

Evaluating Clinical Rating Scales for Evidence-Based Dermatology: Some Basic Concepts

Madhulika A. Gupta, MD, FRCPC^{a,b,*}, Andrew M. Johnson, PhD^c,
Mary-Margaret Chren, MD^{d,e}

^a*Department of Psychiatry, University of Western Ontario, London, ON, Canada*

^b*Mediprobe Research Inc, London, ON, Canada*

^c*Faculty of Health Sciences, University of Western Ontario, London, ON, Canada*

^d*University of California at San Francisco, San Francisco, CA, USA*

^e*Health Services Research Enhancement Award Program, San Francisco Veterans Affairs Medical Center, San Francisco, CA, USA*

A greater emphasis on evidence-based medicine has contributed to the increased interest in the development of standardized clinical rating scales for dermatology [1]. Greater availability of computing technology has made statistical analysis accessible to most practitioners, making it easier to implement these measurement tools. Finally, the desire to measure variables that lie outside the range of traditional biologic outcome measures (eg, the impact of skin disorders on an individual's quality of life) has led to an increased demand for psychometric measures of health.

Because quality of life is a complex and multidimensional construct, most of the measurement tools consist of multidimensional rating scales. Although this ever-increasing choice of assessment instruments makes it more likely that a tool exists for most practitioners' decision-making needs, it makes it all the more important for practitioners to have some guidelines to evaluate instruments for clinical or research uses. This article outlines some of the basic psychometric concepts that need to be considered when evaluating a clinical rating scale.

Measurement of a construct

Measurement is a fundamental activity of science. Within the behavioral sciences, psychometrics has evolved as the specialty that is concerned with the measurement of psychologic and social phenomena, and the instrument typically used is the questionnaire [2–5]. Theory plays a key role in how the variables of the measurement instrument are conceptualized. A concept is an abstraction of an empirical observation, and provides the labels used to describe the environment. Of greater relevance to clinical research, however, is the construct: a concept that is essentially unobservable without some form of systematic definition. For example, the construct quality of life, which addresses the impact of a skin disorder on the physical, social, emotional, and vocational functioning of the patient, has no physical meaning, and no clear unidimensional symptomatic presentation. Despite this fact, however, it can be described in terms of its typical manifestation, and its relationship to other (sometimes similarly unobservable) concepts. This systematic definition is called an operational definition, and it is the first step in translating a construct into a variable (ie, a measurable quantity). It is the development and testing of these definitions that forms the basis of measurement.

* Corresponding author. 645 Windermere Road, London, ON N5X 2P1, Canada.

E-mail address: magupta@uwo.ca (M.A. Gupta).

Basic measurement properties of psychometric instruments

Some of the basic indicators to be considered when evaluating the accuracy of any assay, including psychometric rating scales, are precision, reliability, validity, and standardization [2–5]. A high level of reliability is a prerequisite for validity.

Precision

The ability of an instrument to produce a particular type of measurement is of paramount importance. The precision of an instrument, its capacity to detect small differences, is constrained by its level of measurement. Four levels of measurement are generally used in describing the precision of a variable: (1) nominal, (2) ordinal, (3) interval, and (4) ratio. Categorical data may be either nominal or ordinal, depending on whether an order is imposed on the categories. Nominal data simply classifies subjects (eg, male-female is a nominal scaling classification). In ordinal data, the subjects are ranked but it is not possible to assume any consistent distance between the ranks. For example, if one were to divide the spectrum of symptom severity into three stages (mild, moderate, and severe), the distance between mild and moderate may not be equivalent to the distance between moderate and severe. The use of interval and ratio measurements, however, introduces the concept of continuous measurement, a level of precision at which one may assume that adjacent points on the measurement scale are equidistant. The primary difference between interval and ratio measurements is the presence of a meaningful zero, which indicates a complete absence of the construct. Most psychometric instruments rely on interval-level measurement (eg, when assessing the psychosocial burden of psoriasis, a patient may be asked to rate the degree to which the cosmetic impact of the psoriasis bothers them, using a five-point scale, ranging from “very little” to “all the time”). Because the lowest measure on the scale is not truly indicative of a complete absence of the construct, this is an interval measure. Clinical measures related to duration (eg, “how long have you had the disorder?”) or size (eg, “what is the size of the lesion?”) are ratio-level measures, because a zero represents a complete absence of the construct.

The precision of a variable provided in a questionnaire is an important consideration, because of the fact that it determines the types of statistics that may be done on the measure. Nominal data may only be described by frequencies, because any ordering of categories is wholly arbitrary. Ordinal data may be

described by frequencies or by a median. Both interval-level and ratio-level measurements may be described using means, medians, modes, and standard deviations, but only ratio-level measurements may be expressed multiplicatively. This means that it is not appropriate to express one interval-level measurement as a multiple of another [4]. For example, if the pretreatment score on an interval-level quality of life scale is 6 out of 10 and the posttreatment score is 9 out of 10, one cannot infer that the 3 point or 50% improvement in scores indicates a 50% improvement in the quality of life.

Reliability

Reliability refers to the extent to which a variable may be demonstrated to be measured in a reproducible fashion, or the extent to which the results measuring a stable variable may be relied on to be the same on each subsequent administration. Unfortunately, there is no gold standard for reliability that is appropriate for all testing situations; the measure of reliability on which a good instrument selection decision is based varies according to the purpose for which the instrument is intended. Reliability is generally assessed using one of four methods: (1) test-retest reliability, (2) parallel forms reliability, (3) internal consistency reliability, and (4) interrater reliability.

For the demonstration of temporal reliability (ie, stability across time, provided the subjects have not changed), one needs to look for the test-retest reliability of a measure, a statistic that assesses the extent to which two separate administrations of the same measure are significantly correlated. Test-retest reliability may be raised artifactually if the time between testing periods is too short and there is a carryover effect because the subject remembers the responses from the original testing session. For situations in which this carryover effect is likely, it is often preferable to use parallel forms reliability, where equivalent or parallel forms of items are constructed and subjects take an entirely separate test for the retest.

Another common form of reliability analysis is internal consistency, usually measured with Cronbach's alpha (sometimes called coefficient alpha). The desirability of internal consistency reliability is based on the proposition that if one item is measuring a variable, then the other items must be consistent with this measurement if they are to be considered to be measuring the same variable. Conversely, some psychometricians have argued that if all items are highly intercorrelated (ie, highly consistent), that provides no new information, arguing that the ideal

Download English Version:

<https://daneshyari.com/en/article/9225082>

Download Persian Version:

<https://daneshyari.com/article/9225082>

[Daneshyari.com](https://daneshyari.com)