- X -- O <sub>2</sub> C O <sub>1</sub> ----- O <sub>2</sub>

X: 教育介入, O: 観察か測定, E: 実験群, C: 対照群, R: ランダム割り付け

## 統計学的解析

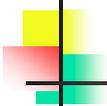
- 記述統計
  - 割合, 分布(ヒストグラム), 平均, 標準偏差...
- 分析的統計
  - 平均の比較
  - 相関
  - 回帰



## 変数

---

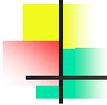
- 連続変数
  - 正規分布
  - 非正規分布
  - イベントまでの時間 (censored)
- 順序変数
- 名義変数 ( $\geq 3$ 群)
- 二区分変数 (dichotomous)



## 連続変数

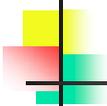
---

- “測定”のデータの多くはこの形
  - テストの点数のように加算してよいか？
- Pros
  - 適切に処理すれば、事象を鋭敏に表現
  - 統計学的検出力も高い
- Cons
  - 欠損値が発生しやすい→多変量に不利
- コツ
  - 分布と欠損値の頻度を常にチェック
  - 数学的処理、順序化、二区分化を考慮



## 順序変数

- 連続変数と二区分変数の中間
- 一見連続変数
- Pros
  - 事象を二区分変数よりも鋭敏に検出
- Cons
  - 欠損値が発生しやすい



## 名義変数(≥3群)

- 順序化できない対象集団の特徴
  - 出身校
- Cons
  - 基本的に数字変換は不可能
  - 統計処理が困難
- コツ
  - 医学的標準、最大標本数の群を標準としてダミーコード化



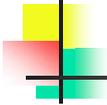
## ダミーコード化

標本ID	出身校	Private	National	Oversea
1	公立 高校	0	0	0
2	私立 高校	1	0	0
3	国立 付属	0	1	0
4	帰国 子女	0	0	1



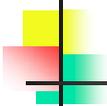
## 二区分変数

- イベントの評価
- 他のタイプの変数の変換型
- Pros
  - 欠損値が少ない ~ 欠損値=“なし”
  - 解析や結果の解釈が容易
- Cons
  - 鋭敏さに欠ける
- コツ
  - 2元変数の組み合わせ~元々ダミー



## 平均の差をみる統計解析

検定目的	パラメトリック	ノンパラメトリック (順序尺度以上)
対応のない2標本の代表値の差	平均値の差のt検定	マン・ホイットニーのU検定
対応のある2標本の代表値の差	平均値の差のt検定	符号検定 符号付順位和検定
対応のないK標本の代表値の差	一元配置分散分析	クラスカル・ウォリス検定
対応のあるK標本の代表値の差	混合型分散分析	フリードマンの検定



## パラメトリックとノンパラ

- ノンパラメトリックな手法は順序／名義変数(カテゴリー変数)に利用可能
- パラメトリックな手法は正規分布, 等分散を仮定して利用可能
- ノンパラメトリックな手法は分布型を問わない。しかし, 検出力が劣る傾向があり, 計算も難しい。濫用は控えるべき
- nが30を超えるような場合, パラメトリックな手法でも頑健性からまず問題なし

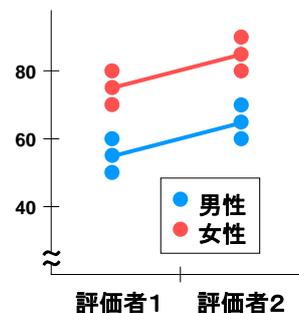
## p値とEffect Size

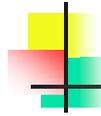
- p値は帰無仮説が成り立つ確率
- Effect sizeは差が大きいかどうか
  - "small,  $d = 0.2$ ," "medium,  $d = 0.5$ ," and "large,  $d = 0.8$ " (Cohen, 1988)
- 差が小さくてもサンプルサイズが大きければ有意になることあり

## 分散分析 = Analysis of Variance

- 反復測定の有無を間違えない
  - 反復測定のないデータ (between subject)
  - 反復測定のあるデータ (within subject)
- Excel, SPSSでは
  - 繰り返しのないデータは縦に重ねる
  - 繰り返しのあるデータは横に並べる

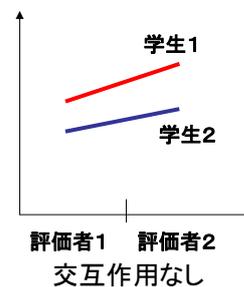
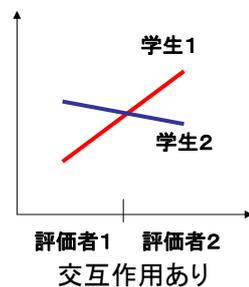
ID	男女	評価者1	評価者2
1	男	50	60
2	男	60	70
3	女	70	80
4	女	80	90





## 2way, 3way ANOVA

- 同時に二つ以上の要因の比較ができる
- 繰り返しは有っても無くても組み合わせ可
  - 繰り返し有る場合は球面性が問題に
- 二元配置, 三元配置の場合, 交互作用も確認



## 研究計画法

- 3要因 ( $2 \times 2 \times 2$ )
  - 直交計画
  - 要因  $1 \times 2$ ,  $1 \times 3$ ,  $2 \times 3$ ,  $1 \times 2 \times 3$  といった交互作用も求められる
  - この場合はいずれの要因においても測定の繰り返し無

	要因 1	要因 2	要因 3
グループ1	+	+	+
グループ2	+	+	-
グループ3	+	-	+
グループ4	+	-	-
グループ5	-	+	+
グループ6	-	+	-
グループ7	-	-	+
グループ8	-	-	-



## カテゴリー間, 二変数間の関係

- 二区分変数, 名義変数: まずはtableに
  - 独立性:  $\chi^2$ 検定
  - 要因による差: odds ratio, relative risk
- 順序変数, 連続変数: まずは散布図を
  - 順序変数: スピアマンの  $\rho(r_s)$
  - 連続変数: ピアソン積率相関係数( $r$ )
  - 3つ以上の変数: 多変量解析へ

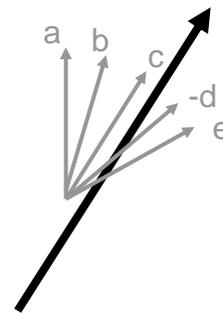


## 多変量解析

- 探索的解析か仮説検証解析か
  - 従属変数, 独立変数の違いは明確か
  - 関連性 (association) と因果関係 (causal relationship) との違い
  - 何が原因で何が結果と考えるか
- データの種類
  - 二区分変数, カテゴリー変数, 連続変数

## 内的一貫性

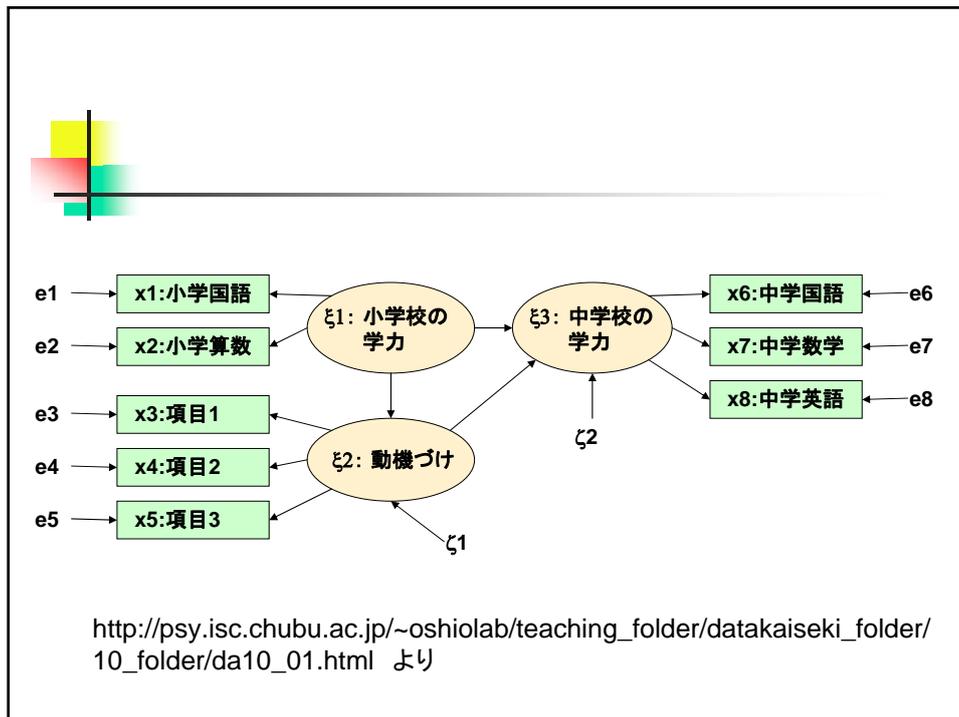
- a~eの測定ベクトルの方向性が似ているならひとまとめに合算することが許容される
- もし複数の測定が互いにあまりに似ているなら, 区別して測定する意味はない
- $0.7 < \alpha < 0.9$ が適正
- dは逆転項目



## 共分散構造分析

(SEM: structural equation models)

- 確認的因子分析など
- 因子分析とは観測変数を潜在変数で説明する点で似ているが, SEMは変数間の因果関係について仮説を立てる点で異なる
- パス解析とは観測変数の因果関係について議論する点で似ているが, SEMは潜在変数を置く点で異なる



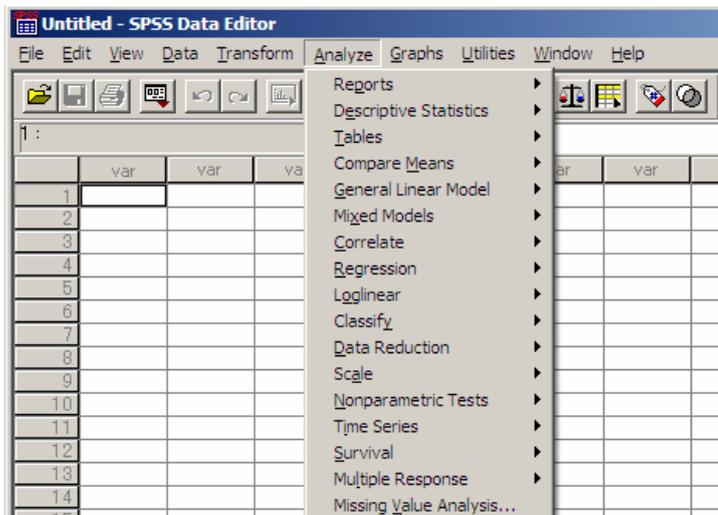
## 信頼性

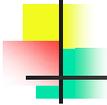
- 評価者間一致度: 評価者間一致率,  $\kappa$  値, intraclass correlation (ICC)
  - 一致率と  $\kappa$  値か ICC の 2 つがなければ単一指標では一致度の判断難しい
  - 3 者以上の評価者がいれば ICC が利用しやすい
- 一般化可能性係数, 信頼度指数
  - 前者は相対評価, 後者は絶対評価に利用される
  - 統計学的には一般化可能性理論にクロンバッハ  $\alpha$  や ICC が含まれる関係

## 回帰分析

- 一つ(単回帰), または多数(重回帰)の独立変数によって, 従属変数の予測式を作る
  - $y = a + b_1X_1 + b_2X_2 + \dots + b_mX_m$
- 従属変数の種類と回帰分析の関係
  - 間隔変数: 線形回帰分析
  - 二区分変数: ロジスティック回帰分析
  - カテゴリー変数: 間隔変数とみなせれば線形回帰. そうでなければカテゴリー変数用のロジスティック回帰分析かプロビット分析など

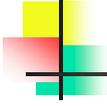
## SPSSの使い方





## 便利なWeb Page(国内)

- 群馬大青木繁信先生のページ  
<http://aoki2.si.gunma-u.ac.jp/lecture/>
- 香川大 堀啓造先生のページ  
<http://www.ec.kagawa-u.ac.jp/~hori/>
- 心理・教育のための統計の初歩  
<http://grace.ceser.hyogo-u.ac.jp/sta/index.html>



## 便利なWeb Page(海外)

- Research consulting (U Texas)  
<http://www.utexas.edu/its/rc/>
- Effect size  
<http://web.uccs.edu/lbecker/Psy590/es.htm>
- Agreement statistics  
<http://ourworld.compuserve.com/homepages/jsuebersax/agree.htm>

## 大西弘高 講演「医学教育研究とデータの種類, 統計解析」

当初の目標にしていたスライドをちょっと枚数を増やしています. いくつかこれは大事だなというふう  
に後から思った分です. では始めます.

スライド2 まず, 研究デザインについて, これは今 追加資料をお配りしました. 事後テスト単独,  
事前事後テストどちらかを選択することになります. 事後テスト単独というのも結構教育のところ  
では行われているのですけれど, 事前事後やると改善度が測定可能なので, その差をとってと  
いうことがよく行われます. それから対照群の有無ですね. 対照群, コントロール群があると当  
然比較ができますからこういうのを心理学の実験の領域では準実験 (quasi-experimental  
design) といっています. それからもし対象群を設けるなら, それはランダム化されているのか,  
割り付けがですねされているかということでランダム化されていれば真の実験 (true  
experimental design) というふうに呼ばれます.

スライド3 それらを用いで事前事後, あるいは事後のみを組み合わせると, 非ランダム化で事後の  
みというのはほとんど意味がないので, この5つがよく用いられているデザインです. 医学教育  
学会のいろんなワークショップでは事前事後やりましょうみたいな感じで随分言っているらしい  
ですが, 事前事後だけをやってコントロール群を設けないのは余り意味がありません. 前実験  
的 (pre-experimental design) といって非常に弱いデザインです. 何故かというこの2回の測定  
の間で何が起ったのかが分からないということですね. ま, 当然学生は成長するというのもあり  
ますし, そのときに例えば, 医療事故で大きなニュースとかがあればですね, 何か自分も行動  
改めないといけないなというふうな感じになる学生もいるでしょうし, そういふうなことでこれは非  
常に弱いデザインとされています. それに比べると, できる限りコントロールグループをおいた  
方がいいということになるのですが, コントロール群をおくとどうしてもお互いに釣り合ったグル  
ープなのかという問題も出てきますし, 分けるということになると「オレはコントロール群に入って  
教育受けられないのか」ということも問題になります. 特にそれはランダム化すると問題になりや  
すいので, 教育学的な研究での一番の難しいところはですね, コントロール群をおいていいの  
かなどうなのかな, もしおいていいのかなどうかということをして学生に訊いたときにどういう反応を示  
すのか, というみたいなことを考えないといけない. そこが臨床試験と違って一番違うところ  
です.

スライド4 統計学的な解析ですけれども, 記述統計あるいは分析的な統計というふうなこと2つに  
分かれます. 記述統計だけでも十分モノがいえるような研究というのたくさんあります.

スライド5 ここからは, 森本先生のスライドで言っていない部分を少し入れていますが, 連続変数,  
あるいは順序変数, 名義変数, そこには二元と書いていますが, すべて二区分に直しました.

心理学とか教育学のところではすべて二区分変数という名前を使っています。それから連続変数、測定のデータの多くはこの形ということですが、実際には質問紙とか使うと、加算していいかどうかということを常に問わないといけません。もし、5ポイントのライカート・スケール(Likert scale)で10問あるような、そういう質問紙を考えたときに、満点が50という形にしてもいいのですが、その50というふうな数字ははたして妥当なのか、信頼性をもっているのか、そういうふうなことを問わないといけません。だから加算していいかどうかという話はまた後で出てきます。

スライド6 連続変数は事象を鋭敏に表現できますし、統計学的な係数が高いということ一番利点ですね。しかし、欠損値が発生しやすい。多変量に不利である。あまりこれは考えたことはなかったのですが、そういわれればそうかなと思います。分布と欠損値の頻度を常にチェックする。数学的な処理で順序化、カテゴリー化するとか二区分化するというのもロジスティック(logistic)回帰分析とかでは必要だということですね。余り僕はこれ考えたことはないですけど。

スライド7 順序変数は、連続変数と二区分変数の中間で、一見連続的だけれども、例えば何かの問いに対して、全くそう思う、そう思う、あまりそう思わない、全くそう思わないというこの4つのカテゴリーは順序はありますけど、間隔が等しいかどうかは分からない。こういうのが順序変数ですね。二区分よりも鋭敏に検出できるということでアンケートなんかではよく使われているわけですが、これも欠損が発生しやすい。

スライド8 名義変数、出身校みたいなものですね。大学を国立・公立・私立などに分けることもありますが、基本的に数字変換の不可能で統計処理が困難であると、ダミーコード化しないといけませんというのがあります。

スライド9 ダミーコード化というのはどのようなものかという、こんな感じですね。公立高校・私立高校・国立附属、帰国子女というのを4つのグループに分けようとするので3つのダミーコードを設けてそれぞれ000, 100とかですね、このように名義変数の種類の数-1だけのダミーコードをおけばいいということになります。

スライド10 二区分変数、イベントの評価、あるかないかですね。欠損値が少ないです。解析や結果の解釈が容易。二区分変数はテーブルに書くということになりますけど鋭敏さに欠けるということ。2変数を組み合わせてもいい、こういうのもでデータを取って統計解析にする。

スライド11 例えば平均値の差を見るという場合には、2標本か3以上の標本数があるのかということで、対応がないのかあるのかということも見ないといけません。パラメトリックを使うならこんなもの、ノンパラメトリックだったらこんなもの、というふうなことになります。ここのt検定は対応のある

ものとならないものでやり方が違います。

スライド12 なんとなくこの話題になってくると、ノンパラというのが何でもかんでもできて安全でいいんだという議論に割と医学領域ではなっていたことがあります。心理とか教育学の方では全くそういう方向には進んでいませんで、パラメトリックがいいというふうな意見が非常に多いです。ノンパラメトリックの手法は、順序とか名義変数に利用可能だということですね。パラメトリックという手法は正規分布とか等分布を仮定している、いろんな制限が多いわけですが、ノンパラメトリックの手法は分布型を問わない。しかし検出力が劣る傾向があり、計算も難しい。検出力は劣るかどうかわからない、やってみるとあまり劣らないということが多いです。計算が難しい、どちらにしてもソフトにいければ簡単です。ただ、多変量とかですね複雑なことができません。それとパワーの計算もできません。ですから、研究計画がうまく書けません。そういうふうなことで、なかなか皆さんあまり使われないようですね。それとnが30を超えるような場合、教育学の研究だと大体nが100とか1クラス全部とかやりますから、30なんか超えることが多いのですけれど、だったらパラメトリックの手法でも頑健性(robustness)というのがあって、大丈夫という話しになります。

スライド13 p値とEffect Size, p値というのは帰無仮説が成り立つ確率ということですね。これが0.5よりも小さければ、帰無仮説が却下されて、棄却されてということですが、Effect Size というのは、差が大きいかどうか平均値を見たときその差が大きいかどうかを見ていきます。基本的にはstandard deviation(SD)との比率がstandard deviation(SD)分、1コずれていますよということEffect Sizeは1みたいなのが大体の目安ですね。Effect Sizeが0.2ズレているというのはsmall, 0.5だとmedium, 0.8だとlarge, というこんなちよつとした指標も出ています。特にこの教育学の領域では、Effect Sizeを非常に皆さん重要視されまして、これって要するにNNTが大きいかわ小さいかわですね。NNTが小さいということは、Effect Sizeが大きいということですから、医学での疫学における議論と似ている部分があります。p値というのは、サンプルサイズが大きければ有意になることがありますから、Effect Sizeが0.1ぐらいでも有意差ありという判断もありうるわけです。サンプルサイズが小さくても有意に出る場合、非常にEffect Sizeが大きいということがいえま

スライド14 分散分析, ANOVA ですね。まず重要なのは要因をみた時に反復測定の有無を間違えない、反復測定というのは、繰り返しのあるデータかどうか、こういうふうな先程森本先生が言っていたデータセットがうまく組んでいるかどうか、これをみると大体その人が分析に慣れているかどうか分かります。たとえばこれですね、評価者1の測定と評価者2の測定を同じID1の学生に2回測定しているわけですが、これは反復測定をしているわけです。そういう評価者1の点と評価者2の点を比べるという分析をやろうと思うと反復測定あり、という要因になります。もう一つ男女の要因がここにありますが、男女の要因は繰り返しがありませんから、縦に積み重ね

ています。こういうふうに繰り返しのないデータは縦に並べて、繰り返しのあるデータは横に並べる、ということになります。まず4つの測定を全部評価者1と2に分けてプロットしてみるとこういうふうになるわけですがそれを男女でまとめるとこういう感じになります。これが ANOVA の一番の基本形です。交互作用なしとかありとかということが分かります。これが交互作用なしの例になります。

スライド15 こういうふうに 2way、3way の ANOVA をやりますと同時に2つ以上の要因の比較ができます。繰り返しはあってもなくても組み合わせることは可能です。ただし、繰り返しのある場合は、球面性(sphericity)ということが問題になります。この辺の議論に興味のある人いますか。ああ訳分かんないですか、はい いいですね。二元配置、三元配置の場合はこういう交互作用というのを確認する必要があります。評価者1が測定したのと、評価者2が測定したのが平行になっていれば交互作用なし、全然逆の測定をしちゃうというのが交互作用あり、といういい方になります。あんまり、医学研究とかで ANOVA やっている、ANOVA が得意だという人はあんまり聞いたことがないのでこれ議論すると皆さん大体ああ〜という顔をしています。

スライド16 研究計画法というのが ANOVA の複雑なデザインを組むときに非常に重要になります。3要因で要因1があるかないか要因2があるかないか要因3があるかないかこれを組み合わせるとグループ8つに分けないとうまくいかない。グループ8つというのが、もし 80 人学生がいたら 10 人ずつのグループに分けるとのことですよ。これをランダムに分けるとなるとなかなか大変ですが、こういうふうにとると、要因1と2と1と3、3、..こういう交互作用なんかもすべて求めますが、3つの交互作用というのは何に意味しているのかはよく分からなくなってきました。この場合はいずれも繰り返しがなしのパターンですね。

スライド17 カテゴリー間、二変数間の関係:ここまで ANOVA とかそういう差をみる検定をやってきましたけど、ここからは少し相関とか回帰の話になります。二区分変数がたくさん出てきてその変数間の関係をみようというときには、まず、名義変数と同様に表に書いて、これをクロステーブルといいますけれども、お互いの関係を見るということが大事です。独立性は  $\chi^2$  検定で分かりますし、要因による差というのは、odds ratio あるいは relative risk みたいな形で出てきます。順序変数、連続変数だったら、まず、散布図を書くと、ほとんどそのクロステーブルと同じような意味合いですけれども、散布図を書いてこちらがX軸、こちらがY軸でプロットしていったら、大体こう右上がりだとか右下がりだとかをみることで相関しているかどうか分かります。一応順序変数だとピアマンで、連続変数だとピアソンということになりますが、これをパラメトリック、ノンパラメトリックの議論なのでどっちでもいい、ということにもなります。3つ以上の変数があっても全部関係見たい場合は多変量解析に行くことになります。

スライド18 多変量、どうやるんでしょうね。という質問になってくるとこれ非常になんとなく哲学的な

問いになることもあるんですが、一番大事なのはまずこのたくさん変数があって、探索的に何かと何かの間に関係がないかな…みたいなですね、そういうものをやるのか、あるいは自分であくまでも仮説をもって、これ証明したいんだというふうに行くのか、まずそこからですね。ときどきこれの区別がつかずに闇雲に有意になるのがありませんかと持ってこられる、最近はないと思いますけど、昔そういうのがありました。この辺の議論をするためには従属変数は何か、独立変数は何かみたいなことが明確になっているかどうかですね、ただときどきこれに対してはこれが独立変数だ、これに対してはこれが従属変数だと、思っているけどまだ次のステップにいくと、関係が変わることは十分あると思います。これらによって、関連性と因果関係との違いみたいな話になってきます。何が原因で何が結果なのか、このときにデータの種類によっていろいろ解析の仕方が変わってきます。ちなみに解析の細かい話はしません。

スライド19 それから内的一貫性の話ですね、信頼性というか、質問紙を作ってみたときに、例えば1番～5番というか、a～eの項目が同じような内容を質問しているかどうか、ということが信頼性とか内的一貫性の話です。よくアンケートみると同じ様な項目訊いているなあ、でもちよつと違うんだけどなあ、というふうなものも並んでいる場合があります。それは信頼性を上げるためにわざと似たような質問をいっぱい並べているのです。ときどきこういうdみたいにまったく逆の質問を入れておいて、答えが5—5—5—5—5と並ばないように5—5—5—1—5になるように工夫されてることが多いです。alpha 係数、先程 0.8 位がいいという話がありましたが、0.9 を超えてくるとあまりにも同じような質問ばかり並んでいてトートロジー(tautology)というかそれはやりすぎだよということになってきます。こういうふうにdをマイナスの方にもっていった一つの方向にまとめあげる。こういうふうにやると、たとえば5ポイントのライカート・スケールで5つの問題に対して 25 点が満点で全部足し合わせてもいいというふうなことを alpha 係数を求めて、0.7 とか 8とかの結果が出たら、よし出してやろうということになります。

スライド20 こういうことに興味持っておられるかどうかわかりませんが、今教育とか心理の中で一番注目されているのはこの共分散構造分析というか structural equation models というもので、SEM という言葉が出てきます。確認的因子分析といって、昔は因子分析というと非常に探索的なものというふうなイメージだったのですが、もう因子の数が決まっている。それでうまく適合するかどうかというふうなことを確認する分析が今非常に流行っています。この辺僕も書いてきただけで分かりません。ほとんどやったことはありません。因果関係がはっきり議論できるというようなことが一番の特徴だといっていいと思います。パス(path)解析も因果関係で議論しますがSEMには潜在変数というのをおくとされていますが、こんなものらしいですね。

スライド21 小学校の国語算数、中学校の国語算数英語の点数とそれから小学校の学力中学校の学力、動機づけみたいなものを全部纏めたモデルを作るとこういうふうな潜在変数が3つくらい出てきて、それぞれは最終的に中学校の実力に結びついていくと、動機づけはここで重要

な役割を果たしている、そういうスタディ法を組んだみたいです。

スライド22 それから先程から信頼性、妥当性の話がいっぱい出てきていました。この次ですね、small group discussion でたとえば質問紙使うというグループ、質問紙を実際に書いて欲しいと思うのですが、評価者間の一致度とか一致率、 $\kappa$  値、ICC (intra class correlation) こういったもので一応表すことになっています。ただ一致率と  $\kappa$  値あるいは一致率と ICC のどっちか2つの組み合わせでなければ一致率の判断は難しいということになっています。とくにその教育学のなかで使われる評価 OSCE なんかであれば、“挨拶ができる”みたいな項目は非常に一致率が高そうに見えるんですけど、ほとんどの人は挨拶しますから単に評価者の一致率だけみたら99%一致しています。ただ一人がずれると  $\kappa$  値0ということもありえるんですね。そういうふうなことで  $\kappa$  値が使いにくい状況があるというふうな論文も私と一緒に研究している人が出したことがあります。それから今注目されているのは一般化可能性理論を用いたこういう係数ですね。信頼度指数というのは、それとほとんど似たようなものですけど前者は相対評価、後者は絶対評価に利用されることになっています。統計学的には一般化可能性理論が分かっていたらクロンバック  $\alpha$  とか ICC というのはその延長上にあるというかその中に包含されるような関係です。

スライド23 回帰分析、ほとんど触れませんが単回帰、あるいは重回帰というような形で独立変数によって従属変数の予測式をつくるのが回帰分析のやり方です。それぞれが意味のある独立変数なのかどうなのかというふうなことですね。間隔変数が従属変数なら線形回帰分析ですし、従属変数が二区分だったらロジスティック回帰ということになります。ここがカテゴリーだと悩ましいのですが、間隔変数とみなせれば線形回帰にすることもあるし、むりやり二区分にしてロジスティック回帰とかあるいはカテゴリー変数の非常にややこしいロジスティック回帰を用いるプロビット(probit)分析を用いるという方法もあるようです。あんまり詳しいことは訊かないで下さい。

スライド24 SPSS では、こういうふうな Analyze の種類がざーっと出てきます。明日統計解析をするときには一応、この辺に SPSS を使えるコンピュータを置いておいて、どうしても使いたい人が利用できるようにしておきたいと思います。

スライド25, 26 便利な Web をあげてみました。僕が使うのはこの辺ということでもっともっとほとんど統計の勉強というのは Web の上でできちゃいますから、そういうのを是非利用して頂けたらと思います。

あまり時間がないので、ここでは質問を取らずにいきたいと思いますが、次のですね 40 分間くらいのところ、もう少し、具体的な測定の内容とかあるいは実際に質問票を使うのであればそれを実際書いてみて、それがどのくらい難しいのかとかあるいは信頼性のあるものにするというのが、どういう意味なのかそういうふうなことを是非知っていただけたらと思います。課題についての質問は

ありますか？スモールグループの、大体やることはこんな感じかなということ見えていますか？それぞれタスクフォースが行って説明するということになると思います。それでは宜しく願い致します。  
6時 20 分までです。