

連載

心理社会的要因の測定(1)

「心理特性 I 信頼性」

産業医科大学産業医実務研修センター 堤 明純

1. 連載の目標

心理社会的要因が健康に及ぼす影響は、公衆衛生の重要な研究かつ実践領域となっている。介入研究も含めて心理社会的状況が健康状態に影響する知見は蓄積しており、種々の行動変容を目的とした行動科学的な手法の適用が実践の場に広がってきた。日本公衆衛生雑誌にも心理社会的要因を扱う論文は多数掲載されている。その中で、心理社会的要因は、単に変数として扱われるだけではなく、公衆衛生活動に使用する尺度の開発に関する研究も投稿されるようになってきている。

連載「心理社会的要因の測定」では、心理社会的要因を測定する道具としての尺度について、その特性と一般的な開発方法について概説することで、公衆衛生雑誌読者の研究・実践活動に活かしていただくことを目的とし、以下のような、全6回の構成での連載を予定している：

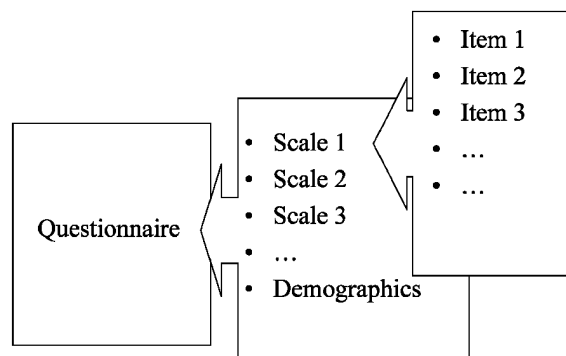
- (1) 心理特性 I 信頼性
- (2) 心理特性 II 妥当性
- (3) 尺度の開発 I 手順と項目分析
- (4) 尺度の開発 II 尺度の編集, 外国で開発された尺度の利用
- (5) 調査票の作成
- (6) 全体のまとめと補足

第1回は、心理社会的要因を測定する調査に関連する用語と心理尺度の特性のひとつである信頼性について解説する。抑えておきたいポイントは信頼性係数の意義と心理尺度が有しておきたい信頼性係数の水準、および研究を評価する際に留意しておきたい信頼性係数に関連する注意事項である。数式は難しいものではないが、読み飛ばしていただいてもかまわない。

2. 心理社会的要因の測定にかかわる用語

最初に、本連載で使用する用語について説明する(図1)。一般に調査票(Questionnaire)は、その調査で測定したいいくつかの事象や概念を測定する尺度(Scale; 心理学ではテスト Test と表現されることが多い)と、性や年齢など被調査者の属性を明ら

図1 調査票の構成



かにする質問群から構成される。尺度は、それが測定しようとする概念を測定する複数の質問群から構成され、この質問を項目(Item)と呼ぶ。しばしば、アンケート(enquête 仏語; inquiry の意)という用語が調査票と同義に使用されることがあるが、本連載では調査票という用語を使用する。

質問紙調査は、言語を媒介として、個人の意見、態度、感情、性格、行動など、人間理解のための資料を得ようとする方法である¹⁾。同様の情報を得るための方法として、他に観察法や面接法などがあるが、いずれの調査法も一長一短があり、研究の目的、規模、予算により選択される(表1)。その簡便性と一度に調査しうるデータ量の多さ(低コスト・データ収集効率の高さ)から、自記式の質問紙調査は集団を対象とする公衆衛生的研究においてもっとも汎用されている調査手段といえる。

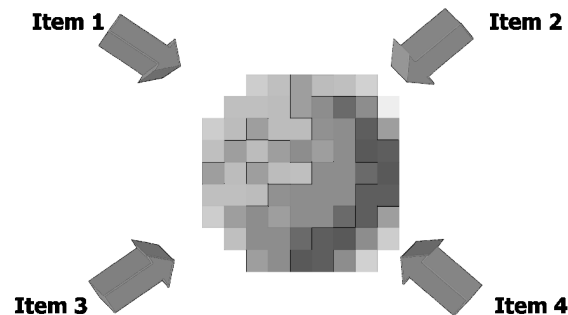
面接法では、面接者によって評価が異なるという評定者間の不一致が短所のひとつとして挙げられる。構造化面接では、評価に必要な項目が目録として事前に準備されており、各項目について該当の有無やその重症度について標準化された方法で評価できるようにルールが定められている。多くの場合、得られた情報から評価結果を出すためのアルゴリズムも提供され、評定者間の不一致が最小になるように設計されている。構造化面接は、自記式の調査だけではその正確な評価が難しい精神障害の頻度の調

表1 観察法および面接法と比較した自記式質問紙調査の特徴

vs 観察法		vs 面接法	
長所	短所	長所	短所
外部的条件と表面的行動の間にあるもの（知覚・欲求・動機・期待など）を捉えることができる 個人の歴史的側面を捉えられる 観察者の観察技量に依存しない 観察者による「観察」が被験者の行動に与える影響が大きい	主観的 被験者の言語能力に依存 行動または事態の経過の同時記録は不可能 定められた質問に対する応答以外のものを期待できない 被験者の観察の精度は他から補強できない（観察者の数・観察側面の増加・十全な結果報告を期待できない）	低コスト（同時性・多数・時間・面接者） 一様性（相対的ではあるが、統一された語法・質問の順序・応答の仕方など cf. 構造化面接） 面接技量の問題・主観性	劣る融通性・妥当性（聞きなおし・矛盾のチェック・分かりやすい問いかけ） 内密な私的感情の掘り下げに劣る 相当程度の教育歴を必要とする

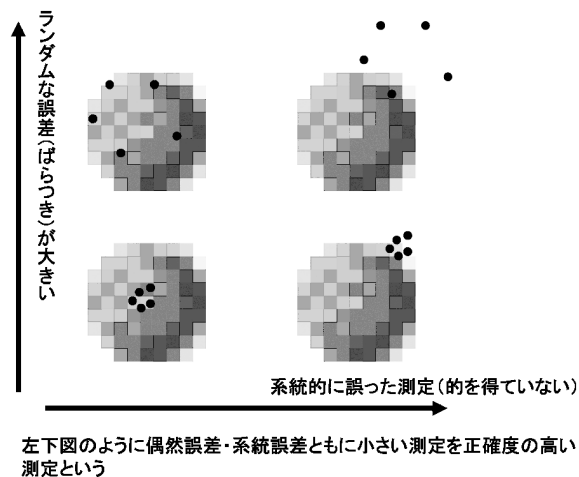
図2 心理社会的要因の測定

測定の基準（ゴールド・スタンダード）がなく、実態のある物質のように、その概念の大きさや量をピンポイントに測定することはできない



その概念を捉えようとする複数の項目（尺度）に対する反応により数値化を図る

図3 測定の正確度



査などに応用されている。

3. 心理社会的要因の測定

心理社会的要因の測定は、人間の意識や行動など、測定しようとする概念を数値に反映する作業で、尺度はそのための道具である。血圧やコレステロールなどと違い、心理社会的な概念には測定の基準となるもの（ゴールド・スタンダード）がない。目的とする概念を捉えようとする複数の項目群（尺度）に対する反応により、数値化を図る（図2）。

心理社会的要因を扱う研究に限らず、測定値は、必ずしも真の値を捉えているわけではない。測定に関連する誤差は、偶然（ランダム）に発生する誤差と系統的に偏って測定される誤差に分類される（図3）。測定誤差が大きくなると、検証しようとする事象間の関連や差を見出すことができなかつたり（主

に偶然誤差による測定の信頼区間の拡大が問題となる）、論理的に間違っただ観察を導いたりする可能性がある（系統誤差の影響は大きい）。調査研究においては両方の誤差が小さい（正確度の高い）測定を目指すことになり、読者側からみると、研究中で正確な測定がなされているか否かは、論文を批判的に吟味する際のひとつの重要なチェックポイントとなる。

4. 信頼性

信頼性 reliability（精度 precision と表現されることもある）は、測定の正確さ accuracy を表す概念のうち偶然誤差の小ささを示す。信頼性という用語は、後述する信頼性の評価の仕方（誤差の捉え方）によって多義的であるが、平たく表現すると、「常に、安定した測定がなされる」ことを示す指標である。

古典的テスト理論 (Classical Test Theory; 本連載で紹介する尺度に関する理論は、主に古典的テスト理論に基づく。最終回に近代テスト理論のひとつ、項目反応理論を紹介する) は、測定値 X は真の値 T と測定誤差 E の和で表現される(1)という前提で立論されている。

$$X = T + E \tag{1}$$

母集団からの無作為な標本抽出を前提とした測定については、測定値の平均は真値の平均に等しく(2)、偶然誤差の平均は0であり(3)、真値と偶然誤差は相関しない(4)：

$$\bar{X} = \bar{T} \tag{2}$$

$$\bar{E} = 0 \tag{3}$$

$$r_{TE} = 0 \tag{4}$$

したがって、測定値の分散は真値の分散と測定誤差の分散の和となる(なぜなら、真値と誤差に相関がない=真値と誤差の共分散が0)(5)：

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2 \tag{5}$$

測定値の信頼性が高いとは、a. σ_E (測定の標準誤差) が小さい、b. r_{XT} (測定値と真値の相関) が大きい、もしくは c. σ_T^2/σ_X^2 (測定値の分散中、真値の分散が占める割合) が大きい、と言い換えることができる。

このうち、cの「測定値の分散の中に占める真値の分散の割合」が信頼性係数と定義されている。

$$\text{信頼性係数} = \frac{\sigma_T^2}{\sigma_X^2}$$

$$\begin{aligned} &= \frac{\sigma_X^2 - \sigma_E^2}{\sigma_X^2} \tag{6} \\ &= 1 - \frac{\sigma_E^2}{\sigma_X^2} \end{aligned}$$

すなわち、信頼性係数は、尺度で測定される値の分散のうち、誤差成分の分散を除いた部分の比率をいう。定義上、係数が1に近づくほど信頼性は高く、0.8以上あることが望ましい²⁾。信頼性係数0.8とは、測定値に対する真値の分散が80%、誤差の分散が20%を占めることを意味する。

5. 信頼性係数の推定法

信頼性係数の推定法には複数の方法がある。信頼性を検討する尺度を、同一対象に繰り返し施行し、両機会で作られた尺度得点の相関を見る再テスト法、尺度の構成項目を二分し両者の関連を見る折半法、折半法の一般化といえる内部一貫法(内的整合法)などがよく利用される(表2)。

再テスト法は、測定尺度の時間的安定性(stability; もしくは再現性 repeatability, reproducibility) に注目した信頼性の検討で、前回の施行の影響を受けない間隔(条件)でテストを繰り返す方法である。しかし、被調査者の記憶や尺度を構成する項目間の相関が厳密な評価を妨げることが指摘されている(一般に、信頼性を過大評価する)³⁾。さらに、状況によって変化する概念(状態不安や新入生のソーシャルサポートなど)の安定性の評価には不向きである。

表2 種々の信頼性係数の推定法(文献4)より改変)

推定法	解 説
平行テスト法 coefficient of parallel test reliability	同一対象に、個々の設問自身は異なるが、設問のねらいや難易度、内容(種類)、量などが同等と考えられ、同一の能力や特性を測定すると考えられる2種類以上のテスト(平行テスト)を作成し、両得点の相関係数を求める
再テスト法 coefficient of test-retest reliability	同一対象に、同一テストをある期間をおいて繰り返し、両得点の間の相関係数を求める
折半法 coefficient of split-half reliability	ある与えられた尺度項目を半分にして二組に分ける：両得点を用いて Spearman-Brown の公式、Flanagan の公式などにより推定する。
内部一貫法 内的製合法 internal consistency	検査の各項目の相互相関や通過率(困難度)をもとに、Cronbach α 、Kuder-Richardson 20 (1-0 型データで適用)を用いて推定する。次式は、クロンバックの α の推定式、 m は項目数、 σ_j^2 は各項目の分散の平均である $\alpha = \frac{m}{m-1} * \left(1 - \frac{m\sigma_j^2}{\sigma_x^2} \right)$
分散分析による方法	同類の検査をいくつか用意し、実験計画法の原理で何回かにわたって実施したデータをもとに分散分析の手法を用いて推定する。 信頼性係数 = 1 - (残余の平均平方/個体間平均平方)

これに対し、テストを一回だけ施行して信頼性係数を推定する折半法や内部一貫法などは、尺度を構成する各要素得点を繰り返しの代用と見ている。高い信頼性係数を得るためには、測定値を構成する要素得点間の相関係数が高くなければならない。すなわち、構成要素がなるだけ類似した同質の課題項目から構成されることが要求され、これらの方法で検討されている信頼性は、尺度に含まれる個々の質問項目の一貫性 (consistency) あるいは等質性 (homogeneity) である。實際上、(一回の) 折半法による部分テストの作成に一貫性を保つことは困難で、すべての項目をランダムに半分ずつに分け、そこから計算された部分得点から得られる信頼性係数の期待値としてクロンバック (Cronbach) の α が算出される。クロンバックの α は、最近では、多くの統計パッケージにプログラムが組み込まれており、容易に算出が可能である。

6. 測定の信頼性および信頼性係数に関する留意点

信頼性係数はその尺度のみならず測定の対象となる集団に依存するという点に留意しなければならない。言い換えると、ある尺度の信頼性が、他の多くの研究で良好と報告されているからといって、自分が読んでいる研究論文の対象で保証されるとは限らない。読者は、研究論文内の測定の信頼性を見て、その研究における測定誤差の大きさと観察されようとしている指標間の関連の強さを評価できるのである。したがって、研究者は、自身の測定の信頼性を確認するとともに、その所見を論文内に記述しておく必要がある。

測定の信頼性が低い研究では、査読の過程で方法論上の不備が問われることがある。場合によっては、尺度の再構成と、それに基づく再解析を求められることもある。しかし、信頼性係数が十分に高くないからといって、その研究が必ずしも受け入れられないわけではなく、測定誤差による限界をわきまえた上での所見の評価と考察を行なうようにする⁵⁾。いずれにせよ、研究者は、その測定の誤差を認識していることが必要であり、読者には、測定誤

表3 心理特性 I 信頼性のまとめ

信頼性は、測定の正確さを表す概念のうち偶然誤差の小ささを示す。
 集団を対象とした測定では、信頼性係数は0.8以上あることが望ましい。
 測定の信頼性は、その測定を行う道具 (尺度) と測定対象の特性に依存する。

差に伴う研究の限界について批判的吟味を加える姿勢が求められる。

7. まとめ

心理尺度は、心理社会的要因を測定するための道具で、自記式の質問紙や構造化面接などで、人間の意識や行動などの概念を測定する際に用いられる。いずれの調査法も万能ではなく、測定誤差の存在を常に認識しておく必要がある。信頼性は、測定の正確さを表す概念のうち偶然誤差の小ささを示すもので、研究のパワーに影響する。集団を対象とした測定の信頼性係数は0.8以上あることが望ましい。測定の信頼性は、その測定を行う道具 (尺度) と測定対象の特性に依存するので、研究者は、自身のデータの測定の信頼性を確認する必要がある。一方、読者は、測定誤差の大きさを確認し、研究が導いている結論を評価する一助とする。本稿のまとめを表3に挙げる。

文 献

- 1) 続 有恒, 村上英治, 編. 質問紙調査, 心理学研究法 9. 東京: 東大出版会, 1975.
- 2) Nunnally JC, Bernstein IH. Psychometric theory 3rd Ed. New York: McGraw-Hill, 1994; 264-265.
- 3) Nunnally JC, Bernstein IH. Psychometric theory 3rd Ed. New York: McGraw-Hill, 1994; 254-255.
- 4) 池田 央. 行動科学の方法. 東京: 東大出版会, 1971; 138.
- 5) Peter R, Alfredsson L, Hammar N, et al. High effort, low reward, and cardiovascular risk factors in employed Swedish men and women: baseline results from the WOLF Study. J Epidemiol Community Health 1998; 52: 540-547.