

# 質問票開発のための統計学1

国立がん研究センター 研究支援センター 生物統計部  
野村 尚吾

# OUTLINE

- 質問票の基礎知識
- 質問票開発の流れ（フェーズ）

# 導入：質問票開発について

- 質問票の目的：**測定対象を高い妥当性・信頼性で測定すること**
  - **測定対象には2パターンある**
    - **正解がある場合**
    - **正解がない場合**
- 質問票開発のメリット
  - 正解がある場合（例：年間の平均食事摂取量）
    - 年間の平均食事摂取量を測定するのは大変なので、より簡便な方法（質問票）を代わりに用いる
  - 正解がない場合（例：患者QOLの測定）
    - QOLを明示的に定義するのは難しいので、「おおよそQOLを反映するであろう」質問票を用いる

# 質問票に求められる性質

## ■ 妥当性 (validity)

- 本当に測定したいものを測定しているか？
- ただし、ほぼすべての状況で絶対的な真値は存在しない

## ■ 信頼性 (reliability)

- 測定値のバラツキの評価
- 同じ結果が再現されるかどうか？

## ■ 感度 (sensitivity)と反応性 (responsiveness)

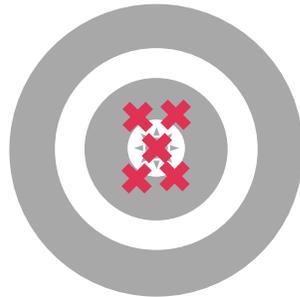
- 感度：集団間での違いを検出できるか？  
(例：一般集団とがん患者でQOLスコアの分布が分かれるか？)
- 反応性：ある患者の病態推移が捉えられるか？  
(例：病状の悪化に伴ってQOLスコアが経時的に変化するか？)

# 妥当性と信頼性の違い

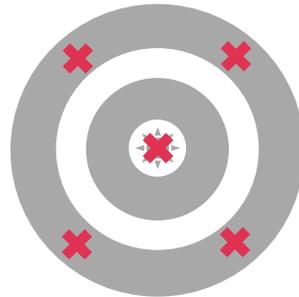
= 「正確性(accuracy)」と「精密性(precision)」の違い



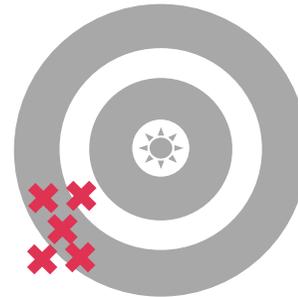
☀: 真の値



正確性 ○  
精密性 ○



正確性 ○  
精密性 ×



正確性 ×  
精密性 ○



- 妥当性評価：正確性に関心がある（興味のあるものを反映しているか）
- 信頼性評価：精密性に関心がある（同じ結果が再現されるか）

# 妥当性を評価する方法

## ■ 正解がある場合

### – 基準関連妥当性 (criterion validity) の評価

- 既に確立された質問票や実際に観察される症状等との一致度を評価する
- よく用いられる統計手法
  - 相関係数  
(一致の強弱を評価できない点に注意が必要)
  - 診断の場面で用いられる手法 (感度・特異度など)

## ■ 正解がない場合

### – 構成概念妥当性 (construct validity) の評価

- よく用いられる統計手法
  - 相関係数、Cronbachの $\alpha$ 係数、因子分析

# 補足：妥当性の種類

※ 内容的妥当性は論文検索や専門家へのインタビューで検討するため割愛します

## ■ 内容的妥当性 (content validity)

- 項目が興味のあるドメインをどの程度反映しているか？

※ 正解がある場合の妥当性評価で重点的に評価すべき事項

## ■ 基準関連妥当性 (criterion validity)

- 既に確立された尺度との関連があるか？
  - 既存尺度を改良する場合は関連の強さはあまり重要でない

※ 正解がない場合の妥当性評価で重点的に評価すべき事項

## ■ 構成概念妥当性 (construct validity)

- 意図された通りに項目・ドメインの関係が成り立っているか？
- 収束的妥当性 (convergent validity)
  - 似た項目は高い相関を示すはず
- 弁別的妥当性 (discriminant validity)
  - 似ていない項目は低い相関を示すはず

# 信頼性を評価する方法

- 正解の有無によって評価方法は大きく変わらない
- 評価者内信頼性（再テスト信頼性）
  - 同じ対象者を経時的に評価して1回目・2回目の結果を比較
- 評価者間信頼性
  - 同じ対象者を異なる評価者が評価して結果を比較
- よく用いられる統計手法
  - スコアが連続変数の場合： 級内相関係数（ICC）
  - スコアがカテゴリー変数の場合： カッパ係数

# 正解がある質問票：妥当性・信頼性の評価

- イソフラボン摂取量の質問票 (Food-frequency questionnaires; FFQ)
  - 大豆のイソフラボン摂取量の年平均を知るのが目的
    - 大豆イソフラボン（ゲニステインとダイゼイン）の測定に着目
  - 食事記録やバイオマーカーからイソフラボン含有量を計算することが可能 (Kimira, et al. J Epidemiol 1998, Arai, et al. J Nutr 2000)
  - しかし、食事記録を長期に渡って作成するのは非現実的なので簡便なFFQがあると便利
  
- 質問票の正解
  - 正解1：4つの時季(春夏秋冬)で調査した28日間の食事記録
  - 正解2：血液中・尿中のゲニステイン・ダイゼイン含有量
  
- 妥当性と信頼性をどうやって担保したのか？

## ■ 妥当性の評価法

– 以下の2つの相関関係を調べること

- FFQと食事記録から計算したイソフラボン摂取量の相関関係
- FFQから計算したイソフラボン摂取量と血中/尿中のイソフラボン含有量の相関関係

## ■ 妥当性評価の結果 (Spearmanの相関係数)

	n	Daidzein		Genistein	
		Unadjusted	Adjusted <sup>2</sup>	Unadjusted	Adjusted <sup>2</sup>
FFQ vs. DR	215	0.57 (0.47–0.67)	0.60 (0.50–0.69)	0.54 (0.44–0.64)	0.59 (0.50–0.68)
FFQ vs. Serum	202	0.31 (0.18–0.44)	0.26 (0.13–0.39)	0.33 (0.21–0.46)	0.22 (0.08–0.35)
FFQ vs. Urine	93	0.29 (0.09–0.49)	0.40 (0.21–0.60)	0.20 (–0.01–0.41)	0.30 (0.08–0.51)
DR vs. Serum	202	0.39 (0.27–0.51)	0.37 (0.25–0.49)	0.42 (0.30–0.53)	0.33 (0.21–0.45)
DR vs. Urine	93	0.43 (0.25–0.60)	0.48 (0.31–0.64)	0.38 (0.21–0.56)	0.36 (0.18–0.55)
Serum vs. Urine	90	0.23 (0.03–0.43)	0.22 (0.01–0.42)	0.34 (0.15–0.53)	0.32 (0.12–0.52)

<sup>1</sup> 95% confidence intervals of Spearman's correlation coefficients.

<sup>2</sup> Adjusted correlation coefficient by total energy for FFQ and DR and creatinine for urine.

正の相関を認めた

相関関係だけでは正解に一致したかは分からないので注意  
(詳細は「質問票開発のための統計学2」で説明)

## ■ 信頼性の評価法

- 再テスト信頼性で結果の再現性を評価
  - 1年間隔を空けて同じ対象者にFFQ調査を実施
  - 1回目と2回目のイソフラボン摂取量が相関するかを調べる

## ■ 信頼性評価の結果 (Spearmanの相関係数)

- 0.75 [95%信頼区間, 0.68-0.82]
- 十分高い再現性（信頼性）が確認された

## ■ 実際のFFQの例

国立がん研究センター 社会と健康研究センターのホームページより

[http://epi.ncc.go.jp/files/00\\_common/questionnaire/japanese/JPHC\\_Q10Cohort\\_1.pdf](http://epi.ncc.go.jp/files/00_common/questionnaire/japanese/JPHC_Q10Cohort_1.pdf)

「みそ汁」についておたずねします。

どのくらいの頻度で飲みますか？

ほとんど飲まない  
  月に1～3日  
  週に1～2日  
  週に3～4日  
  週に5～6日  
  毎日飲む

朝・昼・夕食あわせて、1日におよそ何杯飲みますか？

1杯未満  
  1杯  
  2杯  
  3杯  
  4杯  
 5杯  
  6杯  
  7～9杯  
  10杯以上

どのような味付けですか？

かなりうすめ  
  ややうすめ  
  ふつう  
  ややこいめ  
  かなりこいめ

# 正解がない質問票

## ■ HRQOL (Health related quality of life)の質問票

- 端的に言えば、健康状態について尋ねる質問票
- その他の正解がない質問票
  - 高齢者機能評価 (Geriatric assessment)
  - 満足度調査など

参考 : JCOG高齢者研究ポリシー

[http://www.jcog.jp/basic/policy/A\\_020\\_0010\\_39.pdf](http://www.jcog.jp/basic/policy/A_020_0010_39.pdf)

## ■ 押さえておくべき2つのポイント

- 「**複数の要素/領域**」からなる「**多次元**」的なもの
- 「**主観性**」 = 誰が評価するかによって結果が異なる

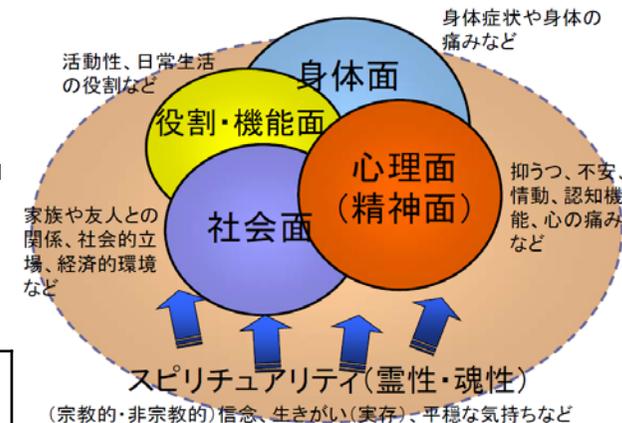
## ■ 妥当性・信頼性評価の事例は第2回目の講義で紹介します

- その前に、HRQOLについて解説します

# HRQOLの多次元性

## ■ HRQOLの4要素 (ドメイン\*)

身体面	身体症状、身体の痛み、性機能など
心理 (精神) 面	不安、抑うつ、認知機能など
役割・機能面	活動性、日常生活など
社会面	家族・友人関係、社会的立場など



\* ドメインを構成概念、因子、潜在 (顕在) 変数と呼ぶことがあります。一般にドメインの数をQOLの「次元」と呼びます。

## ■ その他のドメイン: スピリチュアリティ

- 心の支えや信念、生きがい
- HRQOLのドメインに含めないのが一般的
  - 分化や民族によって考え方の差が大きい
  - ただし、がん医療では重要なドメインとする主張もある

# 質問票の例：QOL-ACD

## ■ がん薬物療法におけるQOL調査票

(Quality of life questionnaire for cancer patients treated with anti-cancer drugs)

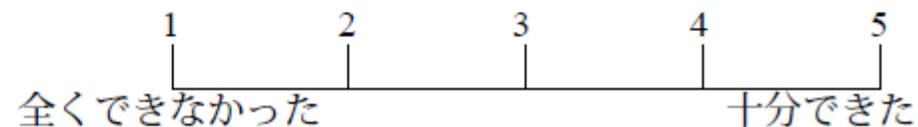
- 患者自記式
- 5つのドメインで構成される5次元の質問票（合計22項目）
  - 活動性 6問（項目番号1-6）
  - 身体状況 5問（項目番号7-11）
  - 精神・心理状態 5問（項目番号12-16）
  - 社会性 5問（項目番号17-21）
  - 全般的QOL 1問（項目番号22）
- 各質問の合計点がQOLスコア（点数が高いほど良い）
  - 単純合計で計算するスコアを「リカート(Likert)加算尺度」と呼ぶ

この調査票は、あなたの現在の状態を正しく理解するために用いるものです。  
 ここ数日間のあなたの状態にあてはまると思われる番号に○をつけてください。  
 (個人のプライバシーが外部にもれたり、治療のうえで不利益になることは決してありませんので、感じたありのままをお答え下さい。)

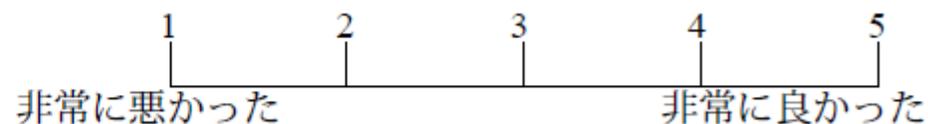
※各質問はカテゴリ尺度

(この数日の間)

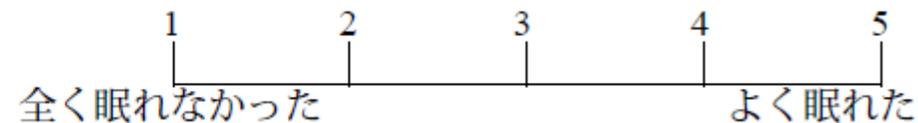
1. 日常生活(活動)ができましたか。.....



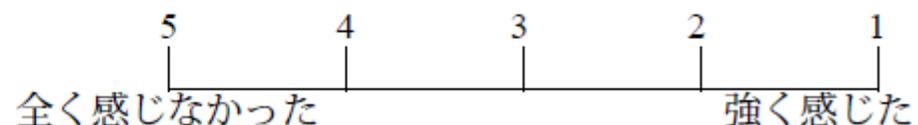
7. 体の調子はいかがでしたか。.....



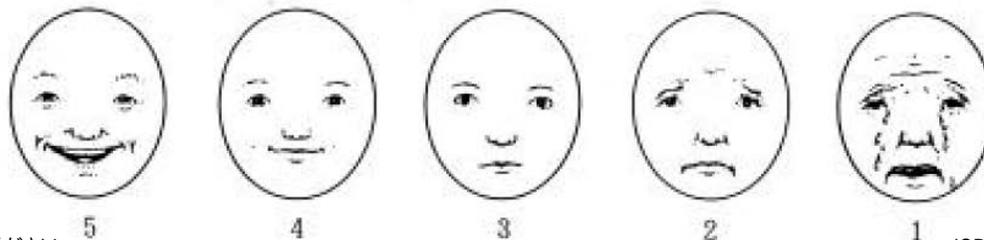
12. よく眠れましたか。.....



17. あなたの病状に不安を感じましたか。.....



22. ここ数日間の状態に相当する顔の番号に○をつけてください。

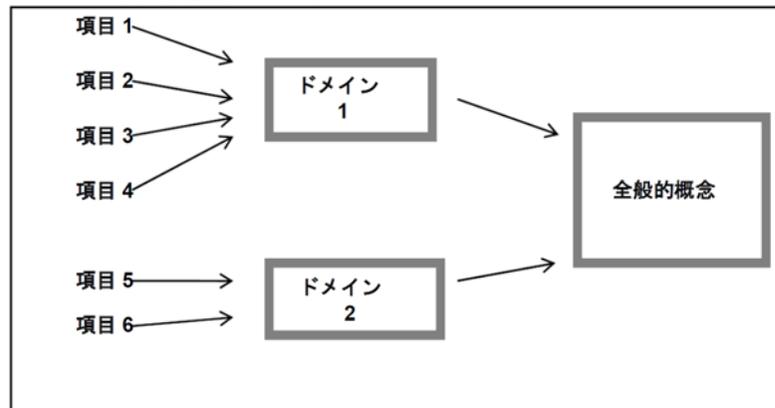


# HRQOL質問票の一般的な構造

## 1. プロファイル型 (QOL-ACDなど)

- 複数ドメインの集合が全般的概念
- 各ドメインは複数の下位尺度 (つまり個別質問) で構成される

図 4. PRO 尺度の概念枠組みを示すダイアグラム



図：FDAのPROガイダンス日本語版より引用  
<http://ispor-jp.org/pdf/kankou/kho.pdf>

- 「疾患特異的尺度」と「一般的 (包括的) 尺度」に分類される
  - 疾患特異的尺度：がん、うつ病での関心事に近づけた尺度
    - 例：EORTC-QLQ-C30, GRID版ハミルトン鬱病評価尺度 (GRID-HAMD)
  - 一般的尺度：一般人や軽症患者を対象とした尺度
    - 例：Short Form 36 (SF-36), WHO QOL

## 2. 選好に基づく尺度 (EuroQOL-5 dimensionなど)

- 効用値測定で使用 (患者自記式)
  - 効用の定義[株式会社岩波書店 広辞苑第六版] : 財・サービスが人の欲望を満たし得る能力の度合
- 個人の健康状態を、0から1の間で選択
  - 0: 想像できる最悪の健康状態あるいは死
  - 1: 想像できる最高の健康状態
- 例: EQ-5D-5L (左: 質問票、右: 効用値への換算表)

各項目において、あなたの今日の健康状態を最もよく表している四角(□)1つに✓印をつけてください

The best health you can imagine

100

95

90

85

80

75

70

65

60

55

50

45

40

35

30

25

20

15

10

5

0

The worst health you can imagine

**移動の程度**

歩き回るのが問題はない

歩き回るのが少し問題がある

歩き回るのが中程度の問題がある

歩き回るのがかなり問題がある

歩き回ることができない

**身の回りの管理**

自分で身体を洗ったり着替えをするのに問題はない

自分で身体を洗ったり着替えをするのに少し問題がある

自分で身体を洗ったり着替えをするのに中程度の問題がある

自分で身体を洗ったり着替えをするのにかなり問題がある

自分で身体を洗ったり着替えをすることができない

**ふだんの活動 (例: 仕事、勉強、家族・余暇活動)**

ふだんの活動を行うのに問題はない

ふだんの活動を行うのに少し問題がある

ふだんの活動を行うのに中程度の問題がある

ふだんの活動を行うのにかなり問題がある

ふだんの活動を行うことができない

**痛み/不快感**

痛みや不快感はない

少し痛みや不快感がある

中程度の痛みや不快感がある

かなりの痛みや不快感がある

極度の痛みや不快感がある

**不安/ふさぎ込み**

不安でもふさぎ込んでいない

少し不安あるいはふさぎ込んでいる

中程度に不安あるいはふさぎ込んでいる

かなり不安あるいはふさぎ込んでいる

極度に不安あるいはふさぎ込んでいる

No	health_state	mobility	self_care	usual_activities	pain_discomfort	anxiety_depression	qol_score
1	11111	1	1	1	1	1	1.000
2	11112	1	1	1	1	2	0.867
3	11113	1	1	1	1	3	0.829
4	11114	1	1	1	1	4	0.771
5	11115	1	1	1	1	5	0.743
6	11121	1	1	1	2	1	0.895
7	11122	1	1	1	2	2	0.823
8	11123	1	1	1	2	3	0.784
9	11124	1	1	1	2	4	0.726
10	11125	1	1	1	2	5	0.699

<http://www.niph.go.jp/journal/data/64-1/supplement.xls>

※ 換算表は国毎に用意されている

Visual analogue scale (VAS)と呼ばれる質問方式

# 質問票：臨床研究での使用目的

## ■ エンドポイント

- 費用効用分析（QALYの分析）も含む

## ■ 患者背景因子

- 適格規準（症例の絞り込み）
- 割付け調整因子
- 予後因子解析や交絡調整解析の説明変数

## ■ その他の用途

- 投与規準の変更
- スクリーニングツール
  - 回答に時間のかかる質問票の前に行う簡易質問票（例：Geriatric 8）

# 小括

## ■ 質問票を開発する上で大事なこと

- 妥当性・信頼性を担保すること

## ■ 正解がある質問票の妥当性・信頼性評価

- 正解との相関関係・一致度を調べるのが一般的
  - 連続変数なら相関係数や級内相関係数など
  - カテゴリー変数なら感度・特異度など

## ■ 正解がない質問票

- 正解が明確に定義できないので多次元の質問票で主観性が入る
- 「意図した領域が測定できているか？」が重要になる

**正解の有無によって妥当性の評価法は変わるので注意して下さい**

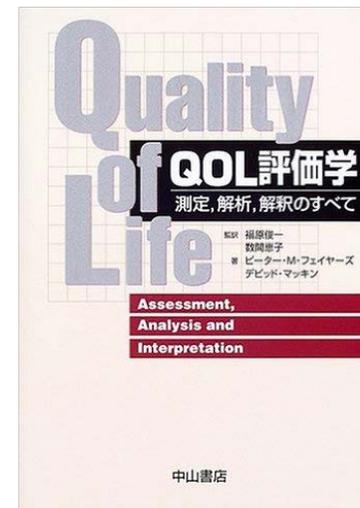
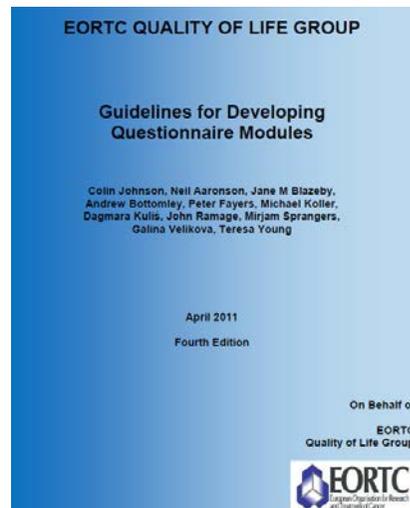
# OUTLINE

- 質問票の基礎知識
- 質問票開発の流れ（フェーズ）

# HRQOL質問票の開発

## ■ 参考文献

- EORTC Quality of Life Group (2011).  
"Guidelines for Developing Questionnaire Modules"  
[http://groups.eortc.be/qol/sites/default/files/archives/guidelines\\_for\\_developing\\_questionnaire-final.pdf](http://groups.eortc.be/qol/sites/default/files/archives/guidelines_for_developing_questionnaire-final.pdf)
- 福原(2005). "QOL 評価学~ 測定, 解析, 解釈のすべて." 7章





## ■ 開発の前に留意すべき事項

- 「包括的評価」と「疾患/治療特異的な評価」のどちらか？
- 調査する疾患の範囲は「幅広い集団」と「特定疾患」のどちらか？
- **尺度の開発・検証は事前に注意深く計画し、  
計画と結果が詳細に記載された文書を作成しなければならない**  
(※ FDAのPROガイダンス、EORTCガイドラインを参照)

## ■ HRQOL開発のプロセス (EORTCガイドライン)

1. 関心のある領域のQOLに関連するすべての事柄(issue)を網羅
2. 関連する事柄から質問項目を作成
3. 質問票暫定版でプレテスト
4. フィールド調査



# Phase 1 & Phase2

※ 使用言語に関する補足

EORTCでは英語圏・北ヨーロッパ・南ヨーロッパのそれぞれ少なくとも1カ国以上で同時にpI, pIIを実施

## ■ Phase 1: 関連するissueの作成

- 文献検索 (MEDLINE, PSYCHINFO, PROQOLIDなどのDBを活用)
- 専門家へのインタビュー
- 患者へのインタビュー
- 最後に収集したissueをドメイン別に取り捨選択

## ■ Phase 2: 質問項目の作成

- 既存の質問票(EORTC QOL Item Bankなど)を参考に質問項目を作成
- 質問構造 (カテゴリー-尺度/VAS) について検討
- スコア化の方法 (単純和/重み付き和) について検討
- 最後に質問票の専門家にレビューを依頼 (質問表現の見直しなど)

<http://www.eortc.be/ItemBank/>

# Phase 3: プレテスト

※ 使用言語に関する補足

EORTCではpI, pIIの開発国に加えて欧州圏から1カ国以上、それ以外の国から1カ国以上を加えてpIIIを実施

- QOLの状態が似ているであろう10～30名の患者集団で実施
  - 例：部位別(口腔, 咽頭, 喉頭, 副鼻腔, 唾液腺)の検討 (EORTC-QLQ-HN35)
- フィールド調査に向けた質問票の見直し
  - 表現の修正（曖昧、言語によって違う解釈になる等の問題に対処）
  - 欠測データに対する対処  
（答えなくても良い or 答えたくないと思われうる質問の特定）
- 妥当性の評価
  - 例：のどの痛み（疼痛のドメイン）が嚥下尺度と相関してしまった (EORTC-QLQ-HN35)

# Phase 4: フィールド調査

- Phase 3のあとに翻訳、翻訳の後にフィールド調査
  - EORTCのcore language:  
Danish, Dutch, French, German, Italian, Norwegian, Spanish and Swedish
  - 順翻訳と逆翻訳により原版との内容・意味の点での対応を確認
- 目標母集団の代表者を最大限含める
  - 十分に大きなサンプルサイズで調査を行う必要あり
- 目的
  - 妥当性・信頼性・感度・反応性の検討
  - データの特徴を考察
    - 欠測データの頻度や特徴について検討
    - 項目スコアの分布の確認（天井・床効果の有無をチェック）

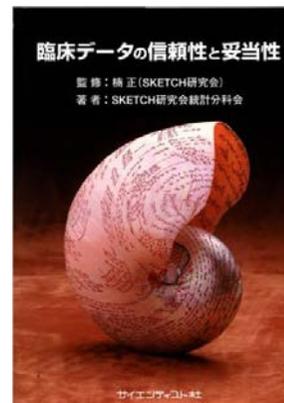
# まとめ

- 質問票の目的：測定対象を高い妥当性・信頼性で測定すること
  - **正解の有無によって妥当性・信頼性の評価方法は変わる**
- 正解のない質問票開発はかなり大変である
  - **「質問票の妥当性・信頼性の証明」は困難**
    - 「予想通りの挙動を示している証拠」、  
「予想通りの挙動を否定する根拠が何もないことについての証拠」  
をできるだけ多く集めてくることが重要
- **必要がないなら自分自身で質問票を開発しない方が良い**
  - まず初めに既存の質問票を使うことを考えるべき 福原(2005), 3.4節、7.12節
  - 新たに質問票を開発するなら、  
**「数年間にわたる非常に大変な仕事と思って準備せよ」**

※本日の講義で取りあげた主なトピックは以下の書籍でより詳しく学ぶことが出来ます



大橋靖雄. 2011  
医歯薬出版  
ASIN: B00I399SY4



SKETCH研究会統計  
分科会. 2005  
サイエンティスト社  
ISBN-13: 978-4860790226

- 正確度と精密度 p.3-8
- 信頼性と妥当性 p.24-32

- 級内相関係数 p.1-33
- カッパ係数 p.34-66

# ※生物統計セミナー【入門編】をより深く理解していただくための推薦図書



浜田知久馬. 2012  
真興交易医書出版部; 新版  
ISBN-13: 978-4880038612



佐藤俊哉. 2005  
岩波書店  
ISBN-13: 978-4000074544



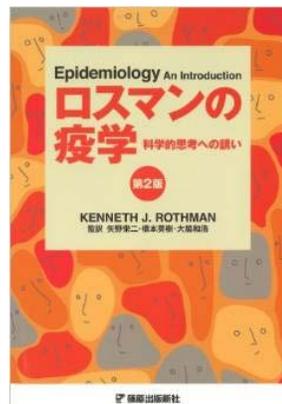
佐藤俊哉. 2012  
岩波書店  
ISBN-13: 978-4000295949



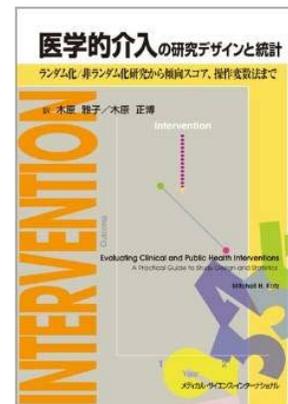
福田治彦 (訳). 2013  
医学書院; 第2版  
ISBN-13: 978-4260018647



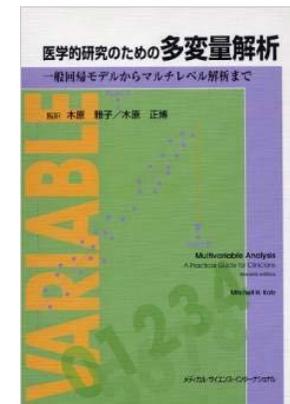
大橋靖雄. 2011  
医歯薬出版  
ASIN: B001399SY4



矢野 栄二 (訳). 2013  
篠原出版新社; 第2版  
ISBN-13: 978-4884123727



木原雅子 (訳). 2013  
メディカルサイエンスインターナショナル  
ISBN-13: 978-4895927574



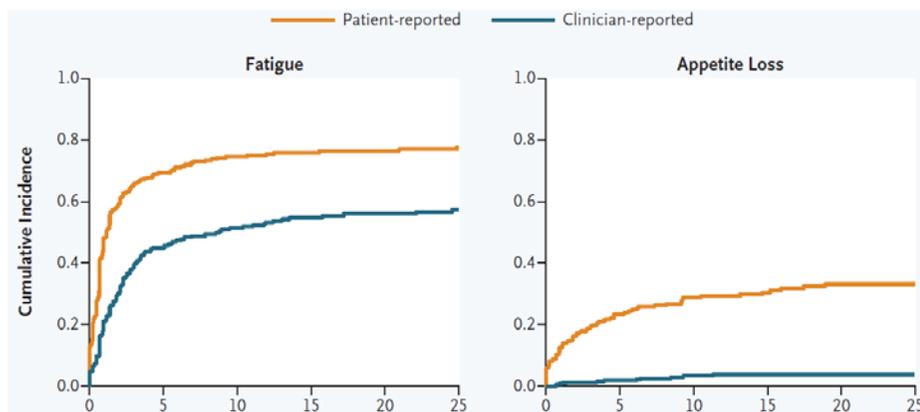
木原雅子 (訳). 2008  
メディカルサイエンスインターナショナル  
ISBN-13: 978-4895925532

# BACKUP

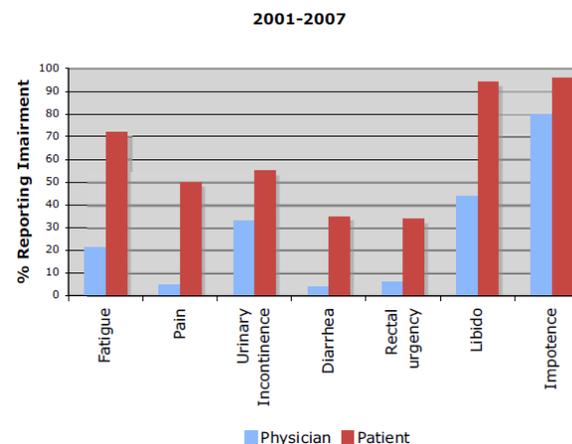


# HRQOLの主観性

- 患者の主観に基づくアウトカム (**P**atient-**r**eported **o**utcome)
  - 「臨床家その他の誰の解釈も介さず、患者から直接得られた、患者の健康状態に関するあらゆる報告」 (FDA PRO guidanceより)  
<http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf>
- 第三者評価は過小評価しがち (評価結果が食い違った事例 ↓)
  - がん患者のCTCAE評価 (左図 : Basch. N Engl J Med 2010 より抜粋)
  - 前立腺がんのSF36 (右図 : Sonn, et al. J Urol 2013 より抜粋)



縦軸：累積発生割合



縦軸：QOL悪化割合