

第4回 医学教育研究技法ワークショップ 記録

日時：2008年10月30日（木）13:00-10月31日（金）12:00

場所：東京大学医学部総合中央館3階333号室

テーマ：「質問紙調査の質を上げる」

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：①医学教育研究において適切な質問紙を用いた研究計画を立案できる

②質問紙を用いた調査で得られたデータのまとめ方を説明できる

2008年10月30日（木）

時間	内容	担当
13:00-13:30	開会挨拶・自己紹介	志村副委員長
13:00-14:00	レクチャー1：医学教育における質問紙調査	大西委員長
	グループワーク1「質問紙作成」の説明	
14:00-15:00	グループワーク1：質問紙作成	
15:00-15:30	休憩，USB回収・印刷	
15:30-15:50	質問紙回答後，回収→各作成グループへ返却	
15:50-16:30	グループワーク2：他グループによる回答結果の比較	
16:30-17:15	グループ発表と討論① 発表要領（作成経緯，集計，問題点，修正点，調査の方向性）	渡邊委員
17:15-17:45	レクチャー2：質問紙によくみられる問題点（質疑応答含）	杉本委員
17:45-18:30	グループワーク3：質問紙の改良	
18:30-19:00	グループ発表と討論②	吉田委員
19:00-	懇親会	

2008年10月31日（金）

時間	内容	担当
9:00-9:30	レクチャー3：質問紙調査の実施前後における問題	森本委員
9:30-10:00	レクチャー4：質問紙調査を用いた医学教育研究の実際	小田委員
10:00-10:20	質問紙回答	
10:20-11:00	グループワーク4：他グループによる回答結果の集計と省察	
11:00-11:45	グループ発表と全体討論③	守屋委員
11:45-12:00	まとめ・閉会挨拶	大西委員長

■ 開会挨拶…………… 志村副委員長

そろそろ始めたいと思います。皆さんこんにちは。第4回の医学教育研究技法ワークショップを始めます。私はこのワークショップの副委員長をしています日本医大の教育推進室の志村と申します。宜しくお願致します。今回のテーマは「質問紙調査の質を上げる」ということを取り上げました。質問紙調査なんて誰しも1回や2回はやっているのですが、簡単だよという声があるかもしれませんが、意外と難しいですね。本当に質のいいものを作るのは、そしてこの古くて新しい質問紙調査の質問紙はすごくいいものを作ることは、フィードバックの基としても大事だし、そしてまた医学教育の質を高めることにもなりますし、またリフレクションの基にもなります。だから本当にいいものを作るのって意外と難しい。この用紙をもとに統計学的なこともやらなくちゃならないし、本当に難しい。だから、このワークショップで皆さんで、いい質問紙の作り方、プロダクトを作って頂きたいと思います。

■ 参加者・タスクフォースの自己紹介

■ レクチャー1：医学教育における質問紙調査 …………… 大西委員長

スライド1：質問紙の作り方

(引き続いて)私、大西がバトンタッチしたいと思います。(自己紹介がまだなので、その辺りから始めますけど、)こちら東京大学の医学教育国際協力研究センターの方に3年前から仕事をさせてもらっています。そうですね、質問紙とか医学教育研究という意味では、佐賀に移った97年の頃から、OSCEとかですねあるいは患者満足度の調査、これは明日お話しがあると思うんですけど、佐賀大学の小田先生が中心になってやっていたのを側面的にサポートしてきたというふうなことがありました。私自身の専門は総合診療、医学教育国際協力ということで、段々手が広がりすぎて訳分からなくなってきておりますけれども、質問紙の作り方あるいは **psycho-metrics** ということに関しては97年当時からずっと関心を持ってやってきていますけれどもほとんど独学ですよ。あまり系統的な教育をやってくれるところがないということは98,99年くらいから大体分かってきて、その頃から心理統計メーリングリストとか東大の先生方とか、教育学・心理学の方が中心となってやっているそういうメーリングリストの中で色々勉強しながらやってきたというような状況です。質問紙だけでおそらく2時間位はいくらでもお話できるのではないかと思うのですが、今日は30分でかいつまんでやります。

スライド2：質問紙作成の手順としては、測定目的の明確化、研究デザインの選定、項目作成、信頼性・妥



質問紙作成の手順

1. 測定目的の明確化
2. 研究デザインの選定
3. 項目作成
4. 信頼性・妥当性

当性というふうなところを選んでお話しします。

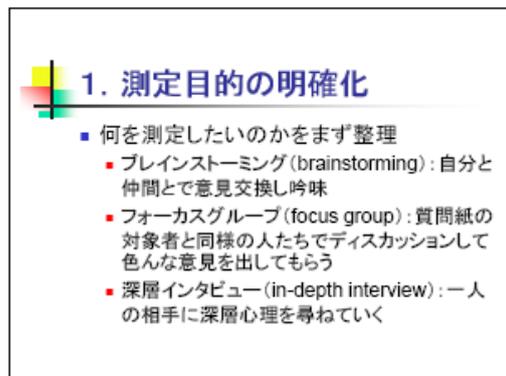
スライド3：まず、測定目的の明確化というところですね。大体ちょっと一晩寝て考えてみようという感じになると、一人で考えているというのはあまりよろしくないですね。大体偏った方向に行くことが多いです。誰かと仲間と意見交換してブレインストーミングして、というところが一番最初のところになるかなと思います。

大体方向性が決まって、今度は対象者に近いような人たちでディスカッションしてもらうようなフォーカスグループをやるといのが最近好まれている手法かなと思います。これによって大体どういうふうな意見の広がりがあるかというあたりをつけて、それをアンケートに落とし込んでいく。あるいはもう少し深層心理に関しての内容であれば深層インタビューという方法もあるのですが、ここまで使うのならもうアンケートじゃなくて、質的な方法をとというような形になることが多いので、そう意味ではフォーカスグループとアンケートというのはわりとセットとしてよく出てきます。

スライド4：構成概念という考え方があります。

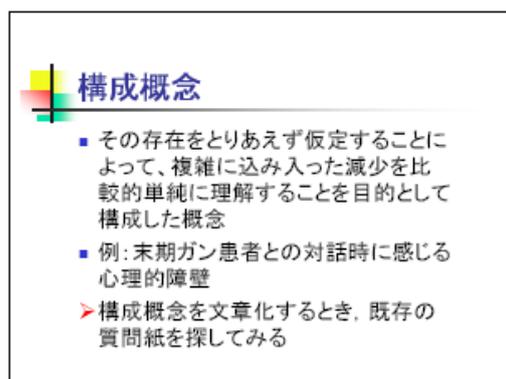
この辺になってくると、皆さん、もしアンケート作ろうと思って何か本を開いて、何だ?この構成概念って、というところに当たった人も多いかと思いますが、(構成概念とは) とりあえずその存在を仮定することによって、複雑に込み入った現象 **phenomenon** を比較的単純に理解することを目的として構成した概念である。何か分かったような・分からないような感じですが、

例えば(今回の参加者では)救急のトレーニングの先生が多かったのでそんな話をしてみると、**ACLS**のトレーニングによって今まで道端で倒れた人がいたら、なんとなく逃げたいなと思ったのが、それが逃げたくなくなりました。そういうことを訊きたいのであれば、道端で行き倒れに対しての恐怖感といいますか、あるいは助けに行くのに対しての危機感というか、そういうものがどうも構成概念なんだろうなという、そういうものをまず自分で仮定してみるわけです。例えばここでは末期がん患者との対話時に感じる心理的障壁というふうなことを挙げてみました。ここです、構成概念を文章化するときに、自分はこれが調べたいのだということが大体まとまってくるわけですが、そこでまず既存の質問紙を探してみてください。誰かがもういいものを作っていて、自分が一から作るというのは非常に無駄です。後ろの真ん中のところに参考図書が7冊ほど置いていますが、そのうちの4つは既存の質問紙についてあげているものです。そういう事典みたい



1. 測定目的の明確化

- 何を測定したいのかをまず整理
 - ブレインストーミング(brainstorming): 自分と仲間とで意見交換し吟味
 - フォーカスグループ(focus group): 質問紙の対象者と同様の人たちでディスカッションして色んな意見を出してもらう
 - 深層インタビュー(in-depth interview): 一人の相手に深層心理を尋ねていく



構成概念

- その存在をとりあえず仮定することによって、複雑に込み入った減少を比較的単純に理解することを目的として構成した概念
- 例: 末期がん患者との対話時に感じる心理的障壁
- 構成概念を文章化するとき、既存の質問紙を探してみる

な本がありますので、そういうものが大分役に立つと思います。

スライド5：もう一度ですね、では質問紙使うべきなの？という質問をやっておいたほうがいいのですけれど、心理学的研究のなかには観察法によるものと言語を媒介にするものがあり、言語を媒介にするものの中にも質問紙法というのは1つであるということですね。

検査法というのは、これは質問紙にかなり近いのかもしれませんが、ロールシャッハ(Rorschach test)というものあってあれは質問紙ではないかもしれません。この質問紙の中には、心理尺度と社会調査というのがあります。我々はこれを両方含んで今日議論しています。

スライド6：質問紙の長所と短所です。長所としては、個人の内面を幅広く捉えられる。対象者のペースでじっくり考えて回答可能である。多人数に同時に実施可能、条件統制容易、費用が比較的安価というところがあります。全国に調査するなんていうのを観察法でというわけにはいきませんから、そういうときにはアンケートをと自動的にになります。但し内面の深い把握は難しいですし、内面を偽るなどして防衛的になる可能性、特に記名式にするかどうかというところで、この辺の問題がいっぱい出てきます。それから年齢や教育歴などに依存するということがあります。先日色々また質問紙をざっと洗い直していたら、小学校6年と中学校3年が今全国学力検査やっていますね。あれにまつわる質問紙が出てきたのですが、ひらがなであるということもひらがなが多いということもあるのですが、実に分かり易い。やはりぱっと見た時にどういうふうに答えていいのかということが構成上すごく良く分かるようになっていました。そういうもの、やはり既存のものを参考にするというのは、そういう意味でもいいのだなと改めて思いました。

スライド7：質問紙とテストの類似性

それから質問紙の話をするとうテストのときも結構同じようなことをやっているよな、ということに気づく人もいるかなと思います。いずれもある構成概念に対して用いられる操作的な測定方法であるという意味では一緒な訳ですね。例えば、ACLSに必要な救急のプロトコル



- 続(1978)による心理学研究の分類
 - 観察によるもの
 - 言語を媒介とするもの
 - 面接法
 - 検査法
 - 質問紙法
 - 心理尺度: 人間の意識や行動を測定
 - 社会調査: 社会に生起する事象の実態把握



- 長所
 - 個人の内面を幅広くとらえられる。対象者のペースでじっくり考えて回答可能
 - 多人数に同時に実施可能。条件統制容易
 - 費用が比較的安価
- 短所
 - 内面の深い把握は難しい
 - 内面を偽るなど防衛的になる可能性(特に記名式)
 - 年齢や教育歴などに依存



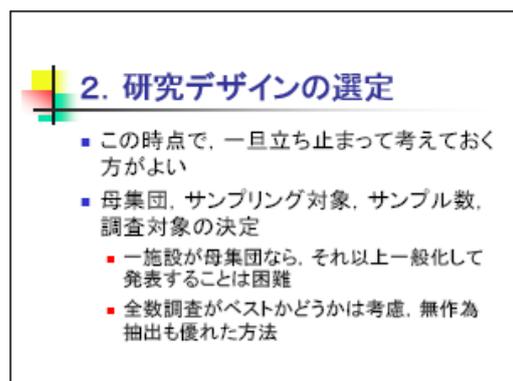
- いずれもある構成概念に対して用いられる操作的測定方法
- 質問紙: ある集団に対して、測定値がどのようであるかを確認する手法。集団間での比較、時間軸での比較なども可能。
- テスト: 測定値により、習得レベルを測定し、フィードバック情報を提供する。テストでの測定内容を詳細設計→blueprinting

を暗記しているのかどうかという話しになると、それはテストやっているのかアンケートしているのか、ほとんど似ている訳ですが、質問紙というのはある集団に対して測定値がどのようなかを確認する手法で、集団間での比較、時間軸での比較なども可能。でもこれもテストでも可能ですよね。テストというと測定値によって、習得レベルを測定。この習得レベルというあたりがいかにも教育的なところでしょうけど、フィードバック情報を提供することが目的である。テストで測定内容を詳細設計するのが **blueprinting** とよんでいますが、**blueprinting** と同じようなことを質問紙でもやるというふうなことを教育関係の人だったらですね、思い浮かべると質問紙で何をやらないといけないのかということがみえてくるのかと思います。

スライド 8 : ここで、あえて研究デザインの話しをちょっと持ってきています。要するに質問紙を作るんだと、この辺の構成概念を測るんだというふうなことが大体まとまった時点で、では自分はというふうな形で質問紙を使って調査したいのか、あるいは研究をしたいのか、ということ一旦考えた方がいい。自分が測定しようとする母集団はどのグループで、サンプリングの対象はどうか、サンプル数がどうか、調査対象はどうか、というふうなことがイメージできると、測ったものが、どのくらいの代表性があって、どこまでは言える、どこまでは言えない、ということは自ずから分かってくるわけです。例えば、自分のところでもいいプログラムができた。そこでプログラムをやって、前後比較して効果があった・なかったという話しをしたときに、あくまでもある大学であるいはある病院でその調査をやったら、大学なり病院なりが母集団で、それが全国に広げられるのかということかなり大風呂敷になってしまうかもしれないわけです。では全数調査をした方がいいのだろうかという話にもなって、例えばその病院での教育に我々のこのプログラムが役に立つのかどうかということで、例えば全国にアンケートしようということを思いついたら、必ず全数でやらないといけないんだというふうな思いに囚われている人がいますが、そうではなくてサンプリングがうまくできるかどうか、そこで無作為抽出ができるのかとか、そういうことで全数調査をすべきかどうかということをお考えた方がいいわけです。

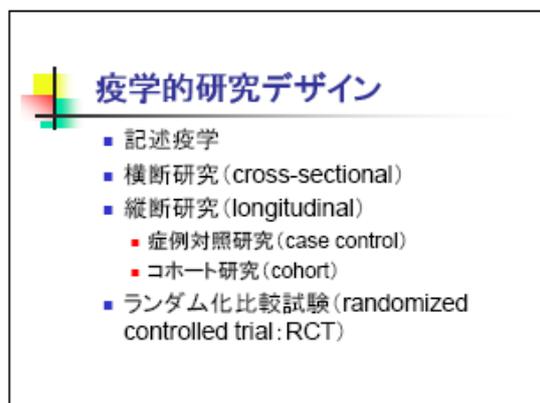
スライド 9 : 疫学的研究デザイン

疫学的な研究デザインとしてこのようなものがあるというのは皆さんご存知なんだろうと思います。そのデザインをお考えた上でアンケートに落とし込んでいく方法も考えることになります。



2. 研究デザインの選定

- この時点で、一旦立ち止まって考えておく方がよい
- 母集団、サンプリング対象、サンプル数、調査対象の決定
 - 一施設が母集団なら、それ以上一般化して発表することは困難
 - 全数調査がベストかどうかは考慮、無作為抽出も優れた方法



疫学的研究デザイン

- 記述疫学
- 横断研究 (cross-sectional)
- 縦断研究 (longitudinal)
 - 症例対照研究 (case control)
 - コホート研究 (cohort)
- ランダム化比較試験 (randomized controlled trial: RCT)

スライド 10 : 教育研究のデザインとしては、先ほどと似ている面もありますけど、違う面もあります。この辺の考え方というのは特に教育学・心理学の研究でよくいわれていることで、ランダム化をすれば、真の実験的な研究だけれども、比較群を置かないとやはり前実験的なモデルになってしまうと。よく事前事後テストをやるということは必須だという話になると思うのですが、事前事後だけやって、比較群がないと実験的なデザインではないわけですね。実験的なデザインではないとやはりかなりもの言える範囲が狭まります。我々がやった教育がいいんだということをこの辺のテストだけでいうのはかなり難しいということが分かります。この事前事後テストだけで、比較群がないときにどういうコントロール・どういうふうな条件があれば因果関係がいえるのかとか、そういうふうなことについてももっともっと深い議論がいろいろできるんですけど。それからランダム化しない比較群をおくというのもこれもかなり準実験というふうなデザインにしかありません。

一般的な教育研究デザイン

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

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

スライド 11 : それから先程縦断研究、要するに事前と事後両方測定するこの辺のデザインですね。この3つのデザインに関しては、縦断研究ということになりますが、教育介入などの結果を測定する場合に、縦断研究的な変化を捉えられるタイプの内容になっているかどうか。例えば「末期ガン患者の話しかけるのに障壁を自覚したことがありますか」、「末期ガン患者の話しかけるのが怖いと感じていますか」というのはちょっと質問の仕方が違います。障壁を自覚したことがありますかというのと、とりあえず今あるといっても、何かトレーニングしても自覚したことは以前あるわけですから、あるわけですね。こっちだと今怖いと感じている、でも教育介入をしてあまり怖いということはあまり感じなくなった、という変化をみることができます。ですから、こういう経験を問うのか、現状認識を問うのかのあたりで、下の方でない縦断研究には耐えられないというふうなことが分かります。よくできた質問紙というので、とくに教育学で使われるのは必ず縦断研究に耐えうるものになっていますよね。

縦断研究と項目の尋ね方

- 教育介入などの結果を測定する場合は、縦断研究的な変化を捉えられるタイプの内容を問うているかが問題に

例1: 末期ガン患者に話しかけるのに障壁を自覚したことがありますか (経験→変化しない)

例2: 末期ガン患者に話しかけるのが怖いと感じていますか (現状の認識)

スライド 12 : 研究デザインの話が出てきたのは、変数をどのようにするのか、アンケートを答えてもらって最終的にどういうふうにデータをま

変数の種類

- 名義変数
 - 東京都-神奈川県-埼玉県-千葉県
- 二区分変数
 - Yes-No, 合格-不合格
- 順序変数
 - 通信簿の1-2-3-4-5
 - 全くなし-少し-多少-かなり-非常に多い
- 間隔変数
 - 授業の出席回数

とめるのか、というあたりにも関わってきます。名義変数だと統計的な解析をするときに結構いろいろ難しい場合もありますし、二区分だとやりやすいという面もありますが、あまり差が出ない。カテゴリカルにある程度段階をつけておいた方が差がみやすいとかそういうこともあります。ただカテゴリカルで1・2・3・4・5とかこういうふうにやっていると、これを連続変数化しない場合には統計学的な解析、特に回帰分析とかは難しいとか、そういうのが出てきますよね。間隔変数だといいいのですけど、教育心理系で間隔変数が純粋に出てくるのはかなり少ないです。ですから何らかの変数変換をやって初めて連続変数になる、間隔変数になるということはよくみられます。

スライド 13:結果がどういふふうな変数で表されるか、によって統計解析が変化するという話です。例えば5件法の項目が10個あって、その合計得点を連続変数にしていいということに関しては、カテゴリ変数を連続変数に変換するわけですが、10項目の内的一貫性が一定以上の α 係数が0.7以上、0.9以下が良いとかいろいろありますが、それが高い信頼性を求める一つの理由になっています。この辺の議論を森本先生とやっていたときに0.9を越えるというのは、要するにredundantな項目が多すぎて、項目減らしてもあまり変わらないという意味なのだろうということですね。ただそのときに因子分析とかをやったとどういふふうな形で項目がばらついていのかとか、本当に同じようなことしか訊いていないのであれば、それは数を減らしていくとかその辺の工夫をどういふふうにやるといいのか、あまりいいガイドラインもないのかなというディスカッションをしていました。

スライド 14:3. 次、項目作成です。項目の数が全体的にどのくらいになるのか、時間はどのくらいかかるのかというのをイメージして全体構成を考えてください。

スライド 15:回答方式の選択。このあたりでよく質問が出てきます。2件法や3件法の方がいいのでしょうか。要するにOSCEとかをよくしている方だと、そのチェックリストというのは非常に信頼性が高いとか、そういうふうに分けると勘違いをするというのがあります。勘違いというのは、観察してやっている・やっていないと判るのはチェックリストがいいし、ある人に認識を問うようなものだと、カテゴリカルにある程度段階をつけ

変数

- 結果はどのような変数で表されるか—これによって用いる統計解析が変化
- 5件法の項目10個からなる質問紙を使い、合計得点を連続変数として利用することが可能な条件
 - カテゴリ変数→連続変数への変換
 - 10項目の内的一貫性が一定以上(α 係数0.7以上など)→高い信頼性を求める理由

3. 項目作成

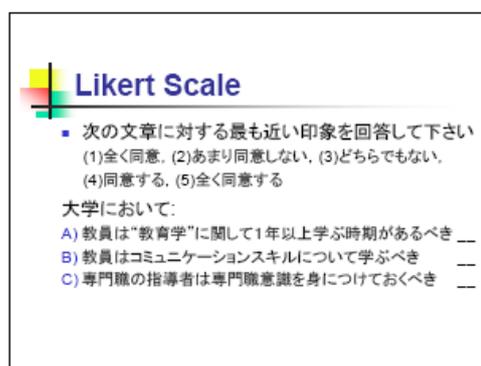
- 項目の数が全体としてどの程度になるか、時間はどのくらいかかるかをイメージして全体構成

回答方法選択

- 2件法、3件法:回答に迷い生じにくい
- 評定尺度(rating scale)
 - 定量化、連続変数化したいならこれがよい
- 多肢選択法:属性など
- 順位法:好きなものから順番に数字を振る

ておいた方がいいみたいなことは一般的には言われていますよね。だから医療面接でこの患者—医師関係はどうもうまくいっていきそう・うまうまいっていきなさそう、みたいなのはカテゴリカルで5件法の方が答えやすいと昔から言われています。そういうことでrating scaleというのが割と使われることが多いのですが、定量化、連続変数化したいならこれが良いという感じがしています。ただ連続変数化をどんどん促していくのは如何なものかということは森本先生から昨日突っ込みもあったのですが、それから多肢選択法ですね。先の名義変数にあたるようなものはそういう多肢選択法でやることになります。順位法、好きなものから順番に数字を振る、いろんな教育技法があってあなたが一番よく使うのはどれですか [1]、一番使わないもの、10個あったら [10] とかですね、こういうものもありますけど、どういうふうに皆さん解析するのか僕分らないので、僕はああいいう方法は使いませんが、一応アンケートとしてこういう方法も紹介はされています。

スライド 16: ライカートスケールというのは大体こういうタイプですよね。次の文章に対して最も近い印象を回答してください。「全く同意しない」、「あまり同意しない」、「どちらでもない」、「同意する」、「全く同意する」、それでこういう質問があります。教員は“教育学”に関して1年以上学ぶ時期があるべき～いろんな意見があると思いますが、こういう項目だなという感じで見ておいてください。



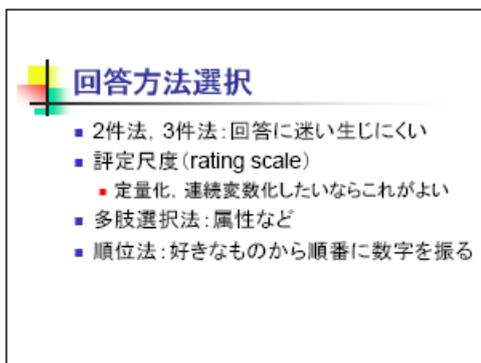
Likert Scale

- 次の文章に対する最も近い印象を回答して下さい
(1)全く同意、(2)あまり同意しない、(3)どちらでもない、
(4)同意する、(5)全く同意する

大学において:

- A) 教員は“教育学”に関して1年以上学ぶ時期があるべき __
- B) 教員はコミュニケーションスキルについて学ぶべき __
- C) 専門職の指導者は専門職意識を身につけておくべき __

スライド 17: ライカート法の件数の問題。何件法が良いかというのは様々な意見があります。件数が多い方が差を際立たせやすいということはありません。7件法・9件法になった方が信頼性は高いというデータはよく出ています。ただ、アメリカでは9件法が結構多いです。割とアメリカの人は慣れているかなという感じはしています。5～9までしかあまり使わないという・・・、3とかつける



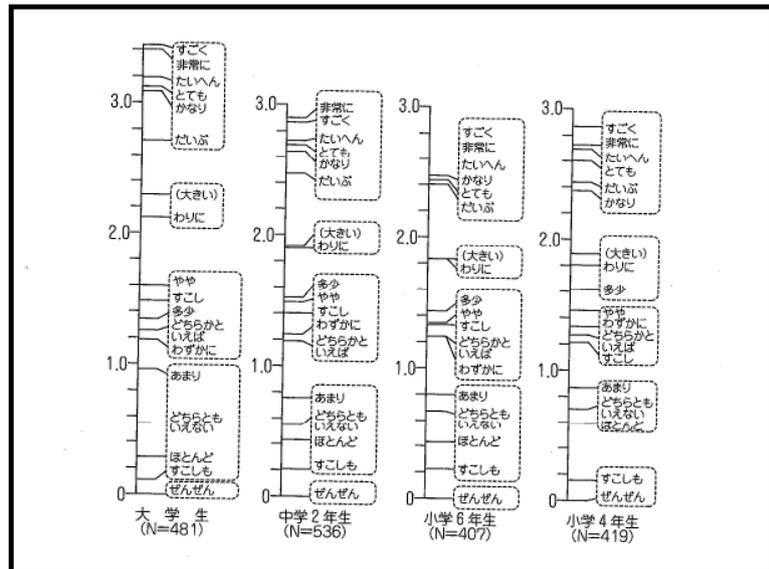
回答方法選択

- 2件法、3件法: 回答に迷い生じにくい
- 評定尺度 (rating scale)
 - 定量化、連続変数化したいならこれがよい
- 多肢選択法: 属性など
- 順位法: 好きなものから順番に数字を振る

人はいないよと聞いたことはありますが、そういうのに付けたい人も結構いるみたいですね。ただ日本でもこういう9件法に対して、下の下～上の上までいくと、3は下の上とかですね、そういうふうに捉えられなくもないので、9件法もいいかなと。ここに下の下とか、下の中とか書いておくと分かりやすいかもしれないですけどね。奇数件数にするというのが割と何となく対照的に見えていいという話もあるのですが、中心偏向性 central tendency, アジア人が特にこの central tendency が強くて、「あなたは何々についてどう思いますか？」—「はい」・「どちらでもない」・「いいえ」にすると、「どちらでもない」が多いという、そういう話ですね。ただ、9件法になってくると、あまり真ん中ということも

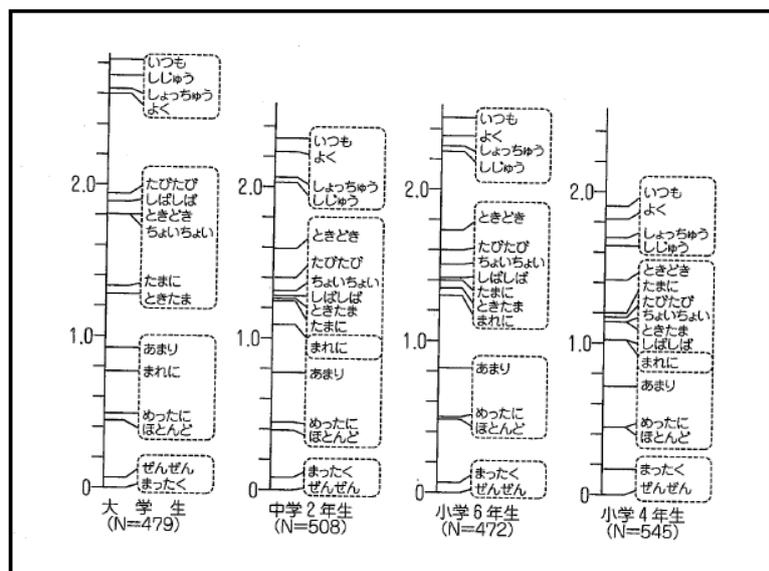
ないんですけど、それからただ数字をつけるときに0~10 みたいにするのか、-2から+2にするのか、こういうのもいろんな意見があります。それから、小田先生とずっと一緒にやってきた患者満足度の調査では元々のアメリカの回答様式が poor fair good very good excellent になっていて、どこが中間かよく分からない。こういうふうにしておくとわりと central tendency は奇数件数でも避けられるということもあるかなと思います。

スライド 18: これは参考図書でお送りした質問紙法の本の中に出てくる図ですけど、80 年位に形容詞とか頻度を表す言葉に関してどのくらいの数値的な差があるのかというふうなことを調査した人がいるんですね。だから「かなり」というのは3 ぐらいで、「割りに」とかは2 に多い、それから1 のあたりだと「あまり」という否定的表現か「わずかに」



とか肯定的な表現にするとか、そういうふうにしておけばある程度等間隔性が保障されるみたいですね。そういうふうな調査をやっていて、これは割りと使えますよね。面白いのは先程、小学生にやろうとしたら難しいという話もあったんですけど、数字的な捉え方が成長によって変わるみたいですね、そういうふうないろんなデータがこれから読み取れて楽しいです。

スライド 19: こちらは頻度に対してです。「いつも」とか、「時々」とかこういうので悩むことも結構ありますけど、これも等間隔性を持たせるためにはどうしたらいいのか。「いつも」と「しばしば」と「あまり」とか「全然」くらい4 つくらい選んでいたら大体等間隔になるみたいなのが分かりますよね。



スライド 20: 尺度変換

尺度変換の話し。間隔変数は順序変数よりも統計学的な解析がしやすいので、順序変数を間隔変数に変換すること、平均などを計算することもよく行われています。

PBL を受けてどう感じたか回答してください。

「大変同意」、「同意」、「中間」、「同意せず」、「全く同意せず」と、一個一個詳しくみていきますと、「私は PBL に熱心に参加した」、「面白いと思う」、「学んだ内容に満足している」、「私は後輩に PBL を勧めない」、ここは否定的な表現になっていますね。「もし、PBL と講義が選べるなら、私は PBL を選ぶ」、みたいな形になっています。

スライド 21：内的一貫性

この回答をみたときに、a から e に対して測定のベクトルというのがあったとしたら、どういふような方向に向いているのかというふうなことをちょっと考えてみますが、一般的にこんな感じかなと。要するに d は逆転項目なので全く逆向いていますけど、他はある程度同じような方向を向いて、でも少しばらついた方向を向いている。すべてのベクトルが同じ方向に向

いているかどうかを見たい訳ですよ。d をまずネガティブなので、逆転する、数字を引っくり返す訳です。1～5 だったので、それを 5 から 1 に引っくり返して、全体をまとめられるのかどうかとか、ということを考えます。

スライド 22：合計得点がもしクロンバッハ alpha がある程度高い、それから因子分析をしてあまり重なりがないというふうなことが大体分かってきたら、5 + 4 + 5 + 4 + 5、ここ 2 に答えていますけど、これ逆転して、6 から 2 を引いて 4 をここに足すわけですね。で 23 というのは合計得点となる。内的一貫性が低いならスコアの合算はできません。それから、回答を

みたらですね、多くの人が全部 5 をつけちゃったということになると、読まずに何となく読まずに、ずっと肯定的に答えてやれと 5, 5, 5, 5 みたいになっちゃう。これだとみんな読まずに回答しているということが分かりますよね。それが分かるように逆転項目を設けておくというのは一つのテクニックになります。

尺度変換

- 間隔変数は順序変数よりも統計学的解析がしやすい
- 順序変数を間隔変数に変換すること、平均などを計算することもよく行われる

PBL を受けてどう感じたかについて回答して下さい。

5: 大変同意, 4: 同意, 3: 中間, 2: 同意せず, 1: 全く同意せず

- a. 私は PBL に熱心に参加した
- b. 私は PBL を面白いと思う
- c. 私は学んだ内容に満足している
- d. 私は後輩に PBL を勧めない
- e. もし PBL と講義が選べるなら、私は PBL を選ぶ

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

尺度変換(続き)

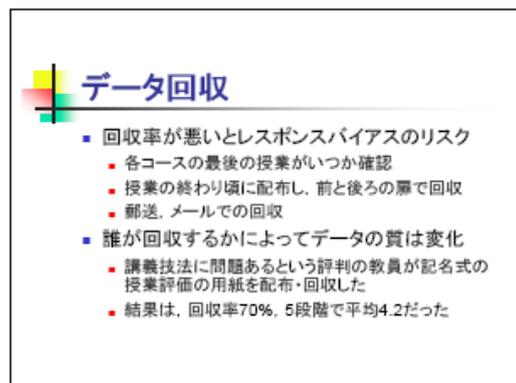
- 5 a. 私は PBL に熱心に参加した
- 4 b. 私は PBL を面白いと思う
- 5 c. 私は学んだ内容に満足している
- 2 d. 私は後輩に PBL を勧めない
- 5 e. もし PBL と講義が選べるなら、私は PBL を選ぶ

- 合計得点: $5 + 4 + 5 + 4 + 5 = 23$ (d は逆転項目)
- もし、内的一貫性が低いならスコアの合算は不可能
- 全てに 5 を付けた場合、明らかに逆転項目は矛盾

スライド 23: データ回収の率が悪いとレスポンスバイアスのリスクがかかります。僕がマレーシアにいたときには、授業評価を集めてその集計をして、というふうなことを取りまとめするオフィスにいたので、そればかりが仕事でした。ただ僕が集めに行くのではなくて集めに行くのを管理をして、回収率というのを僕のターゲットにしてですね、8割というのをどのように95%にもっていくのか、それが一番の仕事になっていました。各コースの最後の授業がいつなのか何時から始まるのか、どの先生がやっていて早く終わるとか終わらないとか、そういうことを全部チェックして間違いなくいくようにするというふうなことをやっていました。授業の終わりくらいに配布して、前と後ろの扉に係りの者を待たせておいてですね、回収するというのがあります。それから僕が行っていたマレーシアの大学では海外に実習に出て行って、その

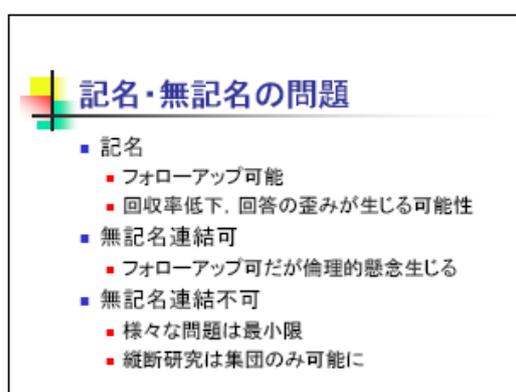
evaluation していたので郵送するあるいはメールで送る、それを回収するというのもやっていました。どのタイミングで再度督促したらいいのか、いろんなそういうこと真剣に考えたことがあります。誰が回収するのかによって、データの質が変化する。これは特に教育なんかでは大事ですね。講義技法に問題のあるという評判のある教員が記名式の授業評価の用紙を自分で配布して自分で回収しました。回収率は7割で、5段階で平均4.2だった。何も分からないというものです。だから、僕らは回収というのを必ず係りの人がやっていて、僕とかが集めに行くというのも許されないというルールになっていました。

スライド 24: 記名だとフォローアップが可能で縦断研究などはしやすい意味では記名であったほうが良い面があります。ただ回収率が低下したり、回答が歪んだりという可能性が大きくなります。一つのテクニックは無記名で連結可能な形にすると、紙の裏に小さく番号を1・2・3・4・5と番号を振っておいて、誰に1、誰に2～がいったのかが分かるようにしておく方法がありますが、フォローアップ可能ですが、倫理的懸念が生じる可能性があります。何となく騙しようになりますからね。無記名で連結不可にすると、様々な問題はあまり起こりませんが、縦断研究は集団にはできませんけど個人でフォローアップしてということはできません。



データ回収

- 回収率が悪いとレスポンスバイアスのリスク
 - 各コースの最後の授業がいつか確認
 - 授業の終わり頃に配布し、前と後ろの扉で回収
 - 郵送、メールでの回収
- 誰が回収するかによってデータの質は変化
 - 講義技法に問題あるという評判の教員が記名式の授業評価の用紙を配布・回収した
 - 結果は、回収率70%、5段階で平均4.2だった



記名・無記名の問題

- 記名
 - フォローアップ可能
 - 回収率低下、回答の歪みが生じる可能性
- 無記名連結可
 - フォローアップ可だが倫理的懸念生じる
- 無記名連結不可
 - 様々な問題は最小限
 - 縦断研究は集団のみ可能に

スライド 25：細かいこと。冒頭で、誰がどうい
う目的で実施する調査なのかとか、 どういう目
的に使われるのかは書いておいた方が（調査協
力者は）一応 “こうこうなんだと協力してあ
げよう” という気持ちになります。 こういうも
のが結構ないという質問紙とか日本では時々見
ますよね。最近ですけど、個人情報云々に関し
ては必ず考慮しており、 データは名前と連結不

可能な形で処理し、とかいろいろややこしいことを書いていますよね。 枕詞ですが、倫理
審査のときにアンケートを先に出すように言われたらこういうのがあるかどうかを見られ
ています。 それから末尾ですけれども、最後に回答してくれたら、最後まで回答してくれ
てありがとうみたいなことを書いておいたほうがいいですよ。

スライド 26：まず最初に、 親しい人とか学生に
見てもらってください。 それだけでも大体思い
込みで自分がちょっと変な文章にしているとか
すぐチェックできます。

パイロット調査を実施するというのはデータ
を確認するためなので、 ある程度数も集めて統
計的な解析もやってということになります。 ま
ず単施設でやって、 多施設でやるときには、 そ
の後で拡大するというのが一つの流れだと思
います。

スライド 27：信頼性は測定の安定性で、 妥当性
は測定したいものが測定できたかであるとい
うことになっています。 妥当性の定義がいろいろ
変わったりとかですね、 最近この 10 年ちょっと
非常にいろいろ議論があるのですけれど、 基本
はこれですね。 信頼性は高いけど妥当性は低い、
信頼性は低いけど、 妥当性は全体としていい方
向を向いているという話になりますが、 信頼性が
ないと妥当性は下がると、 要するに信頼性は妥当性の一条件であるというのは昔からいわ
れていることです。 ただ、 信頼性と妥当性というのは必ずトレードオフ trade-off の関係に
あって、 信頼性を上げようとするとも必ず妥当性が下がるということも分かっていますから、
このあたり難しいですね。

スライド 28：一般的な確認方法

一般的な確認方法。 内的一貫性に関しては α 係数を見ているんですが、 test-retest とか何か

細かいこと

- 冒頭
 - 誰がどういう目的で実施する調査なのか、 どういう目的に使われるのかは書いた方がよい
 - 「個人情報…」の記載は枕詞
- 末尾
 - 最後まで回答してくれた人に謝意を

4. 信頼性・妥当性検証

- できれば、まず親しい人、 学生などに見てもら
- パイロット調査を実施して、 データを確認
- 単施設で調査し、 多施設に拡大

信頼性と妥当性

- 信頼性：測定の安定性
- 妥当性：測定したいものが測定できたか



信頼性は高いが
妥当性は低い



信頼性は低いが全体と
しては妥当な方向

いろいろセオリーが書いてあります。併行テスト法とかですね。この辺はやっていない研究が最近非常に多いのであまりこだわる必要はないと思います。妥当性、内容妥当性については専門家にみてもらってください。これは大体いい質問が全体的にばらついているのかは専門家の目で見てもらうということです。基準関連妥当性は似た内容のものとの相関ということで、併存的妥当性と予想的妥当性ということ

で、時間的に同じタイミングでやっていれば、併存的、違うタイミングだったら予想的（妥当性）ということになります。構成概念妥当性には最近弁別的妥当性と収束的妥当性。弁別的というのは、全然違う内容に関して問うている質問紙と点数が異なるということを確認する。収束的というのはある程度似ていそうな質問表と同じような結果が出るということを確認する、ただこれを相関係数で確認したら OK みたいにしてる論文は、精神科、心理学に多いのですが、それで十分なのかという議論も結構最近あります。

スライド 29 : これはかなり高度な内容ですが、評価に関しては採点結果がどういうふうに使われて、それが教育的にどういう意味を持つのかということも妥当性に含めようという議論が起きてきています。それから構成概念とは何か、ということに関してもいろんな議論があるようですけれど、あまり良くわかりません。

内的一貫性、内容妥当性の確認プロセス、それから弁別的妥当性と収束的妥当性を提示して、後は作った質問紙がどのように広がっていったというふうな流れです。多くの人が使ってみていいんだったらいいんだ、みたいな最終的なことが妥当性の行き着く結論なので、自分が作ったものが広く使われていると思えば、それはかなり妥当性が証明されているといっているんだろーと思います。

スライド 30 : データ入力と解析。これは僕がマレーシアでいつも使っていた様式ですけれども、こういうのに名前とか番号とか色々入ってきますね。ここに回答が出てきます。

一般的な確認方法

- 信頼性
 - 内的一貫性: α 係数
 - test-retestなどはそれが意味あるときのみ
- 妥当性
 - 内容妥当性: 専門家の意見など
 - 基準関連妥当性: 似た内容のものとの相関、併存的妥当性と予想的妥当性
 - 構成概念妥当性: 弁別的妥当性と収束的妥当性の確認がよくなされるが...

妥当性に関する議論

- 評価に関しては、採点結果がどう扱われるかなども含めて、妥当性の議論あり
- 構成概念とは何かについても再度議論
- 内的一貫性、内容妥当性の確認プロセス、弁別的妥当性と収束的妥当性を提示し、後は作った質問紙が使われる中で妥当性が証明されるという考えでもよいかも

データ入力, 解析1

INTERNATIONAL MEDICAL UNIVERSITY
UNIVERSITI PERUBATAN ANTARABANGSA
MELAKA

PMIS EVALUATION OF STUDENTS

Student name:	IMU student no:		
University name:			
Date of transfer:	Period:		
Current semester:	Department:		

Please assess the above student in comparison with other students at the same stage. Kindly circle the appropriate number.

	Major Problem	Minor Problem	Satisfactory	Highly Satisfactory
	1	2	3	4
History taking				

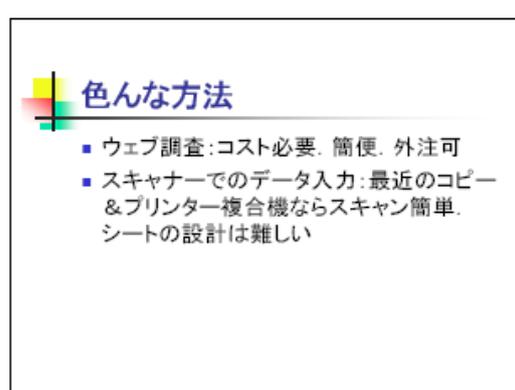
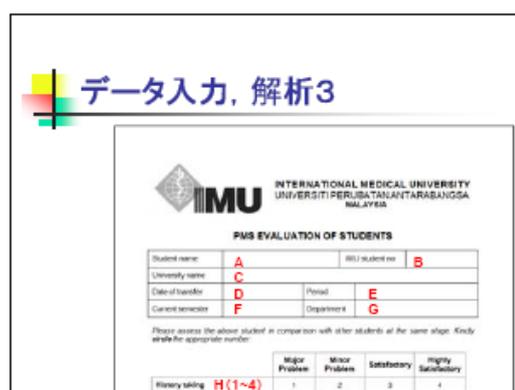
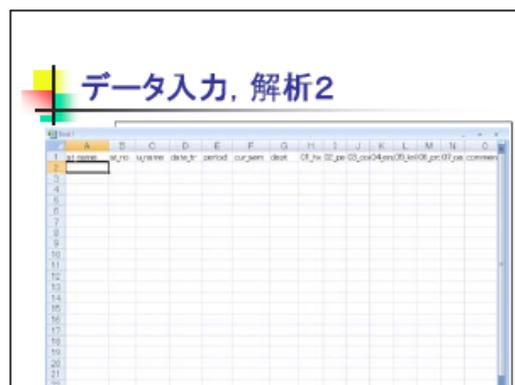
スライド 31 : 解析する前に入力しないといけないのですけど、こういうエクセルの用紙みたいなものを作りますよね。一般的には、エクセルの表で一行目に ID がくるかあるいは名前がくるか、2 行目 No.がきて、後、大学の名前とかいろいろそういうのが並んでいます。

スライド 32 : こういうふうに ABC と並んできたときに僕がいつも励行していることですが、質問紙に ABC ってこういうふうに赤マジックで振っていきます。要するに一枚は入力用に必ず使うわけです。これを一個作っておいて、返ってきたらこの様式に一番上にこれをかぶせておきます。そうすると、他の人に入力をお願いするので、そういうときに大体迷いなくいける訳ですね。この ABCD というのは先程のエクセルのカラムのここに対応している訳です。なので、どこに入力したらいいのかという迷いがないけます。逆に言うと、この ABCD というのがざっとみたときに入力忘れないように、例えばこれだったら、この用紙のある程度効率的な利用を考えて B, E, G とかこういうところにデータが入りますけど、こういうところにあるのを忘れて A の次にここにある数字を入れたりしやすいんですよ。ですから、そういう意味で構成をどうするかというのもこれを先に作っておけばある程度入力する人のことも考えて構成することができます。

スライド 33 : 入力に面倒だということを考えると、ウェブ調査という方法もありますし、スキャナーによるデータ入力という方法もあります。ウェブ調査はコストが若干かかってきますが、簡便で、外注受けてくれるところも増えています。スキャナーでのデータ入力は、最近のコピー&プリンタ複合機なら、スキャンは簡単に出来ます。ただ、シート設計は難しいのとソフトにあわせてうまく合わせていくという作業が必要なので、そんな簡単ではないですけど。ただ、何度も何度も授業評価をやっ

て同じ用紙ばかりを使うんだといえればそういうのも考えてもいいかもしれません。

以上です。どうもありがとうございました。2, 3 質問を受けます。いかがでしょう？



色んな方法

- ウェブ調査: コスト必要. 簡便. 外注可
- スキャナーでのデータ入力: 最近のコピー & プリンター複合機ならスキャン簡単. シートの設計は難しい

質疑応答

- Q: 質問紙の作り方ということなのですが、質問をとった対象に対してどのように結果を返すとか、そういうことは前もって何かしらの形で入れておいたほうがよろしいのでしょうか？
- A: 質問紙そのものにはスペースも限られているので難しいと思うのですが、それは考えておくべきことだと思うんですね。授業評価などでもそうだと思います。
- Q: アンケートの回収率のお話が出ていましたけど、一般的には回収率はどれくらいあれば妥当ですか？
- A: その結果を次に統計学的な解析にやっていくとかの話になってくると、かなり高い方がいいと思いますが、そうではなくて単に記述統計で話を持っていくのであれば、回収率がいくらでこういう結果なんです、ということを書いておいたら、あとは読み手に任せる形になります。
- Q: ライカート法のところの数字の並びなのですが、ネガティブなイメージのものを1とするかポジティブなものを1とするか。結構ほとんどが下から上というイメージだと思うのですが、中には逆のものも見受けるのですが、このあたりの決まりというか何かありますでしょうか？
- A: やはりその直感とズレるような形でやると、間違っって回答する人が増えるというのがあると思います。ネガティブなものが、数字が小さい方が理解しやすいですね。
- Q: 一つのアンケートをとるとき構成概念が複数あるんですか。例えばそういうようなものを作ると、クロンバックの係数はそれぞれ測らないと当然狂いますよね。通常良く見るのは全体に対してクロンバックで α がいくらということを出すということは、1つのアンケートは1つの構成概念で設計するのが正しくて、もし複数やりたかったらそれは別々にやりなさいというふうに理解したほうがいいですか？
- A: そうとも限らないと思います。要するに最初に2つあるというということが明確になっているということはすばらしいことで、それをイメージしてやることも可能です。クロンバック α が2つというモデルをあとから確認することも因子分析でできます。
- Q: 内的一貫性というところで、逆転項目を入れておくと全部5をつけるようなことが見抜けるということですが、逆転項目は入れた方がいいということですか、それともあっても構わないということですか？
- A: 目的によって変わってくると思います。要するに何か教育効果を見たいという場合に逆転項目があると、何か変に探りを入れられているようで、答える方が嫌な気分になるということになればあまり意味がないですし、何となく授業評価とかで、全部5をつけても面倒だから大きな縦丸、そういうものになると意味がないので、そういうのを見抜くためにという場合は必要な場合があります。

次の作業の説明に入ります

■ グループワークの実施要領説明

グループワーク 1 では、質問紙作成を実際に行ってみましょう。グループ内で司会者、発表者を決めます。これはいつもやっていると思いますが、各グループで部屋は移って頂きます。1 グループだけは、ここ真ん中で、同じところでやっていいですけど。2, 3, 4 はそれぞれ別の部屋が準備されています。それからテーマを選んでください。ここが一

番難しいんですけど、ここで時間を使い過ぎると後のワークが進まないという、このあたりが非常に今日の課題の難しいところだと思いますけどね。自分これやりたいんだという人はそれを主張してもらい、周囲の人は「じゃあ、やる？」という流れになるのかなと思います。

次に項目数と内容を考えてください。あまり多すぎないようにしてください。USB メモリを配りますので、誰かコンピュータもっていますかね？各グループで、もっていないグループがあったら後で言ってください。それで、USB メモリ内のファイルに質問紙を作ってください。一応フォーマット作っています。A4 一枚にまとめてください。これはすぐに印刷をするのでその作業の都合上です。

この後の流れを理解しておいてください。各グループが作った質問紙を一応 1 時間後くらいに質問紙を印刷しますので、USB メモリをスタッフに渡してください。15:30 から一旦、全体会の部屋、ここに戻ります。一旦ここに戻ってください。ここで、他のグループが作った質問紙に皆で回答してください。15:50、また各部屋に分かれますがそのときにはデータを回収して、回収率 100%になると思います。その回答結果を見て、問題点や傾向を探ってください。議論の内容をまとめて、16:30 から発表してもらいます、ということです。実際にやってみないとなかなか何が難しいのか議論しにくいので、作ってみようという流れになっております。

■ 質問紙回答

タスクフォース：大西先生

作ってみてやってみて、なかなかうまくいかないという感想だと思うんですけど、やはり難しいですね。ただ、意識して頂いた方がいいかもしれないのは、完璧な質問紙というのはないです。何かこれよりもこれの方がいいとかいろんな感想は出てくると思うんですけど、作られたものをみても、何かこれちょっと変だなと思うものもいっぱい・・・されているものもあります。質問紙って奥が深いものだという認識でいいんだと思います。他のグループのも見ていろんな感想もあるかもしれないんですけど、とりあえず黙々と 20 分で回答してください。お願いします。



グループワーク1

- 質問紙作成を実際に行ってみましょう
- グループ内で司会者、発表者を決めます
- テーマを選びます
- 項目数、内容を考えます。あまり多すぎないようにしてください
- USBメモリ内のファイルに質問紙を作ってください。A4 1枚にまとめること

ここはこうしたらいいのではと、アドバイスいろいろとあるかもしれませんね。そういうのがあったら、書いておいてあげると非常に親切だなと思います。さっき説明するのを全く忘れていたなと思ったのは、いわゆる数値的につけるようなものとオープンに答えるものとのバランスとかいろいろ質問紙常に考えたほうが良くて、そういうのもありました、そうでないものもありました。そうじゃないところ、特にこういうふうな形をつけたほうがいいんじゃないとかオープンなところにいろいろ書いてください。皆さんの役に立つかなと、数値だけだと入力して解析する時間があまりないかもしれないので、そういう意味でオープンなコメントを書いておいてあげるといいです。

残り（回答時間の）5分ですが、大体回答できましたか。そろそろグループごとに、各グループの回答を集めて、それぞれのグループに返さないといけないので、それぞれのグループの机に持って行ってください。すべて今回無記名なんですけど、集めたらまず用紙の右肩に通し番号をつけておいた方がいいですね。あとはどういうふうに解析するのかというのもお任せします。各グループに分かれて 16:30 に集合ですので。

Q：この間にブラッシュアップしていいのですか

A：してもいいですが、その時間はまた別にあります。どういう方向でやろうとか修正してもいいですし、もし修正するならファイルの名前はバージョンⅡとかにしてください。

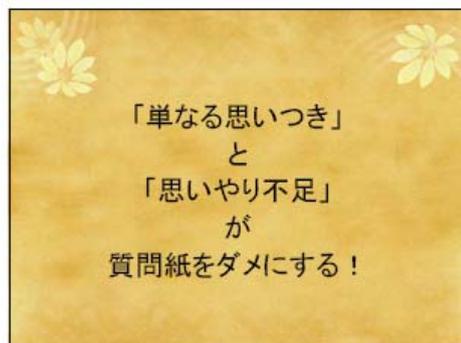
Q：集計することによって、どういう方向性がありそうかを見出すのが目的？

A：そうですね。ただ、入力作業がどれくらい速いかということで、制約があるとは思いません。自分たちはこういう経緯で作ったとか、最初はこういうふうに考えていたとか、まずそこからですね。

■ レクチャー 2 : 質問紙作成によくみられる問題点 (質疑応答含) …………… 杉本委員
25 分ほど遅れて進行しております。今から 30 分ほど 結構お疲れになっていらっしゃるんじゃないかと思います。時間も遅くなってまいりましたので、

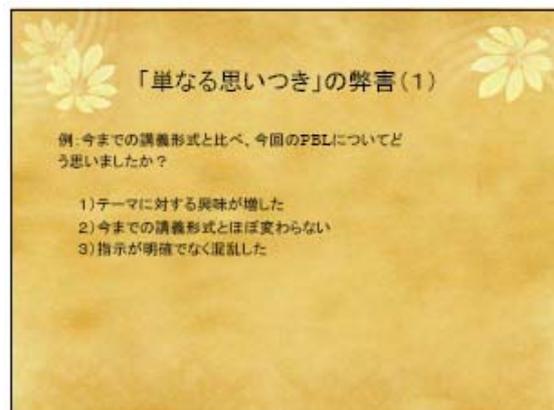
スライド 1 : 質問紙作成によく見られる問題点

「質問紙作成によくみられる問題点」について話をさせて頂こうかと思います。



スライド 2 :

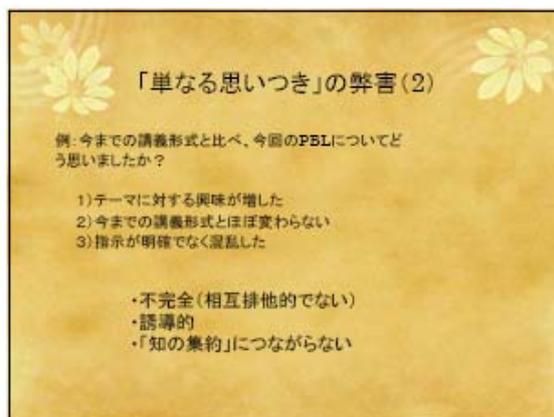
肩の力を抜いてお聞き頂いて、主に私の失敗談を中心にした軽い話ですので、リラックスしてお聞きいただければと思います。先程自己紹介の時に、私先生方に誰にも負けない自負があるのは、ひどい質問紙の犠牲になった回数だと申し上げましたが、それには理由がありまして、私は国際基督大学 ICU という大学を卒業しているのですが、丁度学生だった頃、私の専門である異文化コミュニケーションの中で、日米研究が非常に台頭してきたんですね。ところが当時にアメリカ側の日本研究者は日本語で読み書きが出来ない方が多かったので、日本人のデータを英語で採れるところというのは国際基督教大学しかないということで、そもそも非常に抽象的な概念を英語で、日本人の大学 1 年生に調査することがサンプリング上の問題があると思うんですけど、何度もやらされました。毎授業のようにアンケートに協力してください。あるいは海外で生まれ育った先輩たちが怪しい日本語で二重否定満載のアンケート用紙を作ってきてこれを 5 分で答えるとか、いやな思いをたくさんして、ああこういうふうにつっちゃいけないなということだけは学習しました。今度は正しいやり方を学習しようと思ってイリノイ大学というところに留学したんです。その先生自身がですね、限りなく質的研究に近い量的研究者を標榜している方なので、ものすごく丁寧に質問紙を直して下さるんですね。それでその、レイアウトとか余白だけではなくて、紙の厚さまでいわれました。あまり薄い紙を使って数ページにわたる質問紙だと、2 枚一緒にめくってしまって未回答が増えるからここはケチらないで、いい紙を使いなさい。最初どうしてこんなに瑣末なことを注意されるんだろうと思っていたんですけども、徹底してアンケート用紙を目の前に置いて、相手の気持ちになって考えろといわれたんですね、回答する人の気持ちになって考えれば誤答とか未回答とかが減らせるし、一番意味のあるデータがとれるという教育をされました。そこで、いろいろ ICU での



経験とイリノイに行ってから経験と2つ学んだのが単なる思いつきと思いやり不足が質問紙をだめにする。ということでしたので今日はこの2つを中心に前半後半でお話していこうと思います。

スライド3：「単なる思いつき」の弊害

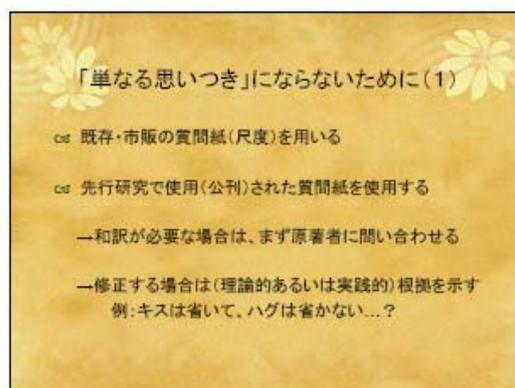
まず、思いつきのほうの例からお話ししようと思うのですが、良くある例です。今までの講義形式と較べると、「今回のPBLについてどう思いましたか？PBLのほうが良いといたい、誘導したい気持ちが非常に表れている。テーマに対する興味が増した。今までの講義形式とほぼ変わらない、指示が明確でなく混乱した」、作った本人としてはポジティブ、ニュートラル、ネガティブな意識なん



ですよ。4段階「はい」「いいえ」「どちらでもない」「・・・」でも全然ねじれを起こしていて、「テーマに対する興味が増す」おそらくここは「今までの講義形式とはテーマに対する興味は変わらない」とそういう意味でいっているんです。その一言が抜けているので、学生がそのまま読んでしまったら、講義形式が変わっているのか変わっていないのかというふうに受け止めますから、変わっているのか変わっていないのかといわれると替わったとしか答えようがなくなってしまうのです。こちらも「指示なく混乱した」というのはネガティブの一つにすぎないのに、これがなかったらネガティブ選ばない人が増えてしまう。結果的にここ選ぶ人が増えたりして、構造的に誘導されるということがよくあります。ということで、そういう問題もあるのですが、誘導の話は先程大西先生もコメントなさっているので良いと思うのですが、最後のところでですね「知の集約」につながらないという問題があります。皆が勝手にアンケート用紙を作って、思いつきで作って、それで調査をしてしまうと、似たような研究同じようなテーマの研究があっても、尺度が違うので、集約して比較したりとか論じることが出来なくなります。すごく研究者の労力の無駄使いになりますので、できるだけ比較検討が可能な方法をとったほうがいいたらと思います。

スライド4：「単なる思いつき」にならないために(1)

単なる思いつきにならないためにどうしたらいいかというお話ですが、先程大西先生から社会心理学の質問紙などの尺度をお見せ頂きました。心理学とか有料なものまでありますよね。一冊200円で買って使ってくださいというテストとかもありますので、そういうものもあります。コミュニケーションに関しては、この一冊が言語コミュニケーションに関する尺度集でこれは



信頼度が計算されたものです。こちらは非言語コミュニケーションに関する尺度集というのがあって、こういうものを使うことが出来ます。これは後ほど後ろに置いておきますので、ごらん頂ければと思います。そういうものがないテーマを扱いたいということで、でも一冊にはなっていないけれども、例えば似たような先行研究があつてすでに巻末に自分たちが使った質問紙をつけてくださっているようなものがあつた場合は、その論文をきちんと引用して、それをもう一回使うのは問題ないと思います。ずっとそういうふうに使っていたのですが、最近ちょっと身近でいろいろ見聞きして事例から思い始めたのが、例えば英語で書かれた論文の巻末に載っていた質問紙を使う、和訳が必要になってきますよね。その場合は一応、本来は原著者の許可を得る必要はないと思っていたのですが、一応問い合わせた方がいいだろうと思うようになりました。日本と外国との交流が盛んになると、日本で複数日本語版ができてしまう可能性があるんですね。勝手にやってしまうと。なので、原著者に問い合わせれば、「ああもう誰それさんがその日本語版作っていますよ」とか「問い合わせがありました」と答えてくださると思いますので、海外で開発されたスケールが1つだけ日本語訳が出来るという形の方がいいと思います。それが必要なと思いましたが、良くあるのが「海外で開発されたスケールなので、日本の実情に合わないからこの項目とこの項目落としました」という論文の記述があります。例えばアメリカの大学生を対象にした質問紙で学部時代に何を専攻しましたかの項目が入っていたとします。日本の場合該当しないのでそれは削除する。それは論拠がはっきりしているので構わないと思うのですが、時々、恣意的に落とされる方があります。これは医学教育とは全然関係がないので何でこういうのが出てきたかと思われるのですが、私どもコミュニケーション学の領域で、親しい人を慰めるときの行動という尺度がありまして、抱き寄せてキスをするという、ハグをするというのがあるのですが、その研究者・著者の方がいて私は聞いている側だったんですね、日本ではキスはしないのでそこは削除しました。その項目落としましたという説明がありました。じゃあハグは残っているんですね。キスは削除されていてハグをするかという議論、その研究会のメンバーは喧々諤々の議論になったのですが、尺度で考えればキスのほうがありえない度が高く、ハグはもしかしたらあるかも知れない。でもどこでそれを切ったかあくまでも恣意的なので、やはりそこはある一人の研究者の判断でキスを省くけどハグは入れますというのではなくて、先程フォーカスグループのお話をなさっていましたが、そういう方たちに最初に見て頂いてここここはないでしょ、そういったことはないでしょうといった意見を実際に予備調査をやってみて、省くとか何か根拠を示して、根拠を示して省く修正するというのが必要だと思います。査読をしていてこれが非常に多いです。勝手に変えてしまったというのが、そうすると新しいものを作ってしまったということと同じことになってしまうので気をつける必要があると思います。

スライド5:「単なる思いつき」にならないために(2)

「単なる思いつき」にならないために(2)

理論から導く

例: 症例プレゼンテーション向上の段階モデル

	プレゼンテーションの状況	プレゼンターの臨床レベル	指導者からのフィードバック
第1段階	症例プレゼンテーションを構成している症状や所見の情報自体の定義や信頼性に問題がある	医療面接や身体診察の基本ができていないため、症例プレゼンテーションを聞いていると整合性のない部分がある。	症状や所見の情報において、どこが不明瞭に感じられたかを指摘する。医療面接や身体診察の教育を徹底する。用語の定義が怪しいと感じられれば質問する
第2段階	症例プレゼンテーションに不可欠な臨床情報が列挙されていない、情報の順序がばらばらであるなどの理由で、聴衆に症例の全体像が十分に伝わらない	現場の指導医に要求されている臨床情報が把握できていない、型通りに臨床情報をまとめられない	プレゼンテーションが長くなってよいので、最低限必要な情報のうち何が足りなかったかを指摘する。プレゼンテーション技法の問題があれば練習させる
第3段階	症例プレゼンテーションに不可欠な情報は網羅されているが、鑑別診断を十分考慮した構成にはなっていない	患者の情報の連絡役は果たしているが、自ら問題点の全体像を把握し、解釈することはできていない	情報が揃っている点は褒める。徐々にプレゼンテーションを短くするよう指示し、不足している鑑別診断についてフィードバックする
第4段階	鑑別診断はある程度できているが、その鑑別診断に必要な陽性・陰性の症状・所見が十分には盛り込まれていない	症状や所見の情報を得る際に、患者の全体像や鑑別診断との十分な関連づけができていない	鑑別診断は一定レベルと褒める。必要な症状や所見について、具体的に何が不足しているのかをフィードバックする
第5段階	鑑別診断と共に、必要な陽性・陰性の症状・所見が盛り込まれている	症状や所見の情報を得る際に、患者の全体像や鑑別診断との関連づけができていない	できている点を具体的にフィードバックする

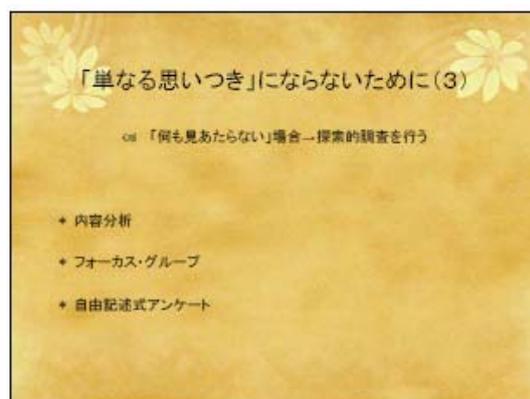
出典：大西弘美(2008). 臨床試験の評価法. 日本内科学会雑誌. 第77巻第10号. 228.

もしそういったものが全くない、先行研究も質問紙もない場合は、なるべく理論から導くのがいいと思います。これは、大西先生が提案なさった症例プレゼンテーション向上の段階モデル。たまたまご自身がなさりたい研究がプレゼンテーションを教えるワークショップを企画して、それを学生にやってみて、プレゼンテーション能力が向上したかどうかという研究をしたかったとします。でもやっていくうちにもしかしたら最初に参加される学生さんのベースラインによって教育効果が違うかもしれない。すごく下手な子だったら、半日でも伸びるかもしれないけれども、ある程度の上手な学生さんだったら、半日位でのワークショップでも伸びないというそういうような研究をしたいと思ったときに、また思いつきでご自分の学生さんの症例プレゼンテーションのレベルを松・竹・梅って判定してしまって、負担感とかその教育効果を判定すると、その松・竹・梅の自分の独自の尺度です。またそれをすり合わせる事が出来なくなる。できればこういったモデルで1段階の学生と2段階の学生と3段階の学生と決めておくで後で擦り合わせ、立ち返って論じることが出来るので、なるべく交換されるものとかモデルに擦りあわせで質問紙を作っていくのもよろしいかと思えます。

スライド6：「単なる思いつき」にならないために(3)

何も見当たらない場合、探してみただけでもモデルも理論も何も見当たらない。よく大学院の頃「何もしません」というふうに報告するとよく先生にしかられました。「あなたがみつけれなかっただけよね。」先行研究が見当たらなかった場合は、探索的調査をすることができます。私自身は日米の謝り方の比較という博士論文を書いたのですけれども、最初は全くスケールがなかったので作らなくてはいけなくて、

内容分析を課されました。日本とアメリカのマナーの本ってあります。手紙の書き方ですか挨拶の本とかマナー本を30冊位、内容分析を日米両方してどれが日米で違いがあるか、読んでいくうちに分析していくうちに、言い訳をするかしないかとか、相手にかけた迷惑を大げさに言うか過小評価するかが日米の違いかなと思いました。次の段階としてイリノイに留学していた日本人の方に10人くらいに来て頂いて、そういったキーになるかなと思う概念についてディスカッションして頂いて日米の違いを出す。最後はできるだけアンケートに使う状況場面をリアルなもの現実的なものにしたかったので、自由記述式のアンケートを一回行って、今まで誰かに謝ってもらいたかったのに謝ってもらえなかった状況はありますかとか、自分が謝った状況ありますか書いて頂いて、こういう状況だったら現実的なんだろうなというのでも12条件抽出しました。そういうやり方があります。医学教育でこれを応用なさるのであれば、内容分析であれば、研修医向けのマニュアル本とかたくさんありますよね。ああいったものを内容分析なさって、こういうところが問題だと考えているんだなというところから議論を始められてもいいのではないかと思います。あとフォーカスグループとしてのやり方も自由記述も同じ。一応それでアンケート用紙が完成したとしてもそれで終わりにはなりません。

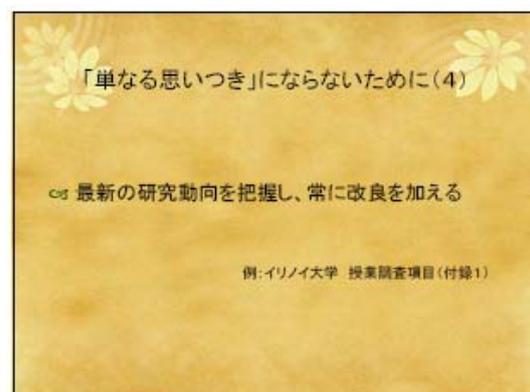


スライド7: 「単なる思いつき」にならないために(4)

単に思いつきにならないための最後ですね。一回完成したアンケートでも最新の研究動向を把握して常に改良を加えるという必要があります。

付録1

その実例をお目につけようと思ってプリントの一番最後ですね、付録1として付けさせて頂きました。イリノイ大学の授業評価の実物です。今は使われていません。電子化されてしまっていて、紙だった頃のもので。私が教え始めた3年目の1990年、秋学期のもんですが、ここのご説明しようと思うのですが、授業評価自体は、皆さん、例えば学年を学生さん



に聞くということは普段なさっていると思います。それ以外に加えられた項目として、例えば4番目ですね。コースイン **Course in** って書いてありますが、この科目が主専攻の科目だったのか、それとも副専攻の科目だったのか、それ以外の他学部の授業だったのかによって、学生さんの興味関心当然違うだろうということがまず加わりました。その次その上の3というところですが、これは必修科目だったのか全く自由に選べる選択科目だったのか、あるいは選択必修の科目は日本でもあると思いますけれど、そういう種類の授業だったのかということによって、学生は嫌々取らされていたのか自分から選んで取った授業なのかによってもバイアスがかかるだろうとして、その要因が加わりました。その後で2番が加わったのですけれども、この授業で履修登録をした時点ではあなたは授業と担当教員のそれぞれに対して、どういう印象をもっていましたかという受講前の印象もやはり授業評価に関連するのではないかということで、そこに学生さんがこれ実物なので、インストラクターに関してはノーオピニオンだった。授業に関してはポジティブだった。というふうに丸をつけていると思いますけど、そういう要素が加わっていきました。実際に取った学生の成績と授業評価の好意度というんですか、そのどれだけポジティブな評価がどれだけ得られたかという評価についての相関を見たら、なんと悲しいことに一番高い相関が得られたのは、どれだけいい成績をとっているかということと教員とか授業に対する授業評価の高さだったんですね。要するに楽勝科目の方が有利なのではないかという話しになったのでそれに補正をかけようという話しになって、楽勝科目の教員が有利だとこれは結果を学内新聞に発表するので、教員の不公平さをなくすために補正をかけるようと研究していくうちに、よく調べてみたら、実際に取った成績よりももっと相関が強いのは、このアンケートを記入したときに学生が自分はどんな成績を取っているかどうか。学生時々勘違いをしますので自分はBだと思ってうらみつらみを書いたら、最後Aがきたということがあるわけです。実際に記入時点でどう思っているかということが補正をかけるのが適切な要因だと分かってこの6番目の項目が加わりました。Expected Grade と書いてあると思うんですけれども、この時点では最後の授業でまだ成績もらっていませんから自分では何がもらえるかと思っていますかという、Aと書いた学生があまり多い教員は、補正されるんですねそのスコア。あなたは楽勝でしょ人気取りはしてはいけませんよとなるんですけれども。そんなふうにして一回完成したアンケートでもどんどん改良が加わっていきます。新しい知見が得られる度に。そういうふうに思って頂ければと思います。では思いつき対策はここまでにして、今度は思いやり不足のお話しをしようと思います。

スライド8：「思いやり不足」の弊害

これは、実例で私の恥ずかしい例なのですが、指導医の先生方の負担感を調べようと思ってこれに似たような形のアンケートをつくりました。

「思いやり不足」の例

ご専門	
臨床指導に経験年数	年
指導の形態 (あてはまるものにし)	<input type="checkbox"/> 一対一 <input type="checkbox"/> 一人の指導医が複数の研修医を担当 <input type="checkbox"/> 複数の指導医が複数の研修医を担当 <input type="checkbox"/> 複数の指導医が一人の研修医を担当 <input type="checkbox"/> その他 ()
研修医指導に費やす時間は	週 時間

大失敗でした。ご専門、後で説明しようと思いますがとんでもないパンドラの箱をあけてしまいました。臨床指導とご経験年数これは分かりやすいと思ったんですね。これ以外に卒後年数というのが書いてありますので比較ができるのですが、この臨床指導経験年数の解釈が皆さん全員違う形で、指導医講習会の会場で取らせていただいたので、認定証をもらうまでは自分は指導医ではない、非常にストイックに考えられた方は、「0」とお書きになるんですね。一方自分は臨床で指導してきたという自負がおりになる方は、卒後年数 25 年、臨床指導経験年数 24 年と書かれていて、2 年目からずっと指導なさっているんだという厚さが伝わってくるんですけども、もう意味はないんですねここんとこの評価項目は、失敗したなと思いました。指導の形態もあてはまるものにチェックをしてみてくださいともしあげましたところ、複数選ばれた方もあったんですけども、これだと全く名義変数、名義尺度しか使えなくて、これむしろどういう事前に指導医の先生に何人かにインタビューして、負担感の順に並べることが出来たら、一対一は結構煮詰まるから人数少なくても大変なんだとか、やはり複数みているほうが大変なんだとかそういう情報があったら、もう少しこの尺度の工夫が出来ただろうなという気がしました。ここが非常に意外に足元すくわれたところですよ。後でご説明いたしますが、実際の指導者の負担感の相関をみたかったので週に何時間の指導に費やしていらっしゃるかと質問をいたしました。これがどうなったかという話しを絡めながら先に進めさせていただきます。

スライド9：「思いやり不足」の例（1）

まず視覚的に理解しにくい回答しにくいというのが思いやり不足のアンケート用紙の実例の1です。レイアウトがごちゃごちゃしているとか改行が途中でなされているとか、改ページが変なところでなされているとか、回答スペースが足りないとかありますが、これは先程の研修医指導に費やす時間は絶対間違わないと思ったんですね。週3時間とか書いてほしかったんですけどこれは致命傷でした。ここに週4時間と書かれておそらくローテーションで、そこには一週間のうちに3回だけ学生さんと研修医が来るということ、更にここに4時間と書かれると、1週4時間なのか全体で4時間なのか1日4時間なのかさっぱり分からないので、非常にまさかこのスペースに何かお書きになるとは思わなかったですけども。これは実際には時間/週としたほうが間違いがなかったんですね。1週当たりと書くとか、そういうふうに本当に思いもしなかったことで足元すくわれました。それから直感的でない尺度。先程の不満というネガティブなものに5をつけていいのだろうか。そういう議論があったので飛ばします割愛しますが、これが実際に博士論文で使ったアンケート用紙です。英語版のほうですけども、私なりに色々工夫はしたつもりだったんですね。例えば表題が



Question on interview professional communication と書いています。対人コミュニケーションに関するアンケートとした漠然としたことしか書いてなくて、実際は謝り方の日米比較だったんですけども、そういうことは一言もでてきていない。なぜ入れなかったのかというとバイアスをかけてしまう。これ状況を与えてあなたはどう反応しますかというふうに訊いているんですね。謝るか謝らないかということも測定項目の一つですので、アポロジューと入れるとそれが誘導になってしまう。ということで外してあります。日米比較ということが変なバイアスをかけてもいけないと思って外しました。男女差は興味がなかったんですけども、男女別々の用紙を配る必要がありました。登場人物を同性にしたかったので答える人と。なので例えばミッキーとミニーになっているんですね、なかの人物が。その男女を間違えて配ってはいけないということで、コーディングでも間違いを防ぐ為に色をつけようと自分ではそれがすごく頭のいい考えだと思った訳です。一見すると、例えばこちらが女性用としたらこちらの先生が答えたらおかしいわけですから、間違いが防げますし入力のときに間違い男性用かいちいちみる必要がないので、自分でもすばらしいと思ったのですが、日本に送る前に何色にするかでもめて、ピンクとブルーではあまりにもあからさまで、反感をかうのではないかという結論になって、黄色とグリーンにしてみたんですね、見事に逆にされました。400名分無駄になりました。400名分男女別に配られてしまって、通年制の大学だったので、その学年では実施頂けなくなって卒業が若干延びました。今でも覚えています。がたまたまパリのユーロディズニーにいたときにその電話をもらって、そこから先は何をしたのか全然記憶にないのですけれども。すぐアメリカに帰ってきて先生の研究室に行ってティッシュ1箱使って泣きました。どうしよう卒業できない。こんなに自分で賢い方法を考えたと思っていたのですが、ピンクとブルーを回避してしまったために、ピンクとブルーにしたら良かったかというところでもないかもしれませんね。それはそれでバイアス与えて、男女差に全然興味がないのにも関わらずそういうのに調査をして、回答者がそういう反応したのかもしれないので、いいのかどうかわかりませんがそんなことがありました。なので直感的でない尺度と色には気をつけましょうという感じですよ。

スライド 10: 「思いやり不足」の例 (3)

今度は内容的に理解しにくいもの、回答しにくいもの。専門用語のお話が出ていました。二重否定「こんな医師になりたくないとは思わない」というのは答えにくいです。順位回答というのは例えば、医療面接には3つの目的があります。情報の収集であるとか、関係の構築であるとか、患者の指導・教育とか、あなたが大事だと思う順番に付けてください。3つくらいなら多

「思いやり不足」の例(3)

☞ 内容的に理解・回答しにくい

- ・専門用語
- ・二重否定 / 後否定 「こんな医師になりたくないとは思わない」
- ・順位回答
- ・限定質問 「問1にあてはまると答えた方にお尋ねします」
- ・項目数
- ・属性項目 「出身高校名を書いて下さい」

分大丈夫だと思うんですけど、5つ7つになると頭の中で数えなきゃいけない、抜かす人が出てきます。なので、順位回答は行いにくい。

それから限定質問、フローチャートのようにになっているものありますよね。問い1ではいと答えた人はこちらに行ってください。それはやはり煩雑なのでそこから先は答えなくなる方がいたりしますので気をつける必要がある。

項目数、先程の逆転項目を作るかどうかということがありました。心理の尺度とかは100問とか平気でありますよね。30分回答にかかるとか。それはやはり疲れてきますし、5、5、5と全部つけてしまう人も出てきますし、時々入れ替えることがあります。

属性項目の訊き方も大変難しく、例えば学生が公立高校の出身か私立高校の出身かの程度のことを知りたいだけなのにこういうふうに書いたりすると、出身高校で差別されるのではないだろうかとか、バイアスがかかってしまうことがあります。属性の訊き方はなかなか難しいのですけれども、どうしても私たちデザインするときには独立変数、従属変数という並べ方をしてしまうわけですよね。アンケート用紙自分の中で。そうすると属性項目が最初に来てしまうことが多いので、属性項目答えやすいのでそう意味ではいいのですが、たとえばこういうような回答者が構えるような項目があるような場合は、一番最後に書いておく手もあると思います。必ずしも最初に聞く必要はないので全部答えて最後に属性がきたときに、属性答えたくないければ抜かしてもいいですけれども、最初に属性があってそこで自分が評価されるんじゃないかと思うと、最初にストップ書かない人って結構ありますので。私は今看護師の教育課程によって回答の傾向が違うかなということを調べているんです。看護学校なのか4大を出ているのかを短大出ているのか、という項目があります。一番下にしてあります。一番トップにははいりません。そういうこともあります。それから、

スライド 11 : 「思いやり不足」の例 (4)

最後思いやり不足、これは自分に対しての思いやり不足の例です。集計・解析のことを考えずに作ってしまった自分の首をしめることになる。今は拝見しましたら問題には皆さんちゃんと番号がついておりましてで大丈夫だと思いますが、付けておかないと入力するときに大変なことになりますね。後はカテゴリー化のことを考えずに自由記述とすると収集がつかなくなって、先程の例ですが、私が対象だった指導医

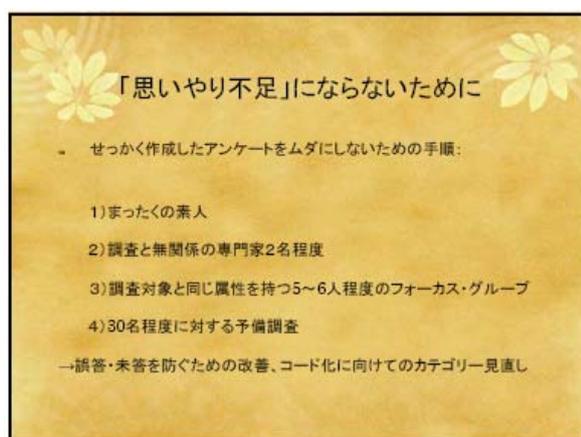


563名の方にご専門と聞いてしまったがために、153通りの回答ができました。リハビリテーション・呼吸器内科、括弧の中はお一人の回答なんですが、医学教育・総合診療ぐらいいいんですけども、救急/外科/内科だって全部させられているのねという感じが伝わ

ってきたりとか、肺癌・乳癌と疾患名を書かれる方がいらっしやったりとか、外科→救急→緩和、変遷を書いてくださったのかもしれませんが、今がここなんです。どう分類していいのかわからなくて、半泣きでしたこれがでてきたときにはどうしようと思って、それを痛い経験を踏まえて、今、こういうふうに変えてあって、主だった診療科が書いてあって、あとは強いて言うならご自分は内科ですか外科ですか、その後も答えられるように改善しましたので分類の仕方を考えてから項目を設定するというのを学習しました。

スライド 12:「思いやり不足」にならないために

じゃ思いやり不足にならないためにどうしたらいいのかというお話ですが、先程何回かフォーカスグループの話が出ていました。これは時系列を追ったものなのですが、まず一番最初に作成したアンケートを全くの素人に見てもらおうほうがいいと思います。むしろ、ご家族とか、これじゃ字が小さくて見えないわよというレベルのところから指導してもらったほうがいいと思います。私自身も大学院のときは、



たまたま西海岸に駐在していた日本人の友達 2 人にファックスを送って、みても全く内容が分からない友達にみてもらって、それこそ字の大きさからコメントもらいました。その後は、調査に加わっていない研究班に関わっていないけど内容には詳しい方 2 人くらいにみて頂いて、その後はフォーカスグループをして、その後は、答えてもらうというよりはディスカッションですね。「これは意味が分かりません」先程出ていたようなお互いに疑義を出しあうセッションを持って、それから 30 名程度に対する予備調査を行って、私は実際に日本にデータをアンケート用紙を送る前に、アメリカに短期留学、夏季の語学研修で来た日本人の方に協力して頂いた 30 人、これをしました。文言が正しくないとかそういうところを修正しました。先程ご質問で、どの程度構成概念とか目的を開示したらいいのか、バイアスになるのかならないのかはこの辺で確かめることが出来ると思うんですよね。一手間に見えますけど、実際の母集団、対象集団から 30 名をロスがでてしまいますけど、でもアンケート失敗して 500 人分失敗することを考えたらこの 30 人というのは非常に重要な保険だと思いますので、こういったことをお勧めいたします。それを踏まえて誤答・未回答を防ぐための改善を行うとか自由記述のコード化に向けての カテゴリ を考える手順がよろしいかと思います。ここまでで、私の説明を終わらせて頂きます。

今の私のご説明と先程の皆様のコメントと注意点を踏まえて、アンケート用紙を改良して頂くこととなります。

グループに分かれる前に、今の私のご説明の中で何かご質問やご意見がありますか。

質疑応答

Q: この実際のアンケート用紙で、上2つが数字でその下がアルファベットになっているのは何か意味があるのでしょうか？

A: 上2つというのはプリントされているものをごらんいただけるといいのですが、全学必修項目といって全員がこの項目をこれを入れなければいけないというふうに決まっているのですね、その次のタイプされた項目というのは、その教員のランクによって自由度が決まっています、大学院のティーチングアシスタントもデパートメントが決めたものを使うしかないのですけど、教授クラスだったら全部自分で選んでいいとか、自分の有利な方に選べるんですね問題を。この教授はタバコを吸わなかった項目が当時ありましたので、そういうものを選ぶこともできるんですけども、最初の2つは全学必修項目でそのスケールにそって全教員が比較されるのでそういったところが入っていると思います。

Q: 逆に言うとそういう自由度によって、数字で答えさせるとか、アルファベットで答えさせるとかそういうのもやはり心がけたほうがいいのでしょうか？

A: どうですかね。これに関しては学生はアルファベットになっていることにそれほど気にしていないと思います。上をみて1~5につけていると思いますけど、むしろ処理上の問題だと思います。

Q: 先生のアンケート用紙の話の中で、色の話しが出てきましたけど、私も外部に郵送でアンケートするときに、何かのどこかの統計の専門家の先生が色をつけると回収率が良くなるということをおっしゃったらしくて、ウチの教授が必ず色をつけよというので、しかもピンク色が一番回収率が上がるというようなことを行ってウチで郵送するアンケートは全てピンク色の用紙に印字しているのですが、そういったこと本当なんですか？

A: それは初めて聞きました。懸賞はがきは色をつけると当たるという話しは聞いたことがありますけど、アンケートは初めて。たぶん受け取られた方女子医大だからピンクなんだと思われているかも。そういう文献は見たことがないですけど、調べてみたいと思います。

Q: 今のお話の中で個々の文献をなるべく見て、似たような質問項目を引用したほうがいいんじゃないかということですが、実際にはなかなか該当するものぴったり見つけてくるというのは難しいのではないかと思います。そういった場合には、やはり、ある程度参考にして自分で改変して作るというふうにした方がいいのか、あるいはむしろ質問を過去のすでに発表されているものに合わせて、発想を変えてやったほうがいいというふうを考えてやっておられるのか？

A: それはケースバイケースで、どれほど自分のなかりたいことと一致しているのか2項目落としてすむものであるのか、半分以上変えなければならないのかによると思うのです

けど、メソッドの論文で、例えば自分に都合の良いようにスケールを変えてそれを検証している論文、もう一步保険をかけるんですよね。自分の研究以外に、そういうものがあるので、そういうのも全体的に研究の流れを見たときには、研究分野に貢献をしますの
で新たなスケールを作るとか、これとこれのいいところ取りをしましたということを公開
するという形ですね。そういう考え方もあるかもしれません。例えば私は内容分析、マ
ナー本 30 冊の比較をしたというのが博士論文の 1 章になっているのですけれども、独
立して出版したのでそういう下準備ではあっても業績にすることは出来ます。

Q: 修正する場合、最初に市販のもの既存のものといった修正は理論的根拠を示してとおっ
しゃったのですが、明らかに実施してみて未回答の人が多いものは省いてもいいのでし
ょうか？

A: それは良いと思います。予備調査をやってみたら、例えば約 8 割の人が未回答にしたと
か、これは分かりませんか書いてあったとか、そのことを報告して、だからドロップ
しましても構わないと思います。

Q: 具体的にダッシュスコアという北米がもとで・・・があるのですけれども、肘の機能評価
で、題目の評価に性生活を普通にしていきますかというものがあって、各国版があるん
ですけれども日本でも委員会がその項目を省かずに作ったら、子どもは答えない、老人は
その質問はととてもとてもみたいな何で肘でそんなことを訊かれるんですか？あまりに
も少なくても、・・・学会にこれはおかしいといったほうがいいのか？

A: 回答率が悪かった時はドロップしても仕方ない。本当は悪かった項目に関してフォーカ
スグループにして理由も探ったほうが良いとは思いますが。なぜ答えなかったかというこ
とを。でないと、単に回答率が悪かったからというだけで切ってしまうと、もしかした
ら改善の余地があったかもしれない。今話の中でご説明しそびれたのですが、海外で
開発されたスケールだとか教育尺度とかを文化の枠を超えて、引用するときのマニユ
アル本みたいなものもありますので、これも後ろに置かせて頂きますのでご参考になれば
と思います。

Q: 調査目的の開示とバイアスということをおっしゃっていたんですけど、先生の例ですと
謝ることですか、それはですね調査者に目的を開示しないこととバイアスをどうする
か・・・どういうふうに判断すればいいですか？

A: 基本的には大まかな方向にぼやかす分には良いと習いました。アポロジーということ
を書かないで、インターパーソナルコミュニケーションとはアポロジーの下位項目だと考
えれば、上のほうに漠然化するのはいい。ただ全然違うことを例えば人を傷つけてしま
った場面での調査みたいなことを書いてしまうと虚偽になるのでいけないとか、例えば
アポロジーよりももっと小さい項目、言い訳の調査と書いて本当はアポロジーを調べて
いるというのは良くない。場合によっては時々その虚偽の指示を出さないといけない実験
とかあるときはあるんですけど、そういうのはディブリーフィング(debriefing)を最後
に必ずしなければいけません。例えばアポロジーの調査を学生を使ってしたんですけれ

ども、義務づけられていて終わった後結果の報告をするんですよね。何ヶ月か経ってからですけれども、皆さんに協力してもらったアンケートはこういう結果が出ました。そのときにあの時はインターパーソナルコミュニケーションと書いていましたが、これは日米の謝り方の比較でしたという説明を加えています。

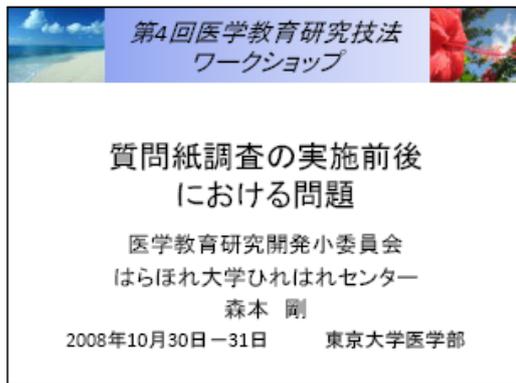
Q：今、項目をどのように・・・ということでしたが、因子でわかれている尺度なので、1つの因子だけで評価したいとかいうときどうしたらいいでしょうか。もし既存の中の2つの因子の尺度を使いたいというときにはどうしたらいいのか？

A：それはそのまま中身を変えないで、そこところでマッピングすることは可能？（可能です）それは論拠がはっきりしていればいいのでは。質問紙が長くなりすぎるとそういう理由ですか？（はい。自分の研究目的に合わない尺度が入っている、必要な一部分だけ使う、そのかわり中身は変えない場合）論拠を明示すればいいのではないかと思います。

■ レクチャー3：質問紙調査の実施前後における問題（質疑応答）…………… 森本委員
このワークショップ，はらほれって，2年前に来ておられた方は2年前のはらほれ大学の学
事データを解析した方もおられるかもしれませんが。

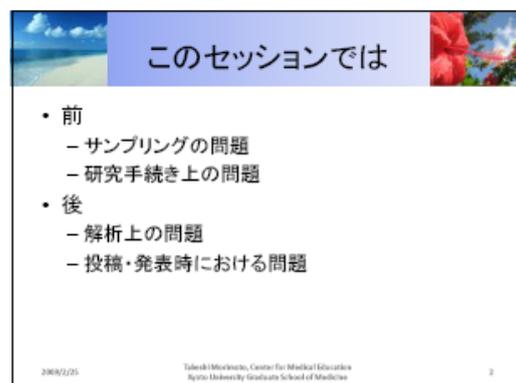
スライド1：質問紙調査の実施前後における問題

質問紙っていくつかパターン，皆さん目的があ
って，学内で調査とか病院内で調査するだけの
目的だけでなく，私どっちかという最終的
にファイナルでペーパーにしたいという人に向
けての少し，前後の問題，中身の問題，質問紙
作るといことに関しては昨日，サウロ先生が
お話くださって，中身の私は前後に関して少
しですね，私は元々臨床統計疫学を教えたりで
すね，やっているのもそっちのパースペクティ
ブ perspective ということで，します。上，これ南の島のような写真ですが，夏行った一週
間ぐらい琉球大学で統計，臨床統計のワークショップで作ったパワポを気に入っているの
でそのまま使っております。



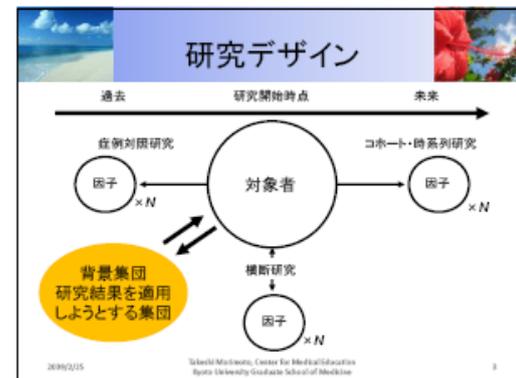
スライド2：このセッションでは

このセッションで何をしようかと思っていて，
一つサンプリングの問題，少し昨日ふれていま
す。それから，研究手続き上の問題，ファイナ
ル的な和文なり英文なり，ペーパーにするため
のゴールですね。解析上の問題を少し触れるの
と，投稿発表時における問題，4 つについて少
しやります。お手元にも資料がありますね。考
えていたこと大西先生，サウロ先生が触れてい
るところと重なることがあると思います。サウ
ロロフス先生、昨日 2 つ目のセッションやった先生のことを我々委員会では杉本先生のこと
とサウロロフス先生と呼んでいるんですけど，
こんな形ですね。



スライド3：研究デザイン

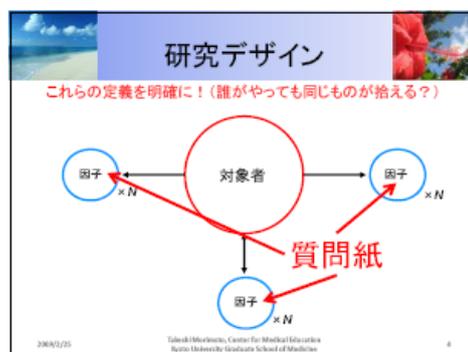
研究デザインでは色んなパースペクティブで
お話しするときにするんですけど，教育でも一緒
で，大体その対象者を確実に固定して，それに対
して後ろにファクターを探すのがケースコント
ロールとかいいますが，その後フォローするのが，



コホートとか・・・時系列であるし、ほぼ同時に対して、いくつかの因子・・・研究デザインですね。もし、3 因子で調べようとするならこの背景集団をですねよく考えるということが大事ですね。

スライド4：研究デザイン

とにかくこの大きな丸をですね、定義を明確にして、誰がやっても同じものが拾える、というのがサイエンスで、誰かがやって違うことが出てくるといのはサイエンスではなくて、ほんとうにやっただけですね。昨日今日のセッションというのはこの因子について質問紙で拾おうというものです。とにかく人によって解析が違えばファクター違ってきてそれは・・・という話しになります。

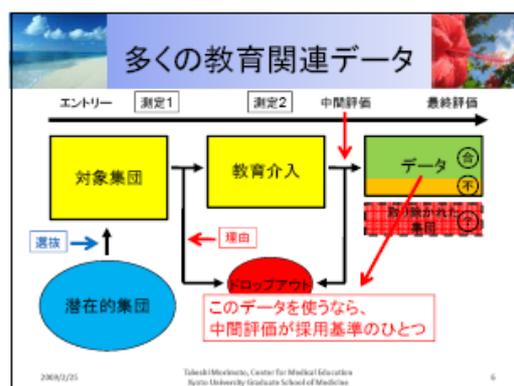


スライド5：サンプリングの問題

サンプリングの問題，少しお話しします。

スライド6：多くの教育関連データ

もう一つは教育関連データというのはこういう解析なんです。私もともと教育研究が専門ではなくて、疫学とか統計なんですけれど、最近教育のことというとい、いくつか潜在的な集団がいて、対象集団がいて、ほとんど考えているのが、レジデント研修医であるとか、学生であるとかがターゲットなんですけど、場合によっては色々な大学でアンケートをしょっちゅう配るとかですね。というのがありますし、もっと

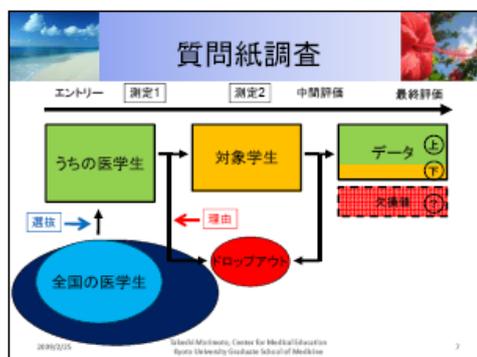


患者さんのパースペクティブで調査する。対象集団というのは色々あります。多くは学校とか病院の対象集団がいて、それに対してある測定をしますね、エントリーして、教育介入をして測定をする。途中ドロップアウトがあつて、中間評価、例えば国家試験だったら、最後に右端にデータをとってですね、合格した人不合格した人、大体ほとんど持っているデータは右端のですね、このこれしかもっていないんだけど、中には当然取り残された集団があつて、一つ分かる。ある程度分かる。この赤い集団というのはわからないんだけど、ほとんど右上のコードだけで解析をして、論文にすると左下のことを思いっきりいっているとかな。必ずシステムティックに違うということは意識してください。当然この選抜の青いプロセス、ドロップアウトした理由とかですね。いうことに対して当然書かなければならないし、場合によってはこのデータ使えなかったら中間評価が基本的採用基準というところがあるのでそこを明確にする必要はあります。今、自分が持っているデータについて、どの集団をいっているのか、それがどう反映していて、そのデータにくるま

でにどのようなプロセスがあつてということは常に意識してください。ほとんどその論文の査読する段階ではそのあたりがターゲットになります。

スライド7：質問紙調査

質問紙調査に少しフォーカスを当てると、おそらくうちの医学生が選抜されて、最初から対象学生、不参加には理由があつて、そこからデータが取れて、データの中に質問紙の中のスコアが高い低いができます。当然欠損値も同じように現れる。全国の医学生は背景にいるんだけど、もし、それにインターナショナルジャーナルに書くのであれば、バックグラウンドにインターナショナルの学生だと、この一番濃いところなんですけど、自分が論文書くときに非常にこの間ですね、自分が今持っているデータと、論文で書こうとしている・・・とどう違うのか、常に意識しないとかなり論文みていると、データの中で突然大上段に大きな話しをするんだけど、それを言おうと思ったら後ろ振り返ってですね、データからどうこの人たちが集まったかということは常に意識してください。かなりシステムティックに違うということは想像に難しくないと思います。



スライド8：医学教育に関連する定量的研究の問題点

次に、医学教育に関連する定量的研究の問題点。これも昨日ふれたのとほとんどですね。評価項目というのは、対象者の性質、態度、行動と、ソフトアウトカムが医学教育研究には結構出てきますね。昨日、大西先生が話した、評価項目の信頼性・妥当性は非常に重要ですね。臨床研究ではハードアウトカム、死亡とかですね、心筋梗塞とかおそらく簡単で、あまり難しくないのだけれどとにかく医学教育研究はソフトアウトカムが多いですね。

医学教育に関連する定量的研究の問題点

- 評価項目(対象者の性質、態度、行動)がソフトアウトカム
 - 評価項目の信頼性・妥当性
 - ⇔臨床研究(ハードアウトカム:死亡・AMI)
- 評価尺度の多くが手製の順序変数
 - Likert型
 - 分布の不整、天井効果/底辺効果
- Nが少ない
 - 1学年100人→選抜→回答

評価尺度の多くが手製の順序変数というのが多いですね。ライカート型で分布の不整とかですね、天井効果、データのほとんど上のが集まってですね、底辺効果というのは下のほうが集まって、これを片っ端から平均出してt検定していくというのが多いですね。昨日たしか1グループのほうはたしか平均出す前に少し分布を見たと思います。他のグループでは分布見ましたか?自分のとっているデータについて。自己調査で、ほとんど分布を見ずにいきなり統計ソフトやって、t検定、平均値出すというのは、かなり misconduct の一つの典型です。まず分布を見るというのを必ずやってください。分布見てこれはt検定がいいのか、・・・いいところと悪いところとにわけてですね、有り無しで検定どっちがいいかわかると思います。まず分布を見ないというのは結構初学者多いところですよ。また、

N が少ないというのは典型的な話ですね。けっこう多いのは1 学年 100 人ですね。今後増えてくれるので、n が増えてきて p が 0.05・・0.04 になったら学生の定員が増えていいのかな。選抜されて回答できたという話になります。

スライド 9：サンプリングの影響

サンプリングの影響。自分のデータですが、これはある病気、疾患の患者さんのデータで、真ん中トータル 760, ある難しいクエスチョンによって、それが available, non-available の人を でこれテーブル I で、テーブル II, III にいくんだけど、何をしたか、最初に調査がデータがある人が、データが何がどう違うかと clarification, 明確にしなければならぬ。これは、ことごとく p が 0.06, 5 を切っている。これは何が言いたい後言うと、この次のテーブル II から話す集団については、全体の集団とはシステムティックに違いますよということをおっしゃっているんですね。これはかなり多くのこと。学生が 100 人いて 50 人だけしか例えば調査に参加しなくてそこに 50 人の何が違うのか調査に参加してくれていないので、細かい調査項目は分からないけれど、少なくとも、大学はいるまでの年数が違うとか、研修医で来るときの前のステップが違うとか、・・・はいくつかあると思いますね。そういう形で何がどうシステムティックに違うかというのは非常に大事な情報です。



TABLE I. Characteristics of survey respondents according to the availability of time (total of 1770) data

Variables	Total (n = 760)	TTO available (n = 467)	TTO not available (n = 293)	p*
Mean age (SD)	46.0 (12.1)	45.2 (11.0)	47.2 (13.0)	0.24
Mean age of onset	47.1 (12.5)	47.1 (11.6)	48.4 (13.0)	0.08
Mean duration of illness (SD)	4.6 (4.7)	5.1 (5.0)	3.7 (4.2)	0.42
Median and 1st quartile (IQR)	4.0 (3.0-5.0)	4.7 (3.0-5.0)	3.0 (2.0-4.0)	<0.0001
Sex				
Male	499 (65.7%)	301 (64.5%)	198 (67.6%)	0.16
Female	261 (34.3%)	166 (35.5%)	95 (32.4%)	
Education				
Elementary school	200 (26.3%)	120 (25.7%)	80 (27.1%)	0.88
High school	303 (39.9%)	177 (37.7%)	126 (43.0%)	
University	257 (33.8%)	160 (34.3%)	97 (32.9%)	
Postgraduate	100 (13.1%)	64 (13.7%)	36 (12.2%)	
Occupation				
Unemployed	115 (15.1%)	67 (14.3%)	48 (16.4%)	0.48
Retired	115 (15.1%)	67 (14.3%)	48 (16.4%)	
Other	530 (69.8%)	333 (71.4%)	197 (67.2%)	
Marital status				
Married	484 (63.7%)	293 (62.7%)	191 (65.2%)	0.68
Single	276 (36.3%)	174 (37.3%)	102 (34.8%)	
Family				
Family size	3.2 (1.0-5.0)	3.1 (1.0-5.0)	3.3 (1.0-5.0)	0.88
Number of children	1.1 (0.0-3.0)	1.1 (0.0-3.0)	1.1 (0.0-3.0)	0.88
Number of grandchildren	0.1 (0.0-1.0)	0.1 (0.0-1.0)	0.1 (0.0-1.0)	0.88
Number of great-grandchildren	0.0 (0.0-0.5)	0.0 (0.0-0.5)	0.0 (0.0-0.5)	0.88
Number of great-great-grandchildren	0.0 (0.0-0.5)	0.0 (0.0-0.5)	0.0 (0.0-0.5)	0.88
Number of great-great-great-grandchildren	0.0 (0.0-0.5)	0.0 (0.0-0.5)	0.0 (0.0-0.5)	0.88

*p-values were calculated for the chi-square test for categorical variables, and the t-test for continuous variables.

スライド 10：質問紙の実施法の選択

質問紙の実施法の選択。これもさっき言いましたけど、目の前の学生とかレジデントに対して、すると結構・・・があるんだけどそれ以外にも、リサーチとしてするんだったらあります。対面：アシスタントが直接ですね、学生とかそのレジデントに行くと対面すると結構長い質問が出来る。そこでアドバイスできるからね。複数また・・・の質問パターンがあります。人がそこにいれば色々分からないことも調整できるからそれで回収率も高い。ですが、非常にコストはかかる。時間が多く必要。状況によって、医学生や研修医相手はプレッシャーとなる場合もあります。典型的な目前での配布と回収ですね。授業評価とかですね。どこかのセミナーにきた後に回収とかあります。これは質問による対応も可能ですしコストが安い、回収率が高い。サンプルサイズが稼げないですね。せいぜいこの後で今回の評価もありますけど、せいぜい数十、100 とかありますね。勿論プレッシャーがかかります。



<ul style="list-style-type: none"> • 対面 <ul style="list-style-type: none"> - 長い質問紙が可能 - 複数の質問パターンの実施が可能 - 回収率が高い - コストがかかる - 時間が多く必要 - (医学生や研修医相手) プレッシャーとなる 	<ul style="list-style-type: none"> • 目前での配布と回収 <ul style="list-style-type: none"> - 質問による対応が可能 - コストが安い - 回収率が高い - サンプルサイズが稼げない - (医学生や研修医相手) プレッシャーとなる
--	--

スライド 11：質問紙の実施法の選択

時々教務に色んなものが来ますね、郵送で。コスト安くて人員も最小。結構対象者が十

分に考えて記載する時間的余裕がありますけど、やはり回収率非常に低い、欠損値もですね、回収はして出来ただけど、項目別に欠損値が入ってくるのが多いです。昨日少しお話が出ただけど、回収率が低いのは、相手の問題かというのはこれやっぱり非常に難しいところで、やはり、やり方が悪いように僕個人的には思っているんですね。とにかく何とかを強制したところで、相手がレス

質問紙の実施法の選択

- 郵送
 - コストが安い
 - 人員も最少
 - 対象者が十分に考えて記載する時間的余裕
 - 対象によっては回収率が低い
 - 欠損値が多い
- 電話
 - 日本では余りされていない
 - 比較的安いコスト
 - ランダムダイアルで一般化可能性が高い
 - データ回収期間が短い
 - 医学教育では？
 - 個人的・微妙な質問には不向き

2006/2/15 Taiheki-Mediatrics, Center For Medical Education Kyoto University Graduate School of Medicine 11

ポンスしなかったらそれはシステムの問題。教育も同じで、僕らが何ぼ教育しても相手が・・・していなかったらそれは学生の責任かという違うような気がします。こっちが悪いリフレクションしなければならない。調査も同じで回収率も悪ければ相手が悪いといっても仕方ないので、上げる手段をいくつか考える。結構郵送というのはコスト安い・・・ですね。電話というのは、電話というのは結構あまりされていない。比較的安いコストだけど、学生相手にはあまりしないけど、一般的な調査ではランダムダイアルでアメリカでよくやっていますが、一般化可能性が高い、データ回収期間が短い。医学教育ではこれほどまでできるかですね。私も疑問です。個人的、微妙な質問に対しては不向きです。

スライド 12：質問紙の実施法の選択

インターネット。これも最近はやっていますね。コスト比較的安くて、回収が早いのと、入力の手間が不要で、ゆっくり考える時間があるのだけれど、代表性にかなり問題がある。これは色んなデータがあります。それから回収率が分からない場合。医学生なら出来るんだけど、学生の授業評価をコンピュータやったことあります？うちはいっぺんやったんですけどほとんど回収率は 10%

質問紙の実施法の選択

- インターネット
 - コストが比較的安い
 - 回収が早い
 - 入力の手間が不要
 - ゆっくり考える時間がある
 - 代表性に問題
 - 回収率が不明
 - (医学生・研修医)回収率悪い

	全国ランダムサンプル (N=1109)	ネットユーザー (N=526)	ネット非ユーザー (N=579)
平均年齢 (歳)		40	56
男性 (%)		52	48
学生 (%)		13	2
通院中 (%)		22	41

(標本、全国ランダム調査 2006)
2006/2/15 Taiheki-Mediatrics, Center For Medical Education Kyoto University Graduate School of Medicine 12

そこらもなく、すぐ諦めたんだけど、わるいところですよこの辺。代表性について少し 2 年前うちで調査をして、ネットユーザーと非ネットユーザーに分けて、ものすごくシステムティックに違うんですね。これ有意に年齢も性別も違う、学生の頻度も違うんだけど、・・・とにかくインターネットのユーザーと非ユーザーではかなりシステムティックに違うのでこれは代表性にだいぶ違うということに意識してください。一般化可能に出来ないというのが我々の結論です。

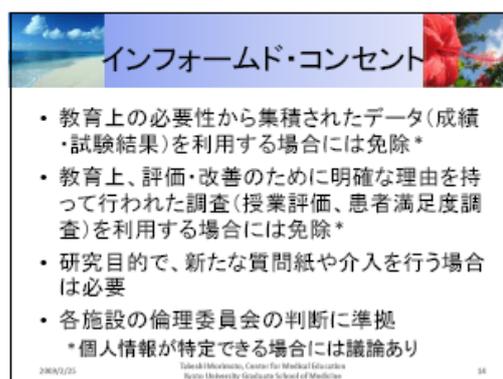
スライド 13：研究手続き上の問題

研究手続き上の問題を少し話します。

スライド 14：インフォームド・コンセント

インフォームド・コンセント。ペーパーにするかしないかはともかくとして、インフォームド・コンセント要るか要らないかについては、教育上に必要性から集積された成績で

ータ(成績・試験結果)を利用する場合には免除。いちいちこの成績を解析に使うって良いですかって学生に聞かないですね。一般的には、教育上必要なデータを解析に利用する、ゴールがですね。教育上の目的、リサーチ。同じくですね、教育上必要な授業評価などありますけど、明確な理由をもっての授業評価、患者満足度調査というような場合免除。これは要するに自分たちの分析。但し、研究目的で、新たな質問紙や介入を行う場合は必要。



インフォームド・コンセント

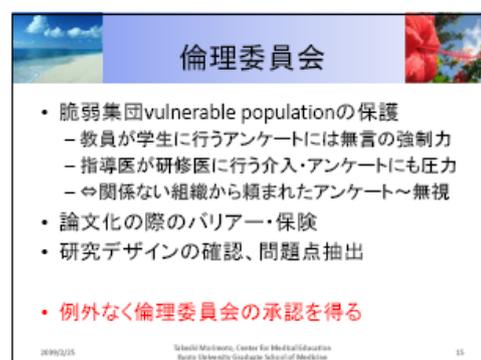
- 教育上の必要性から集積されたデータ(成績・試験結果)を利用する場合には免除*
- 教育上、評価・改善のために明確な理由を持って行われた調査(授業評価、患者満足度調査)を利用する場合には免除*
- 研究目的で、新たな質問紙や介入を行う場合は必要
- 各施設の倫理委員会の判断に準拠
 - *個人情報特定できる場合には議論あり

2019/3/25 School Medicine, Center for Medical Education Toyo University Graduate School of Medicine 14

要。ゴールが研究か要するに実務上に必要性かというところなんです。ただこれに関しても倫理委員会の判断について準拠した方がいいですね。施設で違うと思います。だから成績データとか、教育上のいろんな評価に関しても、個人がかなり同定される場合は議論のあるところだと思います。ゼネラルなルートとしては教育上必要だから、授業評価するときに、アンケートはとつてもいいですかというのはしないと。それをリサーチに使う場合は新たにする場合に要る。では既存のデータの場合はどうするのかという場合に関してはいくつかあって、もう一辺取り直すという倫理委員会もあるでしょうし、もう要らないという倫理委員会もあるでしょう。それはインフォームド・コンセントして。

スライド 15 : 倫理委員会

それはインフォームド・コンセントなしで IRB 倫理委員会に関しては、別で、基本的で **vulnerable population** ですね。脆弱集団を保護するということは大原則で、それから考えると、我々が学生に行うアンケートには基本的無言の強制力が働いています。同じように指導医が研修医に行う介入・アンケートにも基本的に圧力がかかる。そうすると彼らは基本的に **vulnerable population** に該当する。倫理委員会原則としてはこの保護といわれています。関係のない組織から頼まれたアンケートには無視しますね。関係あっても無視しますね。圧力がないから、こっちは圧力あるから、みなさんどうするのか、まずかけてもいいけど、基本的に倫理委員会通す。なおかつ論文書く場合のバリア、保険ですねとか IRB 通したかがかなり問われます。



倫理委員会

- 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

2019/3/25 School Medicine, Center for Medical Education Toyo University Graduate School of Medicine 15

日本での現、緑の雑誌に関してはそんなに厳しくないと思っています。勿論研究デザインの確認問題点抽出です。僕も大学の倫理委員会やっていますが、ほとんど質問紙見てもわけがわからないと思うんだけど、別に研究デザインをチェックすることが倫理委員会の目的ではないので、私はこの一番上の **vulnerable population** が保護されていたら OK で、研究デザイン・・・でも・・・という感じで私は回答しています。それは論文になるとか・・・は研究者の問題なんだけど、倫理委員会の目的はこの **vulnerable population** の保護なので、

これが出来ていたら僕は基本的に通しています。基本的には例外なく倫理委員会の承認を得るということです。では振り返って、過去に研究目的ではなく、学事上の目的でとったデータとか授業評価で採ったデータがあると、これを論文化しようとする場合にはお勧めするのは後付で良いから、倫理委員会出してください。すでに既存のデータがこれこれいうのがあって、今までつかっていたけれども、社会的に非常に意味があるので論文化するので、ということで、基本的にはそのときの条件として個人的な情報は切りますとかですね。そういうことをきちんと書いて IRB に出せば多分それはペーパーになっていくので、緑の雑誌でも、段々緑の雑誌でも厳しくなってくるので、基本的には通したほうがいい。過去のデータを使って個人情報削除された場合に倫理委員会が基本的にごたごたいう問題ではないので、基本的にはデータとかちゃんと管理しておけばいいので、OK なんです。何ヶ月もかかるようだったら、何か倫理委員会に問題があるのであって、本来 1 週間で出して、というのが本来あるべき姿です。

スライド 16：医学教育研究 研究倫理指針（案）

一応医学教育学会のですね、この研究会では一応医学教育研究における研究倫理指針を作りました。去年から少しそういう形であまり見せないですね一応案を作って、この前の理事会に大西先生が出してという形で、この指針は医学教育の質の向上を図る上での医学教育に関する・・・遵守すべき事項を定めることにより・・・という形で作りました。これは近々でるんですよ。（「出るだろう」といわれています）

**医学教育研究
研究倫理指針(案)**

- 日本医学教育学会医学教育研究開発小委員会
- (目的)この指針は、医学教育の質向上を図る上での医学教育に関する研究の重要性と学問の自由を踏まえつつ、個人の尊厳及び人権の尊重、個人情報の保護、その他の倫理的観点及び科学的観点から、医学教育の研究に携わるすべての関係者が遵守すべき事項を定めることにより、研究対象者や社会の理解と協力を得て、医学教育に関する研究の適正な推進を図られることを目的とする

©Tsubaki Murakami, Center for Medical Education, Kyoto University Graduate School of Medicine

びんくの雑誌というのは、今緑の雑誌ですけども色が変わるんです。この小委員会の中では「緑の雑誌」といっていたけれど、今度からはピンクの雑誌とお互いに呼ぶことになっていまして。

スライド 17：解析上の問題

解析上の問題。

スライド 18：質問紙の解析

質問紙を皆さん作っておられて、例えば試験が難しい「はい・いいえ」、出席が難しい「はい・いいえ」、挙手が難しい「はい・いいえ」、予習が難しい「はい・いいえ」、教授が気難しい「はい・いいえ」ですね。講義の評価は「はい」の数の個数で評価していいかこれはある例ですけども。これでいいのか。これある一つの例ですが、考えてください。これでいいのかどうか。

質問紙の解析

- 試験が難しい [はい・いいえ]
- 出席が難しい [はい・いいえ]
- 挙手が難しい [はい・いいえ]
- 予習が難しい [はい・いいえ]
- 教授が気難しい [はい・いいえ]

• 講義の評価＝はいの数の個数？

©Tsubaki Murakami, Center for Medical Education, Kyoto University Graduate School of Medicine

スライド 19：質問紙の解析

ではちょっと変えて rating scale にそって考えて、「いいえ・中・はい」でこれ全部 5・5・5・・・で全部で 25 で、講義の評価はスコアの総数でいいか、良いか悪いかというのは situation によるんだけど、いろいろの問題があると思います。

スライド 20：2 元変数や順序変数の積算

先程の「はい・いいえ」の 2 元変数. とか「いいえ・中・はい」という形の積算・・・することに関しては、すでに確立されている調査票でたぶん次に、小田先生が話すると思うけど、どっかですでにどこかで出ているようなものとかを足して良いようなチェックされている調査票だったら基本的に OK です。そうでない場合、基本的に査読の段階では ALWAYS Suspicious です。その研究者が作った調査項目を足してやっていることに関して必ず怪しまれます。なので、今作っておられるものに関しても、解析の方向性に関してはやはり作る前から検討しましょう。

スライド 21：Rule of Thumb

Rule of Thumb. いわゆる大雑把なルールですね。私が提案するルール。昨日これ Thumb of Rule ルールの指？かといわれて直したんです、クロンバツハの α で方向をチェック、昨日大西先生話しました。全体同じことをチェックしていたら、これミニマムです。OK 何だけど、医学教育雑誌今のところ、それだけではちょっと怪しまれるかもしれません。少なくともマスト。先程有無しがありましたよね。2 元変数を個別にカテゴリカルで解析するのが大事ですね。カイとか Fisher です。足すのも評価するのも良いんだけどその手前でそれぞれ個別の項目を解析する。順序変数を個別に解析するノンパラメトリックで解析する。いわゆる、Wilcoxon で解析する。もとの積算でやっているときの解釈で、例えば教育の集団で、.. こういう積算で解析したものと個別の項目とでやった解析、例えば rating scale でやった解析が、consistency ほぼ同じことをいっているということが非常に大事。これ結果の安定性ですね。ここでカテゴリカルなノンパラで解析で出てくるような解析結果とこっちの足して割った、解析が違ふことといっているのは、これは確実に間違いです。少なくともここまでは絶対しないといけない。もう一つは多変量の従属変数、統計ソフトだと・・・ですね。例えば足したものを別のいくつかのファクターで、定量するような多変量に関しては、多変量の従属変数というの

	いいえ	中	はい
• 試験が難しい	1	2	3-4-5
• 出席が難しい	1	2	3-4-5
• 拳手が難しい	1	2	3-4-5
• 予習が難しい	1	2	3-4-5
• 教授が気難しい	1	2	3-4-5

• 講義の評価＝スコアの総数？

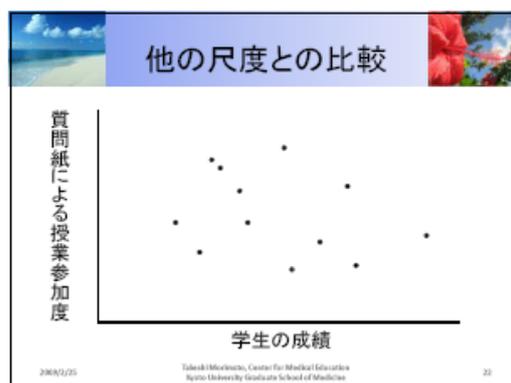
- 既に確立している調査票で積算もチェックされている—OK
- そうでない
 - 査読の段階ではALWAYS Suspicious
- 作成前から解析の方向性を検討

- Cronbach's α で方向チェック—minimum
 - Ummmmmm.....OK (in 医学教育系雑誌), BUT
- MUST
 - 2元変数を個別にカテゴリカルで解析 (χ^2 , Fisher)
 - 順序変数を個別にノンパラメトリックで解析
 - Consistencyをチェック
 - 多変量の従属変数—NO
- SHOULD
 - 主成分分析・因子分析

はかなりそれぞれのロジスティックでも構わないけど、ではいっぱい良く使われる典型解析では、基本的にはサンプル数は沢山あるんですよね。それは基本的なこのレベルでの足しただけではまもられていません。少なくともこっち側の・・・ですね、例えば成績の可否に対してロジスティックな変数に対して、積算変数とか・・・性別こっちが予測因子がこちらは基本的にはモデルでも何でもないので、従属変数サイドに関してはモデルにしても良いんだけど、従属変数に関してはモデルにいくつか条件があるのでこれがたぶん使えない。いくつかの条件をこなせばいいわけですが。勿論条件というのは、主成分分析、因子分析という形でありますし、それぞれの項目の同じウエイトがあって、同じバックグラウンドにある同じ背景にある先言おうとしているという前提があれば、勿論・・・検定に使ってもいいんだけど、その手前だとやはり、使うべきではないというのがマストです。ある程度ですね、エディターサイドも勿論、きれいなところでやるのはベストなんだけれども、・・・そうでないレベルもあるということで、緑の雑誌がどうかはいいませんが。そんな感じかなと思っています。

スライド 22 : 他の尺度との比較

あといくつかの尺度としては、他の尺度ですね、例えば学生の成績というはある程度客観的であって自分が作った質問紙、なんか全然関係がないというのが、これはちょっと尺度おかしい。自分が大学の授業評価と学生成績データと全然合わないというのが、やはりそれは考えなのでそういう他の尺度との比較、これも少し大西先生が話して少しいうとバイアスに関係



してくるけど検討したほうが良いですね。パイロットテストの重要性ですね。10人でも20人でも良いからパイロットテストをやって、ありとあらゆる・・・をチェックしてね。僕、統計家なんだけどあまり統計ソフト気にしないですねどんな状況でも。Nのバランスとか決まってくるので、それで、effect sizeのほうはずっと重要ですね。どう関連するか。統計の人ほどあまりp値気にしないというのは結構トレンドで、effect sizeが重要。どう違うか、例えば、このカードでいうときちゃんと統計的でなくても良いから、これが・・・ばらついていてもいいからなどバランスいいかが重要。

スライド 23 : いい加減な質問紙

例えばいい加減な質問紙の例です。これさっきの(例)、シューっとね(大きな丸をつけてしまう)。(質問文を)読んでいるのかみたいですね、(回答を)飛ばしてしまっってミッシングとかですね。時々こういうの出てきます。

スライド 24 : Solutions

いい加減な質問紙

	いいえ	中	はい
・試験が難しい	1	2	3-4-5
・出席が難しい	1	2	3-4-5
・拳手が難しい	1	2	3-4-5
・予習が難しい	1	2	3-4-5
・教授が気難しい	1	2	3-4-5

2009/2/25 Tohoku University, Center For Medical Education Kyoto University Graduate School of Medicine 23

Solutions として、不真面目な回答ですね。不真面目な回答ははずしたいというのは非常に我々の意向ではあるんですね、僕なんかも授業評価やると、やはり数人「森本先生はふざけている」という自由コメントがくるんですけど、100人中5人ぐらいは、毎年無視しているんですけど、それを無視するということが、論文には基本的には入れなければならぬけれど、どうしてもいい加減なやつをはずす場合は、

Solutions

- 不真面目な回答
 - “いい加減なやつ”の基準—逆転項目
 - 感受性分析
 - メインの解析(フルデータ)とサブ解析(いい加減なやつを除いたデータ)
 - Consistencyが必要
 - 論文に明記
- 欠損値
 - Imputation～勧めない
 - 論文に欠損値の扱いを明記

2019/12/15 Osaka University, Center for Medical Education, Kyoto University Graduate School of Medicine 24

は、いい加減なやつという明確な基準がないといけませんね、完全に逆転項目で、ここに答えるべきではないことに勝手に答えている。全部丸をしているんですね。

そういうのが明確にあればはずしても良いけれど、原則ははずさない。どんな不真面目なやつも。

感受性分析：はずしたい場合、はずさない場合2つあって、基本的にはメインの解析というのはフルデータははずさない、もし、明らかにふざけたいい加減なやつ必ずいて、それに対しては一定の基準があればですね、時々混り気が来るでしょう。質問紙調査をしたらですね。ふざけて書くやつとかですね。除いたデータを使っているんですけどサブデータ、ある程度その結果に consistency が必要です。勿論論文にも明記ですね。これこれの条件に当たった人は、・・・のほうが良いです。欠損値に関しては基本的には原則としては、欠損値は全部使う。欠損値を含んだまま解析するので、ある項目に関しては100人、この項目に関しては85人、この項目に関しては75人とかきちんと書きます。Imputationは統計屋さんに頼んだほうが良いけれど、基本的に勧めない。Imputationというのは別のデータから採ってきて、そのこのミッシングな回答について、予測して入れてしまうというやり方があるのですけれど、これは勧めない。論文には欠損値の扱いを明記しましょう。ということですね。欠損値が例えば5個の項目があって、それぞれに10%・10%・10%の欠損値があった場合、全部多変量に入れてしまうと使える例数が半分になってしまいます。基本的に欠損値が起きないデザインにする、パイロットテストをして昨日出た。素人さんにいっぺん読んでもらうとか、違う人にやってもらうとかという形で、少なくとも欠損値が起きにくい situation を作らないといけませんね。

スライド 25 : 多重比較

多重比較の問題ですね。これちょっと、手元のデータで・・・25項目ぐらいですね。真面目な連中が書いたのにわけてあるんですね。2つぐらい p が出ていることがわかりますね。たぶん聞いたことがあると思いますが、0.05 という・・・なので20回やったら1回ぐらい別に差がなくても有意になるというのが単純比較ですね。2個のサイコロ転がし

多重比較

評価項目 (%)	対象者 (N=122)	正しい (N=124)	不真面目 (N=20)	p値	評価項目 (%)	対象者 (N=122)	正しい (N=124)	不真面目 (N=20)	p値
課題1	87 (71)	83 (67)	13 (64)	0.1	課題14	147 (11)	136 (11)	9 (24)	0.61
課題2	202 (16)	251 (20)	11 (54)	0.6	課題15	165 (13)	158 (12)	7 (18)	0.2
課題3	200 (16)	200 (16)	12 (62)	0.1	課題16	9 (0.7)	9 (0.7)	1 (3)	0.1
課題4	191 (15)	176 (14)	12 (62)	0.063	課題17	43 (3)	41 (3)	2 (5)	0.5
課題5	148 (11)	146 (11)	9 (42)	0.83	課題18	43 (3)	41 (3)	2 (5)	0.5
課題6	131 (10)	125 (10)	6 (30)	0.2	課題19	242 (19)	234 (18)	11 (27)	0.69
課題7	23 (2)	21 (2)	2 (1)	0.89	課題20	12 (0.9)	11 (0.9)	1 (3)	0.2
課題8	131 (10)	125 (10)	6 (30)	0.2	課題21	269 (21)	260 (21)	12 (30)	0.1
課題9	829 (66)	835 (67)	24 (120)	0.96	課題22	23 (2)	21 (2)	2 (5)	0.69
課題10	222 (17)	217 (17)	5 (25)	0.5	課題23	131 (10)	125 (10)	6 (30)	0.2
課題11	108 (8)	102 (8)	7 (35)	0.5	課題24	829 (66)	835 (67)	24 (120)	0.96
課題12	44 (3)	42 (3)	1 (5)	0.4	課題25	222 (17)	217 (17)	5 (25)	0.5
課題13	202 (16)	251 (20)	11 (54)	0.6					

2019/12/15 Osaka University, Center for Medical Education, Kyoto University Graduate School of Medicine 25

て1のぞろ目が出るか6のぞろ目が出るかによって p が 0.05, 何回もやるといつかサイコロがどんなに正確であっても, 0.05 になるので, これは問題ない.

スライド 26 : Solutions

Solutions としては, 予め考慮は必要です. 多重比較に関しては, 多重比較が出てきた場合に, 沢山の項目を比較して1個だけ有意なものを強力に主張するのは困難になります.

ハイリスクが 0.0001 未満とかだと良いですけど, いくつかやって 0.001 みたいなことするのは困難です. もちろん post hoc 補正ができるのであればベストです. 0.0001 未満なので, 補正できればベスト

です. いくつかパターンがあっていくつかの Bonferroni があった, 最近遺伝子の解析が入っていてとにかく片っ端から測って行ってチェックすると, かなりの率で偶然でリスク・False Discovery Rate・時々補正したりとかですね. 後臨床試験とかたくさんの方の心筋梗塞とか, 色々なことに関してチェックするのが overlapping adjustment, 色々あります. 今日はここには触れません.

MUST としては少なくともアウトカムの順位付けが論文に書くときは, 少なくともこれがメイン. アウトカムになるという形で, 本当は事前からね, 後からこれが重要だと書くという人いますけど, 補正の有無, ・・・に関しては, たくさんあって, どうするんだ. 補正をするのかしないのか, しない場合には, どういう justification 要するに理由をつける. それから勿論ペーパーの最後に limitation 書くというのが必要です.

スライド 27 : 世界三大宗教

連続変数で, t 検定すぐに出す平均値とか多変量解析とかということを書いています.

スライド 28 : 投稿・発表時における問題

投稿・発表時における問題.

スライド 29 : 原則

原則ですね. 何をやったかが, 全く研究のことを知らない人が読んで, 一目瞭然. clinical の content はともかくとして, method に関して何をやったかについては一目瞭然でないといけません. オリジナルの質問紙に関しては例えば論文でテーブルを設けるとか, figure とか appendix で記載する. それから妥当性の検証したかしなかった場合それなりの理由付け, コメント

Solutions

- 予め考慮は必要
- 強力に主張するのは困難
- Post hoc 補正ができるのであればBEST
 - Bonferroni
 - False Discovery Rate
 - Overlapping adjustment
- MUST
 - アウトカムの順位付け
 - 補正の有無への理由 Justification/Limitation

2019/2/25 Tsukuba University, Center for Medical Education, Sports Science Graduate School of Medicine 26

世界三大宗教

- 「連続変数」教
 - 神経ステーション 22/37(59%)
 - 自己評価 8/10
 - 出席日数 21日
- 「平均値±SE」教
 - 平均自己評価点数 3.7±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 ($p < 0.05$)

2019/2/25 Tsukuba University, Center for Medical Education, Sports University Graduate School of Medicine 27

原則

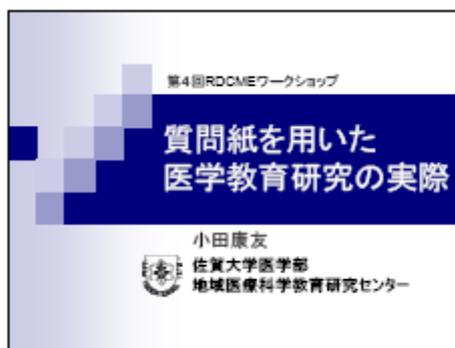
- 何をやったかが, 全く研究のことを知らない人が読んで, **一目瞭然**
- オリジナルの質問紙
 - 論文での記載 (Table/Figure/Appendix)
 - 妥当性の検証・コメント
 - 積算尺度への justification

2019/2/25 Tsukuba University, Center for Medical Education, Sports University Graduate School of Medicine 29

レクチャー4：質問紙調査を用いた医学教育研究の実際（質疑応答）… 小田委員

スライド1：質問紙を用いた医学教育研究の実際

森本先生に引き続きまして、話をしていきます佐賀大学の小田と申します。今日、今からお話するのは、質問紙を用いた医学教育研究の実際ということで、経験談と申しますか、振り返ってみますと、こういう研究を大西先生と10年以上もやっているんだなということにこれを作りながら愕然としたのですが、その中で色んな失敗ですね、皆様方にこういうことをすると失敗しますよ、同じことを是非



されないようにして下さいという形で具体的なお話をさせて頂きたいと思っています。特にポイントは、質問紙調査をするときには、質問紙の出来ということに関して最新の注意を払われると思いますけど、勿論それはそうなんです、それ以外の要素でも落とし穴がたくさんありまして、そのために全然違った調査になってしまったという経験を沢山しましたのでそれについて、お話ししてお役に立てていただければと思っています。

スライド2：本日のお話

今日の内容ですが、何の経験をお話するのかと申しますと、私は佐賀医科大学卒業でありまして、総合診療部卒業で今は医学教育の専任になっていきますけれども、総合外来診療というところで、ずっと診療及び教育、研究をやってまいりました。その中でずっと目標の一つとして掲げていたことは、医学生、研修医の臨床能力、実習とか大学で学んだことの到達度評価を臨床能力、現場での臨床能力評価でやりたいということがありました。試験やレポートで実習を評価するというのはおかしいだろうということで実際の診療能力で調査できないかと、そのための評価表を開発できないかということをやってまいりました。そのために色んな調査票を用いたのですが、最初はですね、この受容度調査ということをやりました。これは先行研究といいますか、僕らの先輩といいますか、上司の先生がやっておられたものを最初は引き継いでやったのですが、これは自分たちで作ったオリジナルでした。それを米国の内科専門医会の患者満足度評価表というものに途中から切り替えてやってきた、という流れがあります。PSQ (Patient Satisfaction Questionnaire)と申しますけれども、それを用いて色んな研究をやってまいりました。パイロットスタディをさんざんやりまして、日本語版のPSQを作ったと。それからPSQを用いて学生の臨床能力の評価をやってみた。頓挫しましたが、それからプログラム評価、大学のコミュニケーション教育プログ

The slide is titled '本日のお話' (Today's Talk). It contains a bulleted list of topics:

- 佐賀大学総合外来における、医学生・研修医の臨床能力評価法開発のための取組
- 「受容度調査」から「ABIM-PSQ」へ
 - オリジナル調査票の長所と限界
 - American Board of Internal Medicine, Patient Satisfaction Questionnaire
- ABIM-PSQを用いた諸研究
 - Pilot Studyの数々、日本語版PSQの作成
 - PSQを用いた臨床能力評価／プログラム評価
 - PSQの信頼性・妥当性の検討から360度評価へ

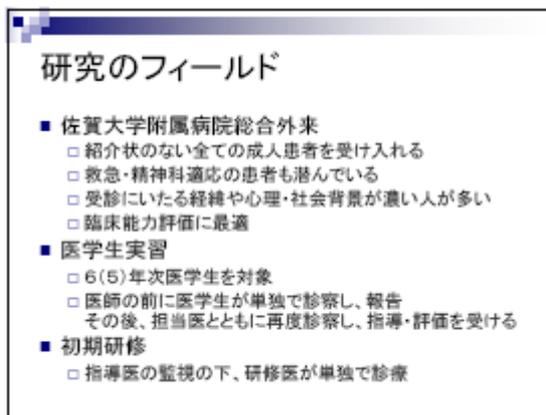
ラムの評価をやってみたり、総合外来実習の評価をやってみようということに使ってみたり、いろんなことをやりまして現在に至っております。今、うまくいったのかといいますと、まだ悪戦苦闘の途中ですけれども、そのお話しをするということです。

スライド3：研究のフィールド

まず、研究のフィールドについてお話しします佐賀大学附属病院総合外来というと非常におもしろいところで、佐賀大学は元々研究者ではなくて臨床医を育てるということを目的としていますので、非常に実践的な教育を佐賀医科大学が作られた当初から実践的な教育をということで、臨床実習はかなり実践的にやられていました。私も5年生に上がったとき、・・・心電図も取れないのかといわ

れて怒られて・・・そんな・・・という感じでしたし、当時からルンバールはしますし挿管もしましたし、今ではとても許されないことを沢山していました。その中でも総合外来の実習というのは、総合外来というのは紹介状のないすべての成人患者を受け入れています。佐賀大学は「誰でもいらっしゃい」ですね。紹介状なんて関係ない、総合外来で診ますから。必要があれば我々が専門家に紹介しますよ。遠慮なくいらっしゃいということになっています。ですから紹介状のないすべての成人患者を受け入れています。その中には救急とか精神科適応の患者も潜んでいます。精神科的な心の病んだ人も非常に多いですね。

それから受診にいたるまでの経緯や心理的な、社会的な背景が非常に濃い人が多いという点。そういう点で何が来るか分からない、問題が身体にあるのか、心にあるのか、社会的な関係にあるのか、全意図の関係にあるのかというあらゆるパターンがありますので、臨床能力を総合的に評価するという点では非常に最適なフィールドであります。その医学生実習では何をしているかと申しますと、5、6年次、主として6年次の学生を対象にして、5年で全科のクラークシップを終えた6年生がやってきて総合外来をやるのですが、医師が診る前に医学生に入ってもらって、カルテを渡されて医学生がその患者さんにご挨拶して「診させて頂いてよろしいでしょうか」と承諾をとって。学生が自分で診察室に、個室に2人で入って、「どうなさいましたか」そこから始まる。診終わった後、患者さんは待合室で待っていて、学生はまとめて、指導医のところに「先生こんな状況でした」ということを報告する。その後に担当医と一緒に診察しまして評価を受ける、そんなふうな面白い実習をやっています。昔は共用試験、OSCEとかなかった頃、学生もあまり慣れていなかったもので本当に学生も緊張していて、話もできない途中で緊張してがたがた震えて汗だらだらかいて、患者さんの方から『大丈夫ですか?』と言われて、失礼しますと帰っていく者もおりましたし、非常に何かチャレンジングな実習を昔からやっていたなという印象です。佐賀だから患者さんも受け入れたということもありましたが、非常におもしろい実習をや



研究のフィールド

- 佐賀大学附属病院総合外来
 - 紹介状のない全ての成人患者を受け入れる
 - 救急・精神科適応の患者も潜んでいる
 - 受診にいたる経緯や心理・社会背景が濃い人が多い
 - 臨床能力評価に最適
- 医学生実習
 - 6(5)年次医学生を対象
 - 医師の前に医学生が単独で診察し、報告
その後、担当医とともに再度診察し、指導・評価を受ける
- 初期研修
 - 指導医の監視の下、研修医が単独で診療

っていました。その後初期研修が始まった後は、指導医がカーテンの後ろで聞いておりまして、監視の下に研修医が単独で診療していますし処方まで出すというような、これがフィールドで、これが今日お話するのは医学生実習のお話しです。

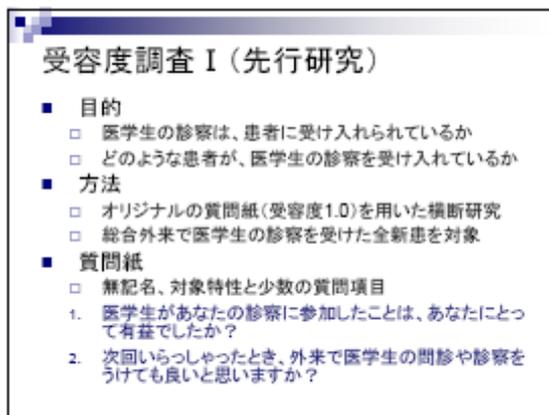
スライド4：受容度調査Ⅰ（先行研究）

まず受容度調査ということ。これは私の上司がやっていた研究だったのですが、そういうチャレンジングな実習をしておりますので、この実習は本当に患者さんに受け入れられているのか、本当は嫌がられているのではないだろうか、というようなことを上司が非常に気にしておりました。目的としては、医学生の診察は、患者に受け入れられているか、あるいはどんな患者さんが嫌がっているのか、どんな患者さんが医学生の診察を受け入れているか、その辺を調べて受け入れてくれそうな人に渡そうと、学生を担当させようというような目的でやられました。オリジナルな質問紙を用いた横断研究で、総合外来で医学生の診察を受けた全新患ですね。新患には必ず学生がつきます。要するに、新患の全数調査を対象としてやりました。質問紙は、残っていないのです。マックユーザーで取り出すことが出来ませんでしたので記憶の限りですけども、これは無記名で患者の特性は非常に細かいですね。性別と職業の有無、病院まで来る時間とか待ち時間とかいろいろ書かれています。内容としては、

医学生があなたの診察に参加したことは、あなたにとって有益でしたか、というような質問。次回いらっしゃったとき、外来で医学生の間診や診察を受けても良いと思いますか、また受けても良いですか、これは受容度という観点で非常に的を得た質問だと思います。このような3つか4つの少数の質問項目と対象特性ということを中心として行いました。なぜならばどんな患者が受け入れてもらえるのかことを目的としたことでしたからです。

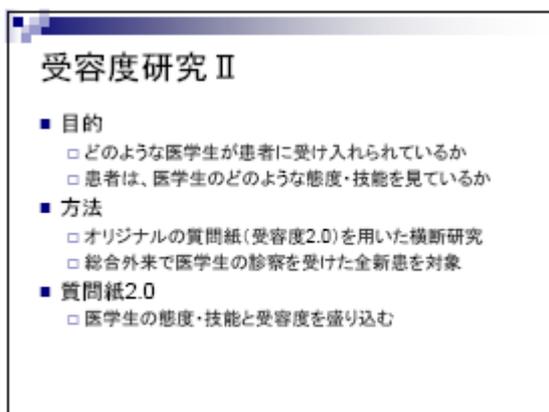
スライド5：受容度研究Ⅱ

これに対して当時若かったのか、おかしいのではないかと文句をつけました。文句をつけたのは僕と大西先生だったと思いますが、どんな患者が受け入れてくれるかではなくて、医学生をどういうふうに教育すべきか、医学生のどういう態度とか技能が患者さんに見られているか、そっちを調べるべきというふうな提言をしましたら、この上司が非常に心の広い方でそれはそうだと正しい筋が通っているお前やれといわれ、全部投げられまして、ここで得た教訓は「調子



受容度調査Ⅰ（先行研究）

- 目的
 - 医学生の診察は、患者に受け入れられているか
 - どのような患者が、医学生の診察を受け入れているか
- 方法
 - オリジナルの質問紙(受容度1.0)を用いた横断研究
 - 総合外来で医学生の診察を受けた全新患を対象
- 質問紙
 - 無記名、対象特性と少数の質問項目
 - 1. 医学生があなたの診察に参加したことは、あなたにとって有益でしたか？
 - 2. 次回いらっしゃったとき、外来で医学生の間診や診察を受けても良いと思いますか？



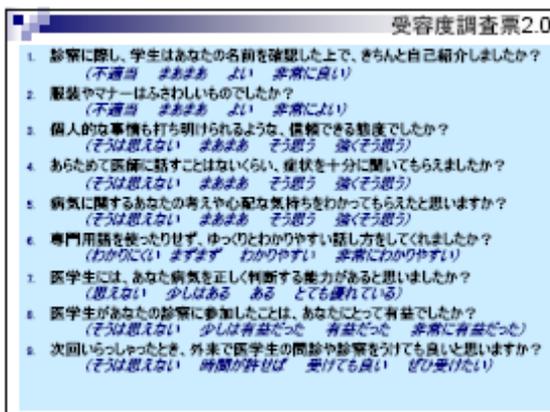
受容度研究Ⅱ

- 目的
 - どのような医学生が患者に受け入れられているか
 - 患者は、医学生のどのような態度・技能を見ているか
- 方法
 - オリジナルの質問紙(受容度2.0)を用いた横断研究
 - 総合外来で医学生の診察を受けた全新患を対象
- 質問紙2.0
 - 医学生の態度・技能と受容度を盛り込む

に乗って言うと仕事が増える」ということでした。それで、質問紙の 2.0 を作りました。これまでの受容度に加えて医学生の態度とか技能を加えた質問紙表がこの 2.0 です。

スライド 6：受容度調査票

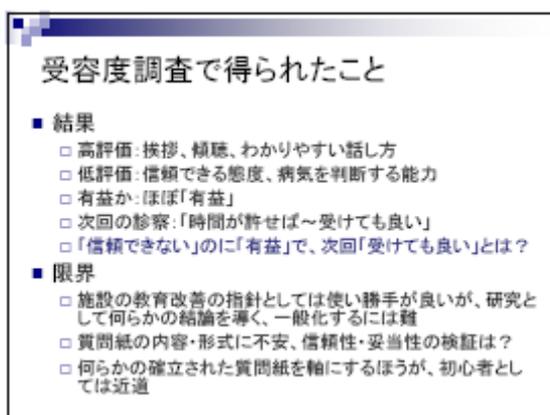
いろいろ試行錯誤して、最終的に出来上がったものですからちょっとあれなんですけど、診察に際して学生はあなたの名前を確認して自己紹介したか、服装やマナーはふさわしかったか、信頼できる態度だったか、しかもそれは個人的な事情も打ち明けられるようなというようなことを枕詞をつけています。あらためて医師に話すことはないくらい症状を十分に聞いてもらえたか、この辺、これはあとでつけたんですけど



も、ただ、信頼できる態度か症状を十分聞いてもらえたか、これだけだと答えにくいという話があったのでこういうこと、あくまでも、医師としての力量をみているのですということです。病気に関するあなたの考えや心配な気持ちをわかってもらえたか。専門用語を使わずに、ゆっくりと分かりやすく話しをしていたか、あなたの病気をあのこの 6 番目までがですね、特に様々な医学生の技能や態度を訊いていて、7以降が総合力を判断しています。医学生にはあなたの病気を正しく判断する能力があると思いますか、あなたにとって有益でしたか、それから、問診や診察をまた受けてもいいですか次回、という質問にしています。これが 2.0 のオリジナルですね。

スライド 7：受容度調査で得られたこと

これから何が分かったかと申しますと、学生に対して高い評価が得られているのは、挨拶とか、症状を十分に聞いてもらえたかということは、非常に高い評価をもらいました。あるいは、分かり易い話し方をしていたかも高い評価でした。ところが評価が低いのは、この信頼できる態度だったかというところで、せいぜい「まあまあ」という反応でしたし、病気を判断する能力については「少しはある」というくらいのところにありました。有益だったかという、「有益だ」という答えが非常に多かったんです。次回の診察については、時間が許せばあるいは受けても良いというところで、「どちらかという」と受けてもいい」というような感じのスコアになっていました。



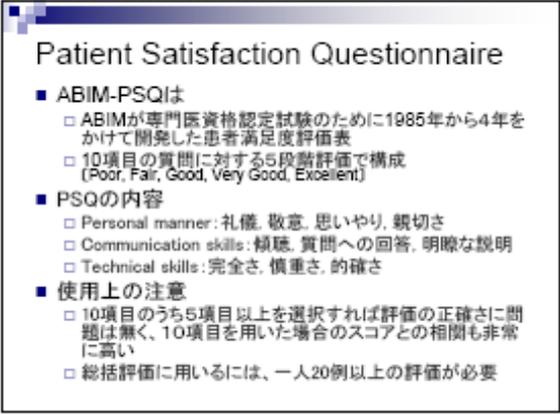
不思議なのは信頼できないし、判断する能力はないのに、有益でまた受けてもいいとい

うことはどういうことだということで、色々と聞き取り調査とか、インタビューもしてみただけですけど、要するに患者さんにとって大学病院というのは非常に壁のような場所なので、学生さんに色々話を聞いてもらってそれを主治医に伝えてもらえるということが、非常にリラックスできる。それが非常に助かる、それが本当に伝わっているということで非常にメリットがある。先生にいきなり話すというのは非常に葛藤があるんだと分かりました。特に前の病院から逃げてこっちに来てという方は話しぶりということもあって、実は学生が入ってくると助かることもあるんだということが分かってきました。

これでどうしたかといいますと、医学生の傾聴の姿勢ですね、どうぞおっしゃってくださいという姿勢はどんどん強化していっていきことによってこの実習は続けられる。けどやはり病気を判断する能力に関してはやはり非常に乏しいと思われている。信頼できる態度かといわれると、そうでもないというふうに思われている。ではそっちをもっと強化していこうというふうな形で指導の方針が出てきたわけです。ですからこの調査は、自分のところの教育改善の指針として非常に役に立ちました。けど、これを研究として何らかの結論を導いたり、一般化できるのかということとなかなか厳しいものがある。佐賀医大でやっているこの実習だからやれることであって、他では一般化できないだろう。また、質問紙の内容ですけれども、それなりに考えて作ったんですけども本当にこれいいのかな、これで本当に学生の臨床能力が評価できるのか、これで可否判定出しているのか、成績判定しているのか、大体この項目は4つにしています。5つではなく4つにしたのは真ん中に寄らないようにしたんですけども、まあまあ・・・とかいちいち1個ずつ違うわけなんですよね。たとえばこれを統計にするときにはこれを、1・2・3・4と割り振ったりするんですけども、これって間隔って同じかな？というさまざまなことが不安になってきて、これって研究にはならないよねというようなことで、非常に悩みました。まず研究者としても初心者でしたので、やはり何らかの確立された質問紙を軸にしてやっていったほうがいいんじゃないかということで、新しいそういう評価表がないか探しました。

スライド8 : Patient Satisfaction Questionnaire

そのときたまたま出会ったのが、この患者満足度評価表ということで、当時もう一人の先輩がハーバードの留学しておられたんですけど、そこで、向こうで公衆衛生をやられていたのですけども、何かデータが要るんだけど、小田君何か送ってくれない？といわれて受容度のデータを送りました。「君って天才じゃないか。アメリカの内科学専門医が作った質問紙表とそっくりだよ。センスあるね」といわれまして、そうですかということで、「それ送ってください、こっちで使ってみます」そういうことをや



Patient Satisfaction Questionnaire

- ABIM-PSQIは
 - ABIMが専門医資格認定試験のために1985年から4年をかけて開発した患者満足度評価表
 - 10項目の質問に対する5段階評価で構成 (Poor, Fair, Good, Very Good, Excellent)
- PSQIの内容
 - Personal manner: 礼儀、敬意、思いやり、親切さ
 - Communication skills: 傾聴、質問への回答、明瞭な説明
 - Technical skills: 完全さ、慎重さ、的確さ
- 使用上の注意
 - 10項目のうち5項目以上を選択すれば評価の正確さに問題は無く、10項目を用いた場合のスコアとの相関も非常に高い
 - 総括評価に用いるには、一人20例以上の評価が必要

っているうちに、結局データを持っていかれました。このときの教訓はおだてられるとろくなことはないというのが「おだてには気をつける」というのが第二の教訓でした。そういうことで、これに出会いまして、それに切り替えて、まずはやってみようと、ちゃんとしたものでやってみようと、勿論、アメリカで使われていたものですから日本で使えるかどうか分かりませんが、まずはやってみた方が質問紙調査についても勉強になるだろうということで、こっちをやりました。どうやって作られた調査表なのかというところもみましましたけども、愕然とするんですね。このくらいあるんですよ。作られるまでの4、5年かかって作っておりますし、フェイズ1の調査、フェイズ2の調査、それから10項目の質問に対して5段階の評価をするんですけども、これを他のものに作るとどうなるかもいろいろなこととして作っているんですね。質問紙一つ作るのに、どれだけやるんだろうか。特にこれが、専門医認定試験のための総括評価のための評価なので非常に厳しい。ここまでやるんだなと愕然としまして、これをこのまま使うことにしました。

PSQの内容はですね、マナー、コミュニケーションスキル、テクニカルスキルのこの3つの内容が含まれるというもので、使用上の注意としては、10項目、これは最初24項目ぐらい挙げられてそれを絞り込んだもので、そのうち5項目以上選択すれば問題ない、必ずしも10項目やらなくてもいいことになっています。それから総括評価に用いるには、一人20例以上の評価を必要とする。患者さんいろいろ違いますから、3、4人の難しい人に当たった場合は難しい。総括評価に用いるには一人20例以上の評価が必要であるというのは注意書きがありました。

スライド9：ABIM-PSQ（米国内科専門医認定委員会患者満足度票）

これがPSQの原文です。これに1個1個に、5段階の評価項目がついております。こういうものになっております。日本語に直訳したものを、次の頁に載せています。

ABIM-PSQ	
1.	Telling you everything; being truthful, upfront and frank; not keeping things from you that you should know.
2.	Greeting you warmly; calling you by the name you prefer; being friendly, never obnoxious or rude.
3.	Treating you like you're on the same level; never "talking down" to you or treating you like a child.
4.	Letting you tell your story; listening carefully; asking thoughtful questions; not interrupting you while you're talking.
5.	Showing interest in you as a person; not acting bored or ignoring what you have to say.
6.	Warning you during the physical exam about what he/she is going to do and why; telling you what he/she finds.
7.	Discussing options with you; asking your opinion; offering choices and letting you help decide what to do; asking what you think before telling you what to do.
8.	Encouraging you to ask questions; answering them clearly; never avoiding your questions or lecturing you.
9.	Explaining what you need to know about your problems, how and why they occurred, and what to expect next.
10.	Using words you can understand when explaining your problems and treatment; explaining any technical medical terms in plain language.
(Poor, Fair, Good, Very Good, Excellent)	

日本語版PSQ(1-5)	
1.	あなたに全てのことを話してくれましたか？ 話し方には誠実さと率直さ、親しみやすさを感じられましたか？ あなたが知るべきことを知らせないようなことはありませんでしたか？
2.	あなたに心のこもったあいさつをしましたか？ あなたへの呼びかけは適切で、親しみを感じるものでしたか。気難しい、または失礼と感じる態度はありませんでしたか。
3.	医師はあなたと対等な立場で接していましたか？ 態度が高慢だったり、子供扱いしませんでしたか。
4.	医師はあなたに話をさせてくれましたか？ 注意深く耳を傾け、質問しましたか。あなたが話している間、さえぎることなく自由に話をさせましたか。
5.	医師はあなたに対して十分な関心をもった態度で接していましたか？ 退屈な様子や、あなたの話を無視するようなことはありませんでしたか。
〔不十分、まあまあ、良い、非常に良い、特に優れている〕	

スライド10：日本語版PSQ（1－5）

1つ1つの質問項目が長いですね。1番はすべてのことを話してくれましたか、誠実さと率直さ、親しみ・・・非常に長い。2番は、こころのこもった挨拶をする、親しみを感じるか、対等な立場、いろんな・・・注意深く傾聴していたか、と。5番は、十分な関心をもっていたか、

スライド 11：日本語版 PSQ (6-10)

6 番は診察に対する説明, 7 番は治療や検査の選択に関する話し合いですね. 8 番は, 医師はあなたに質問がないか尋ねられました, 質問に対して分かりやすく説明・そんなことがあって, 9 番が病気について知っておくべきこと, 10 番が専門用語を使わずに分かりやすい言葉を使っていたかということでした. 確かに僕が作ったものに似てるっちゃ似てるんですけど. 微妙に違う. この 10 項目ですが, 10 項目は使いませんでした. というのは僕たちは特に学生を対象としていましたので, 治療や検査の選択とか, 病気について知っておくというのは 7 番とか 9 番はほとんど話題にのぼらない. こういうのは削除して, 最終的にこの 6 項目を選んで, 少し文章を簡潔にして使いました.

日本語版 PSQ (6-10)

- 診察をする際には, 何故, どのように診察をするのか, あらかじめ説明がありましたか. また, 診察した所見について説明がありましたか.
- 治療や検査の選択について, あなたと話し合いましたか? どのような方法があるかの説明があり, あなたの決定に助言が得られましたか. なにかをする前に, あなたの考えをきかれましたか.
- 医師はあなたに質問がないか尋ねましたか. 質問したことについてわかりやすく説明してくれましたか. あなたの質問を無視したり, 説明がなかったりしませんでしたか.
- あなたの病状について, あなたが知っておくべきことの説明がありましたか. その症状がどのように, どうして起こったのか, 今後どんなことが起こり得るか説明がありましたか.
- あなたの症状や治療について話をするとき, 専門用語をつかわず, わかりやすい, 簡単な言葉を使っていましたか.

【不十分, まあまあ, 良い, 非常に良い, 特に優れている】

スライド 12：Pilot Study の数々

この 6 項目あるいはさっきの 10 項目全部を使ってですね, 最初は, パイロットスタディを色々やりました. 一番困ったことは, 医学生の臨床能力をこれで本当に反映させることができるだろうかということですね. というのは受容度の研究のときに思っていたんですけど, 色んな要素があるんです. 質問をそのままストレートに答えてくれるわけじゃなくて, 悪いことを書いたらこの学生さん可哀相だとか, なんでまた受けてもいいんですかということ, 普通はあまり話を聞いてもらえないのにこんなに話を聞いてもらえたとかですね. 寂しい度調査? 孫に似ていたとか, 中年男性が女子学生に甘いとか, これは違うだろう. それ以外の要素があまりにも入ってはいないか. もう一つは, 当時は受容度調査票を学生が自分で「お願いできますか」といって患者さんに渡していたんですね. ここ何が困るかと言うと, 患者さんはそうやって渡されたらですね, やっぱいいことを書いてしまうということもありますし, 学生自身が評価内容を知っているわけですね. だから, そのときだけそれに沿った対応をする, そういうこともありえるわけです. だから, 単純にただやっただけでは, 質問紙をこちらに変えた, いいのに変えた, 今までと同じようにやり方でただやった. そうしたら, 万万歳のはず. そんなうまい話はない. どうやったらこの回答に医学生の能力を見て書いてもらえるかというようなことに関して, いろいろパイロットスタディをやりました.

Pilot Study の数々

- 医学生の臨床能力を忠実に反映させるために
 - 調査のタイミング: 医学生の診察直後・医師の診察前
 - 臨床能力以外の要素がどの程度影響?
 - 患者性別・年齢 × 学生の性別
 - 来院に要する時間, 待ち時間, 診察時間
 - 受療行動, 大学病院での診療歴
 - 疾患の重症度 (主観的), 患者の求める患者医師関係
 - 質問紙の構成, 配布・回収方法
 - 答えにくい質問があると, それ以下の回答は減り
 - 質問紙が裏に及ぶと, 裏面の回答は逆順に増し
 - 配布・回収方法の程 (重要)
- Back Translation

調査のタイミングですね。これは医学生が診察をします。その後指導医と一緒に診察をしますので、指導医の診察が終わってしまったら指導医の診察が混じってしまいます、患者さんの評価に。指導医の診察ではなくて医学生のことを書いてくださいといっても、それは無理な話で、ですから、医学生が診察直後で医師の診察が行われる前に配らなければならない。そうすると配る人、専門の人が一人要るということです。雇うのに年間100万かかる。それで臨床能力以外の要素がどの程度影響してくるのかということ先程申しましたけれども、患者さんの性別とか年齢と学生の性別これは、やはり影響がありました。遠くからわざわざやってきて大学病院で診てもらおうとやってきたら、専門家に診てもらおうと思ったら総合外来にまわされた、これで一つ文句がある。なおかつ医師じゃなくて学生が出てきた。こういうことも不満だろう、待ち時間、診察時間が短かった、学生が15分話を聞いた。20分話を聞いた場合と2、3分しか聞かなかったそういうことはだいぶ違う。あるいは受療行動、大学病院での診療歴、大学ってこんなもんだと・・疾患の重症度、主観的な重症度ですとか、患者さんが求める患者医師関係はどういう関係を求めているのかとか色んなこういうことを色々調べました。色々調べましたけども最初はやはり、影響あるな、やはり女性のほうが優しいな、高齢者の方が評価非常に高いな。やはり男性は女性に甘いな、とかいろいろありました。

ところがいろいろ工夫していくごとに、不思議なことにこれが段々とあまり差がなくなってきたんですね。それはどうしてかといいますと、配り方、どうやって配るかによって患者さんの回答が全然違う。全然とは言いませんが、かなり違う、これは質問紙の構成や配布、回収方法について詳しく書いていますけど、例えば一つの例を挙げますと、最初はただポンと配って、あそこの箱に入れておいてくださいというふうな形でやっていました。配る秘書さんとか雇っていなかった。そうすると、適当につけて出される。だけれどある秘書さんを雇って、しっとりした感じのいい感じの女性でその方が「これは調査のためにやっていることで医学生の成績には関係ありません」ぜひ医学生の診察を振り返って率直に書いていただけますかみたいなことをキチンと言う。その上でお渡しして、もし分からないことがありましたら声をかけてくださいと言って。なんとなく離れて見ている。そうすると患者さんは適当に書けない、欠損値もぐっと減りました。こんなに違ってくるんだなと。患者によってあまり影響されないな、安定してきたなという気はしました。数年前に僕の科研費が切れて、秘書さんが雇えなくて、ある大学院生がいて、すごく元気な女医さんなんですけど、茶髪のボーイッシュな感じで「書き終わったらお願いします」と配ったら、全部よしよしという感じでした。ところが、クロンバッハ α が0.96。回答が項目にかかわらずほとんど同じだ。こんなに違うのかと。そのときはたまたま質問紙が表裏になったんですね。質問項目が多くて、そういうのもあったと思うんですけど、配布とか回収の方法、どれだけこちらが真剣に調査をしているのかということは伝わって、同じPSQでも全然違うなという経験をしました。答えにくい質問があるとそれ以下の回答は、雑になります。診察終わって患者さん体調悪い人が多いですよ。質問紙が裏に及ぶと裏の回

答は途端に雑になります。雑になるという意味は、全部同じなるということが多いですね。答え易い質問から答えるということと出来れば表にメインの質問を置いて、裏は患者特性ということにした方がいいのかもしれませんが、ちょっとわかりません。とにかく、PSQによって、ちゃんとした質問紙さえ使えばちゃんとした調査が出来るというのは全然違うな、とつくづく思いました。

スライド 13 : Back translation

もう一つの問題は、Back translation です。これは日本語に訳しているのですが、これを日本語版 PSQ でいいのかという問題がありました。この作業は主に大西先生にやってもらったんですけども、まず PSQ を忠実に訳してもらって、使いました。ところが日本語のニュアンスが伝わらないとか、長すぎるとかあったりして、必要に応じて微調整を加えたものを日本語版 PSQ として作ります。それを日本語表現に造詣が深く

Back translation

- PSQを忠実に日本語訳して試用し、必要に応じて修正を加える ⇒J-PSQ
- 日本語版PSQを、翻訳家(日本語表現に造詣が深く、英語を母国語とする外国人)に依頼し、逆翻訳版(英文) PSQを作成
 - Original PSQを知らないことが条件。見せてはならない
- ABIMに逆翻訳版(英文) J-PSQを送付し、日本語版としての試用許諾を得る

で、英語を母国語としている外国人の翻訳家に依頼して逆翻訳、すなわち我々が作った日本語版を英語にしてもらいました。この人はオリジナル PSQ を知ってはいけません。それを出来た逆翻訳版の英語 PSQ を ABIM の専門家に送って日本語版として使って良いでしょうかという許可を得ました。この作業はほぼ大西先生にやってもらいましたが、どのくらい時間がかかったのでしょうか？（翻訳は結構お金がかかったのと時間がかかりました。米国に送って承認が出るまでに半年くらい）こういうことをやりました。大手を振ってやりました。別な評価表です、これと同じことをやって向こうに送ったけれど、使用の許可を得ようとしたけど何年も返事がないというのもあったそうです。

スライド 14 : 佐賀で使用した学生用 PSQ

この 6 項目を使いました。この 6 項目を使ってですね、学生の評価を色々やりました。

佐賀で使用した学生用PSQ

1. 挨拶は丁寧でしたか？ あなたへの呼びかけ方は適切で、親しみを感じましたか。気難しかったり、失礼な態度はありませんでしたか。
2. 対等な立場で接してもらえましたか？ 態度が高慢だったり、子供を扱っているような感じはありませんでしたか。
3. 十分に話を聞いてもらえましたか？ あなたの話をささげることなく、注意深く聞いてくれましたか。あなたへの質問は適切でしたか。
4. あなたに対して十分な関心を持った態度で接していましたか？ 退屈したり、あなたの話を無視するような様子はなかったですか。
5. 質問がないか尋ねられましたか？ あなたの質問にはわかりやすく答えてもらいましたか。質問を無視されたり、一方的な説明をしたりしませんでしたか。
6. 症状や治療についての話では、専門用語を使わず、分かりやすい簡単な言葉を使っていましたか？

評価：不十分=1 まあまあ=2 良い=3 非常に良い=4 特に優れている=5

スライド 15 : PSQ を用いた研究の概要

その概要と結果こんなふうに役立ったということをお話ししますと、一つは PSQ を使ってですね。個々の学生の臨床能力の総括的評価をしました。実習の合否判定に使いたい、優劣を付けたいというようなことをやりました。これは頓挫しました。なぜかというと 2 週間しか実習がない中で、信頼維持に必要な症例数を確保できない、20 例以上診るのは必ずしもできないということで大体

だめになりました。ではこれはもう使えないのか、使えないのかとがっかりしていたのですが、そのときに大西先生の発想でぱっと切り替わったのが、個々人の学生の評価に使うからだめなんだ。プログラム評価、佐賀大学の医学部のコミュニケーションプログラムの評価として使える。一人ひとり分けずに全部まとめてですね、大体3~400例のPSQがとれるんですけども、これを全体使って佐賀大学の医学部の学生全体の傾向として使うことによって、プログラム評価をやろうというふうにやりました。これは非常に役に立つ研究が行われました。一つは全体の傾向を見るところと、プログラム評価(2)で書いていますけども、異なる年度のスコアを比較して、どういうふうに改善しているかを見る、ということですね後は、プログラム(3)として総合外来実習2週間でこのPSQがあがっていることを示そうとしたんですけどもこれは頓挫しました。詳細は省きます。

PSQを用いた研究の概要

- 個々の学生の臨床能力の総括的評価
 - 信頼性維持に必要な症例数を確保できず⇒頓挫
- プログラム評価(1)
 - 全患者のスコアを一括して分析
 - 学生の能力の傾向を把握し、大学のコミュニケーション教育プログラムにフィードバック
- プログラム評価(2)
 - 異なる年度のスコアを比較
 - 改善の傾向を分析し、必要な教育介入を明確化
- プログラム評価(3)
 - 総合外来実習2週間のスコアを前半・後半で比較
 - 総合外来実習プログラムを評価しようと試み⇒頓挫

スライド 16 : プログラム評価の試み (1) 1998年

簡単に結果をお示ししますと、すごく昔にやったんだなとつくづく思うんですけど、98年の調査ですが、この6項目のうち全体としますと、3.38くらいで、大体「よい」から「非常によい」くらいの評価になりました。ところが項目によって非常に差があるということがわかりました。明らかに低いのは、5番：質問がないか訊ねられたか、あなたの質問は分かりやすく答えてもらったかいうところが非常に低いということが

プログラム評価の試み(1)1998年

質問項目	全スコア (n=405)	男子学生 (n=224)	女子学生 (n=181)
1	3.51(0.74)	3.47(0.75)	3.56(0.73)
2	3.57(0.76)	3.60(0.78)	3.53(0.73)
3	3.50(0.81)	3.54(0.85)	3.45(0.76)
4	3.55(0.76)	3.56(0.78)	3.53(0.73)
5	2.95(0.97)	2.93(0.99)	2.98(0.96)
6	3.22(0.93)	3.26(0.94)	3.18(0.91)
平均	3.38(0.66)	3.50(0.72)	3.43(0.68)

Cronbach α 係数0.89

がわかりました。6番の症状や治療についての話では専門用語を使わずに分かり易い簡単な言葉を使っていたかというところがやや低くなりました。これは有意差もありましたけど、低いということが分かっています。男子も女子も同じように低いと、同じ傾向があることがわかりました。それでプログラムを振り返ってみますと、確かに質問に答える訓練も確かに質問を促したり質問に答えたりというトレーニングは全然していないということで、どうしたらいいかということを考えるためにさらに学生の一部に聞き取り調査をやったりもしました。実際クラークシップで病状説明のトレーニングはどういうふうに経験しているかなということを20人くらいですか電話取材を調査していきますと、学生が患者に説明することなんてありえない。これはきちっとオリエンテーションで患者さんに情報提供してはならないということを各診療科のオリエンテーションで言われている。だから情報提

供は決してしないのが基本だと捉えている学生が多い、ということが分かってきました。だから答えられる簡単な質問に対しても拒否反応を示してしまう、という現状があったのでした。というのは昔、佐賀大学では患者さんから誘導されて病名を告知してしまったという事件がありました。『俺はガンというのは分かっているのだけれども肺ガンにも色々あるんやろ、治療法が違うやろ？俺はどのタイプね』といわれて「あなたは・・・ガンで・・・」と学生が言ってしまったという事件がありました。そういう事例もあって、学生は怖がっていて、ではこのトレーニングどうする？かということでプログラムを改善しました。

スライド 17：プログラム評価の試み（2）1998年～2001年

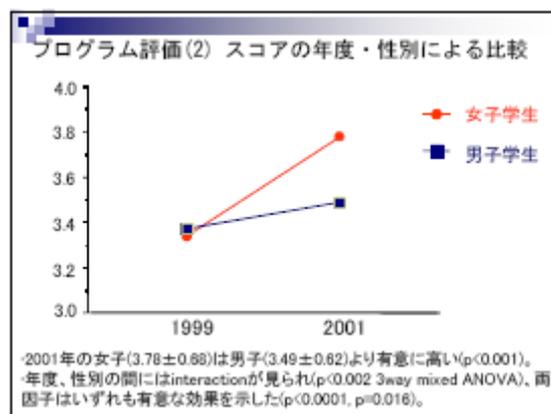
それでどうなっていったかということを経験年と2001年を比較してみますと、スコアでは劇的な改善がみられました。詳細は省きますけども、99年と2001年で全体のプログラムを変えましたし、総合外来の実習でのトレーニングもきちんと質疑応答が出来るようなプログラムを入れました。やりましたけど明らかに改善が見られたということで効果がありました。

質問項目	合計 n=660	1999年 n=287	2001年 n=373
PSQ 1	3.66 (0.75)	3.49 (0.74)	3.59 (0.73)
PSQ 2	3.62 (0.77)	3.54 (0.75)	3.61 (0.75)
PSQ 3	3.65 (0.79)	3.48 (0.82)	3.69 (0.73)
PSQ 4	3.68 (0.76)	3.54 (0.74)	3.66 (0.72)
PSQ 5	3.29 (0.98)	2.92 (1.00)	3.43 (0.81)
PSQ 6	3.41 (0.92)	3.19 (0.95)	3.52 (0.79)
平均スコア	3.55 (0.69)	3.36 (0.66)	3.58 (0.65)

・ 6項目のCronbach α係数 0.89
・ 2001年の平均スコアは1999年より有意に高い(p<0.0001)。

スライド 18：プログラム評価の試み（2） スコアの年度・性別による比較

ただ、効果を見てみると男子学生より女子学生に明らかに差があるという interaction (交互作用) があるというおもしろい結果が得られました。女子は確実に上がっているのに男子のほうはあまりあがっていないということが見られたのです。



スライド 19：プログラム評価の試み（2） スコアの改善の度合い

スコアごとに Effect size, 項目ごとのスコアのこの2年間改善の度合いをみていきますと、女子は全部の項目で、かなりの改善特にこの5番に関しましては高度の改善を示している。それ以外にも中等度の改善、あるいは軽度の改善が見られている。ところが男子学生は、ほとんどの項目で改善が見られていないところがマイナスになっているところ、ただ、質問5に関しましては弱いなが

	PSQ項目	スコアの差	Effect size(d)
男子学生	PSQ 1	0.070	0.096
	PSQ 2	-0.029	-0.039
	PSQ 3	0.081	0.104
	PSQ 4	-0.003	-0.004
	PSQ 5	0.403	0.442
	PSQ 6	0.170	0.195
	平均スコア	0.115	0.175
女子学生	PSQ 1	0.214	0.288
	PSQ 2	0.258	0.339
	PSQ 3	0.428	0.583
	PSQ 4	0.359	0.490
	PSQ 5	0.759	0.854
	PSQ 6	0.611	0.712
	平均スコア	0.438	0.670

Effect sizeはCohen's delta (d)を用いた。
高度の改善 d>0.8, 中等度の改善 d0.5, 軽度の改善 d0.2

らも改善が見られているということで、これだけ差がでてきているということ。改善はしているんだけど、やはりこれだけ差が出ている。ではこれは何故なのかというようなことを考えていくようなきっかけになりました。非常に有用なデータが得られたと思います。詳細は省きます。

スライド 20 : PSQ の妥当性検証のために

ここでもう一つ付け加えたいことは、そういう調査を進めていって実際にプログラム評価教育の改善に非常に効果があるし、ペーパーにもなるということで自信も出てきたのですが、一つひっかかってきたことは、患者にアンケート取ったものでは学位はあげられない。アンケートをばっとやっただけで、そんな学位が取れるならそんな気楽なことではない。そのような噂が伝わってきました。

PSQの妥当性検証のために

- OSCEのスコアとの相関無し (×)
 - OSCEのステーションが少なすぎて一般化できず
 - 4年次までの臨床技能プログラムの未熟のため、マニュアル的演技能力評価
 - OSCEスコアが劣悪な学生は、PSQも低い傾向
- 指導医による外来診療内容の評価 (○)
 - 主として診断を目的とした情報収集能力を、症例プレゼンテーションを通して評価⇒良好な相関
 - 適切な診断仮説に基づいた論理的な問診は、患者の満足度にも好影響を与える
 - 欧米の報告とは異なる所見

当時僕はこれで学位をとろうとしていたので、こりゃいかんというふうに思いました。だからこれだけ単独研究でずっとやってもこりゃいかんだろうといかに患者さんの評価が「あて」になるものなんだということを、示さないといけない。ということもありましたし、そこで妥当性を検証していかなければいけないと思いました。色々悪戦苦闘をしました。臨床の評価つまり OSCE と相関がないかどうかやりました。結論で言いますと OSCE とは全然相関はありませんでした。OSCE が共用試験あるいは共用試験が始まる前佐賀大学では OSCE やっていた OSCE の結果とは全然相関が全くありませんでした。なぜかと申しますと、OSCE 自体にステーションが少なすぎて一般化出来ない、あるいは 4 年次までの臨床技能のプログラムが非常に未熟なので、どちらかという演技能力になっている。そういった面もありました。ただ、OSCE スコアが劣悪な人は PSQ も低い傾向にある。やはり OSCE というのは、足りには良いですが、ここでダメな学生は、後からダメダメな学生ということが分かるくらいで、これ以外は全然相関はないということでした。

では指導医がみたらどうか、指導医が直接診療内容を見る。医学生が患者さんを診察する、その行為に対して患者さんが PSQ で評価する。一方、指導医はその医学生の報告つまりプレゼンテーションを聞いてそれに優劣をつける。これは見事な相関がありました。つまりこれは欧米の報告とはだいぶ違うんですね。患者の視点と指導医の視点というのは相当異なっている、からこそ患者の視点を入れなくてはいけない。細かいデータは出しませんがそれでも非常に相関が高かったですね。

スライド 21 : 総合外来実習評価表

だから、指導医はきちんと診断仮説を挙げてちゃんとした情報収集をしているかという観点で訊きました。それが評価表です。4 段階 ABCD に分けたのですが、一番だめなのは患者とのコミュニケーションが成立していないということ。その次は患者の訴えは一応整

理はできているけれども積極的な情報収集はできていない。もう一つは、鑑別のための情報つまり診断仮説を挙げて積極的な情報収集はしているけれども、不十分である。Aは初期情報の評価は得られている、ということで診断のための情報収集能力に視点を置いたプレゼンテーション評価、比べてみると非常に相関するということが分かった。患者さんの評価は当てになるんだということが分かりました。

スライド 22 : 今後の展望

ただこの評価表は自分が勝手に作ったオリジナルなんですね、ですから、これではちょっと弱いということで例えば臨床評価能力として確立したもの、RIMEとかMini-CEXとか色々ありますけど、それでやってみようやってみて、失敗したり悩んだりというのが今の状況です。総括的評価として使用するためには、PSQ単独では困難であるということで、RIMEとかMini-CEX

とかそういう臨床能力評価と合わせて360度評価表の開発ということは今試みています。もう一つは、ちょっとこれはそうではないかという仮説なのですが、直接に患者の診察風景を見て指導医が評価するよりもむしろ、医学生がプレゼンテーションをしているのを見てつけた評価の方がきちんと評価が出来るというか、患者さんの評価とも反映するし、他の評価表とも相関するなという印象、手ごたえを持っています。現場では色々行ったりもしますが、かえってプレゼンの方が、能力は分かるな、ということが最近感じることで、そういった方向でこの研究を進めていこうかということも考えています。

以上、質問紙を用いた調査をやってきました、簡単ではないということを思います。それから質問紙のできの良し悪しもありますけれど、それ以外の要素がたくさんあって、しかもやはり調査をするというのも一つの技能と申しますか、良し悪しと言うか、やり方のうまい下手があつてやはりきちんと調査しないとろくな結果は出ない、おかしい結果が出てしまうということを注意しなければいけない。もう一つは、森本先生も言われていましたけど、研究者はそのフィールドで患者さんがアンケートを書いているところを見たいですか、勿論データを見てその傾向を把握することこれは絶対だと思います。結果論、データの平均値とか標準偏差だけ見て云々と言うのは問題だと思いますけれども、データも見ないといけない。でもフィールドに出てやはり見ないとここはおかしい、なんでこの

総合外来実習評価表

- 評価基準
 - A) 初期評価が可能なだけの情報を得ている
 - B) 鑑別のための情報を積極的に求めているが不十分
 - C) 患者の訴えを整理したのみで積極的な情報収集はできていない
 - D) 患者とのコミュニケーションがうまくとれていない
- 評価方法
 - 学生は一例毎に指導医の評価をもらう
 - 2週間で10例以上がノルマ
 - A+, A, A-/ B+, B, B-/ C+, C, C-/Dの10段階

今後の展望

- 確立した評価表による検証(進行中)
 - RIME, Mini-CEX: 診療の直接観察による指導医からの評価
 - Mini-PAT: 振り返りによる同僚評価
- 総括的評価に使用するために
 - PSQ単独では困難
 - 上記指標と併用した360度評価表の開発
- 新たに着目している指標
 - 症例プレゼンテーション能力の評価

質問票に 15 秒くらいで答えられるのか、15 秒くらいで答える。こういうことは自分でフィールドに出ないとわかりません。あえてこういう患者さんは除外すべきではないか。例えば健診の精査で来た人、もう情報収集もないわけですね。だからこういう人の満足度というのは入れるのは問題がある。こういうのは除外すべきだということも見えてくる。そういった意味で研究者がフィールドに出てみないと、なんともいえないというか、おかしなデータ、たまたまそれが都合の良いデータが出てしまうとそのまま結論に行ってしまうそういう仕事になるので、研究者はよくフィールドを知らないといけないということを強く思います。以上、10 年こんなことをやってきているのか、途中、さぼったりも挫折もしましたけれども、今度こそうまくいくように頑張ろうと思っております。皆様のご参考になればと思ひましてお話ししました。ありがとうございました。

質疑応答

Q：学生用 PSQ は、患者さん自体には、何か感想伺ったりしたんですか・・・質問紙の中身に関して・・・

A：そういうことをした年もありました。答えにくいとかいう質問も頂いたので少し簡略化もしました。

Q：結構長い文章の内容なので、例えば、「呼びかけ方は適切だったけど、親しみを感じなかった」というのはどう扱っているのですか。特に確立されたものであればそれでいいのかな？と思いますし、自分で作る時に迷ってしまうんですが。

A：挨拶に関しての項目においても似たような内容が並んでいるのが PSQ の特徴といえます。例えば挨拶に関して、大体そういうニュアンスに受け取ってもらえるんじゃないかということで、そのまま使っています。5 年生に使っているのは簡略化しているんですね。文章も長いですし、もう一つはアンケートですねフォントが、字体が 12 ポイント以下だと年寄りにはもう見てくれません。ですから 1 枚に入る分量が非常に限られます。これは教授会とかでもそうで、12 ポイント以上で作るとというのが教授会に出すときのポイント。

Q：誤差を減らすために先生がおっしゃった秘書さんを雇うという方法、自分が出て行って誤差を少なくする方法が挙げられていましたが、他に何かファクターはありますか？

A：僕が自信を持ってできていることというところですね。後は主旨を説明するためのイントロを質問項目につけることですね。何の評価なのか、このためにこういう調査をしていますということ、3,4 行入れてですね、ということじゃないかと思います。

Q：先程、森本先生のお話にもあったことで、ちょっと混乱しているんですけど、配布する係りの人が変わったら、データが変わったということになって、そのあたりのところと、横に近くに立っているとある意味、圧力といいますかね、圧力をかけてはいけないうけど、圧力をかけるからしてくれるという部分もあるので、そのあたりはどのように判断したらいいのでしょうか。それと知らないような細かいところはかけてもいい

のかな横に立って書いているということは・・・あなた真面目に書きなさいよという圧力もあるような部分もあるような気もするし、でも適当に書かれると分からないから困るといふ、どういうふうに整理して考えれば・・・？

A: 配布者によって変わるということについて言葉が足りなかったんですけど、配布者はこれまでに何人も代わっています。僕の秘書さんはバーンアウトして1年で辞めるんです、1年でやめる、毎年変わるんです。ただ、このPSQをずっとやっていたので、ノウハウの継承はずっとなされていきました。その間はきちんと安定したデータがとれるようになっていた。横に立つのはやはりだめですね。配った後は端の方で離れたところで視界に入るのもだめです。診療した本人が配るのもだめですし、横でやはり見ているというのは相当のプレッシャーです。ただ、中には読めないとかいうこととか手が震えて書けないという人とかの場合は、もし、助けが必要でしたらということ尋ねることもあります。そういう場合はこういう補助をしたということを書いてもらわないといけません。やはりよく書けないおばあちゃんが、秘書さんが手伝って書いたら多少(回答が)良くなるというのは結構あると思います。

Q: 例えばウチの大学で研修医が付き添いとかやった場合、最後封筒に入れて回収する方法をとったことがあるんですけど、封筒に入れただけでは、実は順番においていけばやはり書いた人を同定することはできるわけで、・・・どこまでしていいのか・・・

A: まず無記名であること。診療内容には一切関係しない。担当医にはこの結果は伝わらない、あなたがここですごく低い評価を学生につけたことは学生にも伝わらないし、指導医のほうにも伝わらないから、今後のあなたの診療に関してなんのデメリットも受けない、というふうなことはやはり最初にきちっと書いておくようにします。

タスクフォース：森本先生

ウチの外来でそんな患者満足度というか医療安全の調査をするんですが、僕らがデパートに行った店員の対応を考えると、店員にすぐついて欲しい場合と、放っておいて欲しい場合があるといくつかのパターンがあるので、僕がやっているのはいくつかのパターンを用意しておいて、目とか手とか質問が読みにくいとか困る場合は横に立っておいて、ついてあげるときはついてあげるし、一人で放っておいて欲しいといわれれば・・・、ちょっとここでゆっくり考えたいという場合には返信封筒をお渡しして、持って帰って返してもらう、というように1パターンじゃなくて複数パターンするほうが患者さんもハッピーだし、回収率上がると思います。

Q: 気になるのが佐賀の診療、大学で診てもらいたいという患者さんの背景なんですけども、佐賀大学では学生さんが大学以外の診療所で、患者さんを診ることがありますか、そこで回収させる場合と外来でさせる場合と差があるんじゃないですか？

A: そういう診療所で調査をした場合、おそらく相当違う結果が出るのではないかということとは推測していますし、その調査を別な施設ではじめたこともあります。中規模、診療所です。結局これはですね、ある事情があつて頓挫しまして、出来ていません。で

すから単一施設での経験でしかないということは非常に弱いところです。そのあといくつかの大学でこれを使ってやりたいというお話があつて、何箇所かにお送りしたのですが、何らかの結果として報告がきたことがないので分かりません。

Q: 患者さんの心理を質問紙で測る場合、患者さんの心理をよく考えないとどうしても患者さんは大学病院で診てもらいたいと思っていて、いやなことは出たくない。このアンケートが診療に何らかの影響があるのではないかと考えてしまうと思いますが…

A: デメリットはないということを繰り返し、繰り返し強調していくこと。やはり求めているものによって、満足度は違うということがあります。全くおっしゃるとおりです。

Q: 日本語版の PSQ の 10 項目はどこかにアプリケーションされたり載っていたりしますか?

A: ペーパーになった分には原文があります。

Q: 先生の出された 5 項目、6 項目もありますが、10 項目の正式な日本語版というのはまだ存在していないのですか

A: 存在しています。それも承諾を得たものはあります。

Q: どこかにパブリケーションされていますか? 参照することは?

A: どこかに出されたような気がしています。たぶん。必要であれば送ります。

Q: 今回は学生さん対象でしたけど、研修医、指導医と比べてどうかというのは

A: 今やっています。研修医と患者さんだけです。指導医に対する患者満足度を見るのは微妙ですよ。全部に配って、教授だろうが准教授であろうが…

タスクフォース: 杉本先生

今日ご説明頂いたのは、**back forward translation** というタイプの **back translation** です。もう一つ **forward translation** というやり方があるんですけど、投稿なさるとか・読者から **back translation** したんだったら、**back translation** で・・・を報告しなさいということがあります。それは一致率なので、元のオリジナルなものと翻訳された英語の・・・その一致率を求められて、その一致率は 6 割が満足といわれているんですけど、8 割ぐらいがレポートされていて、査読していて 9 割以上越えていると、逆にまずいなと思うんですけども、というのは **back forward translation** と **back translation** は一長一短あつて **back translation**・・・を上げようと思つたら、直訳になればなるほど一致率があがるので、英語のものだけの前後の比較を日本語のクオリティが低くても一致率が上がっちゃうんです。非常に変な日本語のものが高いので一致率 9 割を超えた場合では特に日本語のものを見せてくださいとお願いするんですけども、なので私個人は、**forward translation** を使うんですけども、**forward translation**・・・を客観的に担保できるというか、報告できる数値のようなものがないので、**forward translation** しましたというと必ず **back translation** をかけてからもう一回・・・を報告しなさいといわれることがあるので、・・・を最初に上げたければ文レベルの一致率ではなくて、文節レベルの起こせばいいんですけども、文一本全部ロスになっちゃうので、本来違うところがあつたり、単位を細かくすれば、さすがに単語レベ

ルにするとおかしくなってしまうので、・・・ forward translation というのはまず誰かのものを翻訳して、・・・フォーカスグループ・・・2人が・・・例えば複数のものだったら並行して、不一致点をフォーカスグループで話し合っ解決していく、一番いい訳はどれかという話しで戻さないんですね。それで終わりというやり方。それは話し合いの過程を記録するしかなないので、数値の報告が出来ないのでそれは査読の時には好まれないのが特徴です。

Q：結果を見たいという学生さんにどのように対応されますか。

A：出てきた PSQ をいちいち学生に「これが今日の君の結果評価だよ」と見せることはしていませんが、学生はそれを見たがります。だけど見せませんでした。というのはそれに即した行動を採ってもらおうとやはりマニュアル的になってしまいますので、実習の最後に、実習の中間に見せずに簡単なコメントをしました。一番最後に 2 週間通してこんな感じだったよと、コメントしました。細かい質問の内容とか言い回しとかは伝わらないようにしました。これは見せた方がいいのか悩んだのですけども、研究としてやっているということで見せませんでした。本当は見せたほうがいいのかなどというものがあります。こちらがマニュアル的にやっても、相手の反応しだいですので、何とか見せずにやっています。