Clinimetric Criteria for Patient-Reported Outcome Measures

Clinimetric Criteria for Patient-Reported Outcome Measures

Danilo Carrozzino[a], Chiara Patierno[a], Jenny Guidi[a], Carmen Berrocal[b], Jianxin Cao[c], Mary E. Charlson[d], Kaj Sparle Christensen[e], John Concato[f,g], Carlos De las Cuevas[h], Jose de Leon[i], Ajándék Előry[j], Marcelo P. Fleck[k], Toshi A. Furukawa[l], Ralph I. Horwitz[m], Andrew A. Nierenberg[n,o], Chiara Rafanelli[a], Hongxing Wang[p,q], Thomas N. Wise[r,s,t], Nicoletta Sonino[u,v], Giovanni A. Fava[v]

[a]Department of Psychology "Renzo Canestrari", University of Bologna, Bologna, Italy; [b]Department of Surgical, Medical and Molecular Pathology, and Critical Care Medicine, University of Pisa, Pisa, Italy; [c]Changzhou first people's Hospital and Psychosomatic Gastroenterology Institute, Soochow University, Changzhou, China; [d]Division of Clinical Epidemiology and Evaluative Sciences Research, Department of Medicine, Weill Cornell Medicine, New York, USA; [e]Research Unit for General Practice, Department of Public Health, Aarhus University, Aarhus, Denmark; [f]Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD, USA; [g]Department of Medicine, Yale University School of Medicine, New Haven, CT, USA; [h]Department of Internal Medicine, Dermatology and Psychiatry, Universidad de La Laguna, Canary Islands, Spain; [i]University of Kentucky Mental Health Research Center at Eastern State Hospital, Lexington, KY, USA; [j]Department of Family Medicine, Semmelweis University, Budapest, Hungary; [k]Departament of Psychiatry, Universidade Federal do Rio Grande do Sul, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil; [l]Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan; [m]Lewis Katz School of Medicine, Temple University, Philadelphia, PA, Pennsylvania, USA; [n]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA; [o]Harvard Medical School, Boston, MA, USA; [p]Division of Neuropsychiatry and Psychosomatics, Department of Neurology, Xuanwu Hospital, Capital Medical University, Beijing, China; [q]Beijing Psychosomatic Disease Consultation Center, Xuanwu Hospital, Capital Medical University, Beijing, China; [r]Department of Psychiatry, Inova Health Systems, Falls Church, VA, USA; [s]Department of Psychiatry, Johns Hopkins University School of Medicine, Baltimore, MD, USA; [t]Department of Psychiatry and Behavioral Sciences, George Washington University School of Medicine, Washington, DC, USA; [u]Department of Statistical Sciences, University of Padova, Padova, Italy; [v]Department of Psychiatry, University at Buffalo, State University of New York, Buffalo, NY, USA

Running title: Clinimetric criteria

Corresponding author:
Danilo Carrozzino, PhD
Department of Psychology "Renzo Canestrari",
University of Bologna
Viale Berti Pichat 5, 40127, Bologna (Italy)
Tel: +393343204782
E-mail address: danilo.carrozzino@unibo.it

Number of Tables: 1

Number of Figures: 0

Word count: 4.040

Keywords: Assessment, clinimetrics, criteria, indices, patient-reported outcome measures, psychometrics, rating scales, sensitivity, validity.

Abstract

Patient-Reported Outcome Measures (PROMs) are self-rated scales and indices developed to improve the detection of the patients' subjective experience. Given that a considerable number of PROMs are available, it is important to evaluate their validity and usefulness in a specific research or clinical setting. Published guidelines, based on psychometric criteria, do not fit in with the complexity of clinical challenges, because of their quest for homogeneity of components and inadequate attention to sensitivity. Psychometric theory has stifled the field and led to the routine use of scales widely accepted yet with a history of poor performance.

Clinimetrics, the science of clinical measurements, may provide a more suitable conceptual and methodological framework. The aims of this paper are to outline the major limitations of the psychometric model and to provide criteria for Clinimetric Patient-Reported Outcome Measures (CLIPROM). The characteristics related to reliability, sensitivity, validity and clinical utility of instruments are critically reviewed, with particular reference to the differences between clinimetric and psychometric approaches. Of note is the fact that PROMs, rating scales and indices developed according to psychometric criteria may display relevant clinimetric properties. The present paper underpins the importance of the clinimetric methodology in choosing the appropriate PROMs. CLIPROM criteria may also guide the development of new indices and the validation of existing PROMs to be employed in clinical settings.

Introduction

Alvan R. Feinstein introduced clinimetrics [1-3], a clinically based evaluation method that has been defined as the science of clinical measurements [4]. Feinstein described criteria to be used for developing clinimetric indices: “We must create new scales or new criteria for old scales that deal with the different roles of clinical data in identifications, classifications, and temporal distinctions. In identification, criteria of existence provide clinical rules for denoting the presence of such entities as arthritis, carditis, or rheumatic fever. In classification, we need criteria of gradation to demarcate such entities as the severity of arthritis, the degree of cardiac enlargement, or the rating of functional impairment. Each of these types of criteria can be applied in different temporal circumstances, referring to a single state, a transition or a prediction” [1, p. 2]. Such criteria were further refined in a subsequent publication [5], where the main methodological differences between the psychometric model and the clinimetric approach were outlined.

Classification systems in psychiatry, such as the Diagnostic and Statistical Manual of Mental Disorders [6], are essentially based on diagnostic criteria that reflect a clinimetric approach [1, 7]. The use of rating scales and indices is an essential integration to the diagnostic configuration [8-11]. Such integration appears to be particularly important in randomized controlled trials (RCT) and for measuring the effects of mental care [12], where there has been growing interest in the use of rating scales in the clinical process of psychological and psychiatric assessments. The importance of Patient-Reported Outcome Measures (PROMs), self-rated subjective experiences of symptom burden and psychological well-being in relation to health conditions and/or treatments, has gained increased recognition [13-27].

In the present paper, we discuss the major limitations of the psychometric model, particularly as to the assessment of validity of PROMs, and provide alternative methodological recommendations based on clinimetrics. The criteria for Clinimetric Patient-Reported Outcome Measures (CLIPROM) are presented. Such criteria apply to the development of new clinimetric

indices and scales, as well as to the assessment of properties (e.g., sensitivity) and of clinical utility of existing PROMs.

Patient-Reported Outcome Measures (PROMs)

PROMs are self-reported scales and indices specifically developed to improve the detection of the patients' subjective experience [13-27]. Originally designed to measure treatment outcomes, PROMs have a number of applications. In clinical practice they can be used as screening tools for case identification and to evaluate the severity, burden and impact of symptoms on quality of life and psychological well-being and to assess functioning in daily life, productivity, and emotional stability. In addition, PROMs can serve as monitoring instruments, particularly to detect the personal experience of care, including the subjective perception of change after treatment [14, 22, 23, 25, 28]. In this regard, PROMs have been used to enhance patient engagement (e.g., promoting a shared decision-making process) and to facilitate patient-clinician communication, clarifying the patient's subjective values and priorities for care [14, 22, 23, 25, 28, 29]. A major long-term advantage of PROMs is to facilitate shared identification of goals and priorities between physician and patient faced with complex, chronic, and multifaceted problems [29]. Two main types of PROMs can be identified: some were designed to be used for evaluating the impact or burden of a specific disease such as cancer or depression, whereas others that can be used across disorders have a more general focus on the assessment of general perceptions such as quality of life [14, 22, 23, 25, 28].

The importance of evaluating the reliability and validity of such instruments for inclusion in trial protocols has been highlighted [30-38]. Criteria on the basis of the psychometric model have been developed [13, 15, 18, 30, 39-44]. The "Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT-PRO)" statement provided an evidence-based list of PROMs recommended for inclusion in trial protocols [32, 34].

Psychometric Guidelines

Focusing on traditional psychometric criteria, Bombardier and Tugwell [45] were among the first authors to indicate a checklist for determining the reliability and validity of PROMs. Since then, various publications have suggested standards for the development of new PROMs and evaluation of their consistency and validity [39, 40, 42, 46, 47]. One of the most comprehensive checklists was developed in 1994. The Medical Outcomes Trust, a complex mix consisting of nonprofit organizations, academic researchers, public sector agencies, and commercial firms, established a Scientific Advisory Committee, with the mission to define a set of criteria to be used for the assessment of reliability and validity of outcome measures covering the domains of health status and quality of life [39]. The resulting criteria, which represent general guidelines, were set according to psychometric methods and covered the evaluation of the following domains: conceptual model, reliability, validity, responsiveness, interpretability, respondent and administrative burden, alternative forms, cultural and language adaptations [39]. In 2004, under the aegis of the Food and Drug Administration (FDA), the US National Institutes of Health (NIH) launched the Patient-Reported Outcomes Measurement Information System or PROMIS initiative (www.nihpromis.org), a 5-year cooperative group program of research designed to develop, validate, and standardize PROMs [13, 15, 18, 30]. In the position paper produced for the launch of the PROMIS initiative, Reeve et al. [31, p. S22] declared their mission as providing "[…] a unique opportunity to use advanced psychometric methods to construct, analyze and refine item banks, from which improved patient-reported outcome (PRO) instruments can be developed". Another similar initiative, also strongly rooted in the psychometric tradition, was promoted by a panel of international experts (psychologists, epidemiologists, statisticians, and clinicians) who published the COnsensus-based Standards for the selection of health Measurement INstruments, the COSMIN criteria [41, 43, 48, 49]. They provided a checklist containing standards for evaluating the methodological quality of studies on measurement properties of PROMs [41, 43, 48, 49].

A number of other consensus papers were published, initiatives launched, international conferences organized, working groups established, methodological recommendations provided, and guidelines introduced, in which standardized criteria for the evaluation of measurement properties of PROMs according to psychometric principles were suggested [13, 30, 32, 34, 36, 44, 50-52].

Clinical Inadequacy of the Psychometric Model and Development of the Clinimetric Approach

The development of psychometrics has taken place outside of clinical fields, mainly in the educational and social areas [5, 53]. It took a long time before the issues of reliability and validity were approached in their clinical meanings with the introduction of clinimetrics [1-3].

The psychometric strategy consists of finding and combining multiple items that are relatively homogeneous, as they are all intended to measure the same single attribute [5]. The assumption of homogeneity of components, which represents the methodological framework of the psychometric model, refers to the degree of positive correlations between items of the rating scale under examination [4, 5, 9, 54, 55]. As Feinstein noted: "The customary psychometric goal is to achieve a unidimensional construct, in which the relatively homogeneous components all measure essentially the same idea. Components that seem substantially different are eliminated if they do not correlate well with each other" [56, p. 126]. However, as it has been widely demonstrated [4, 5], the same statistical analyses that give a scale a high score for homogeneity of components may obscure its clinimetric properties, particularly sensitivity to change.

In the clinimetric approach, neither homogeneity of components nor unidimensionality are required, and what matters is the clinical utility of the rating scale in the process of assessing a wide range of clinical issues [3, 5, 53]. Such issues encompass: patterns or types, severity, duration and sequence of symptoms; rate of progression and staging of illness; effects of comorbid conditions, including the evaluation of iatrogenic comorbidity; aspects of illness behavior, functional capacity,

remission, recovery and lifestyle; response to previous treatments; mental pain, psychological well-being, and euthymia; and many other clinical features that demarcate major prognostic and therapeutic differences in patients who appear to be deceptively similar since they share the same diagnosis [3, 54, 57-61]. Clinimetric indices that could function as PROMs have been developed [28, 59, 62, 63] and the clinimetric properties of previously developed scales have been evaluated [64-74]. However, the clinimetric perspective has not been considered by consensus reports on standardized criteria for the assessment of PROMs [13, 30, 31, 34, 36, 44]. It has been argued that exclusive reliance on psychometric criteria has impoverished the clinical process of psychological assessment [7] and represents an obstacle to the progress of clinical research [53]. There is, therefore, a need for a consensus to define criteria for evaluating the properties of PROMs according to clinimetric principles.

Methodological Differences between Psychometrics and Clinimetrics

Reliability

The term "reliability" was originally introduced by Spearman [75, 76] and refers to the assessment of accuracy and stability of test results [77]. Clinimetric criteria for evaluating the reliability of PROMs differ from those conventionally used in the psychometric model [78].

Internal Reliability or Consistency

A crucial psychometric characteristic, perhaps the most critical criterion, is internal reliability or consistency, that is a measure of the extent to which items in a rating scale are correlated [42]. All items in a psychometric rating scale are weighed the same [9, 42]. As recently stated in a publication on classical psychometrics: "all items on an instrument are equally relevant in evaluating a person's true score" [79, p. 9]. Such a psychometric assumption, however, clashes with the clinical reality: certain symptoms may be more troublesome or incapacitating than others; mild, moderate, and severe symptoms should be differentiated [3, 7, 80].

Cronbach's alpha is the most widely used index of internal consistency in the psychometric model [81]. Cronbach himself [82] warned against increasing the number of items for improving the coefficient and provided a method of correction that is a measure of correlation among items. Unfortunately, such a method has been ignored. Feinstein [3], Bech [54], and Fava et al. [53, 83] further questioned the utility of pursuing an improvement of Cronbach's alpha by increasing the number of items in a rating scale. The longer the scale, in fact, the more problematic is its clinical use.

Another major problem related to the use of Cronbach's alpha coefficients has to do with redundancy (i.e., items that do not add clinical information as they actually measure parallel forms of the same symptom). The clinical utility of rating scales including redundant items is questionable, since they capture only a narrow part of the clinical condition under examination [83, 84]. Furthermore, the redundant nature of items may increase the internal consistency of the rating scale but is likely to decrease its sensitivity [5, 53, 83].

In the clinimetric approach, not all items necessarily carry the same clinical weight and the property of internal consistency is not required. In addition, items are not assumed to be inter-correlated as they should be locally independent (i.e., the probability of a positive score of an item should not depend on the positive score of any other item) and, most importantly, items should not provide redundant clinical information, as they should be multidimensional [9, 83]. "The multiple items are not expected to be homogeneous because they indicate different attributes of a complex clinical phenomenon" [5, p. 1203]. Bayes Information Criterion (BIC) is a statistical model that can be used in the clinimetric approach to avoid redundancy across items [85]. Specifically, this method may apply to the detection and selection of locally independent and multidimensional items providing distinctive clinical information.

Test-retest Reliability or Reproducibility

Test-retest reliability refers to the ability of a rating scale or index to produce the same results over time, assuming that the clinical dimension under assessment has remained unchanged [9]. However, in a clinical population there may be day-to-day variations in the parameters under analysis. In clinimetrics, this measurement property is not considered to be as important as other features such as sensitivity [83].

Sensitivity

The concept of sensitivity was introduced by Kellner in 1972 as the ability of a rating scale or self-reported questionnaire to differentiate patients from control subjects, to discriminate between different groups of patients suffering from the same illness (e.g., depressed inpatients and outpatients), and to reflect changes in experiments in therapeutics such as drug or psychotherapy trials [86]. Sensitivity may also extend to the ability of a clinimetric index to differentiate between wanted and unwanted effects of treatments and to discriminate between an active drug and placebo or between a specific psychotherapeutic treatment and attention placebo or clinical management [83]. In clinical trials that did not differentiate active treatments from placebo [87, 88], this failure might be due to poor performance of the treatment under consideration, but it might also be due to the lack of sensitivity of rating scales developed using psychometric criteria only [53]. PROMs may be valid and reliable according to psychometric principles, but they may lack sensitivity in a clinical context.

Sensitivity is crucial when treatment effects are small, in studies with limited sample size, and in the clinical process of assessment of recovery, with particular reference to residual and subclinical symptoms [89].

The concept of responsiveness refers to only one aspect of sensitivity (response to change) of a scale and not to its discriminating properties between different populations [4, 83].

Validity

For evaluating the validity of PROMs, clinimetric criteria differ from those conventionally used in the psychometric model and include the assessment of clinical, construct, biological, predictive, incremental, and concurrent validity.

Clinical Validity

The ability to discriminate between subjects with or without a condition is a core clinimetric criterion for evaluating the usefulness of an index [3, 53, 73, 74, 80, 84]. Such discrimination may also be achieved with the use of cut-off scores [9, 90, 91].

Construct Validity

The concept of construct validity was originally introduced by Cronbach and Meehl [92] and refers to the extent to which a rating scale adequately measures the underlying construct that is intended to measure [93]. In the psychometric model, construct validity is often evaluated using Factor or Principal Component Analyses [9, 94]. The utility of such psychometric analyses has been questioned [3, 9, 83]. Bech noted: "Factor analysis is a psychometric method that reveals a structure in an assessment scale, but not whether it is a dimension in which the total score is a meaningful expression of the severity of a condition" [9, p. 23].

According to the clinimetric approach, unidimensionality is not a specific requirement of PROMs. A method for exploring and understanding the dimensionalities of the index is represented by Item Response Theory (IRT), with models such as Rasch and Mokken analyses [9, 73, 74, 80, 95, 96]. In the Mokken analysis, the coefficient of scalability is used to investigate the construct validity of PROMs [9]. The clinimetric concept of scalability differs from the psychometric assumptions of homogeneity of components or unidimensionality [9, 73, 74, 83]. Testing the scalability means examining to what extent each item in a clinimetric index provides distinctive clinical information and determining whether symptoms included in a clinimetric index belong to an underlying clinical syndrome [9, 73, 74, 80].

Biological Validity

Gummel and Wildner [97] coined the expression "biological validity", but it was Bech [9, 98] who applied this concept to rating scales. Biological validity refers to the extent to which items of PROMs correspond to or reflect biological processes of the clinical condition under examination [9, 99].

Wright and Feinstein [5] noted that since in the psychometric model the items are combined only according to high statistical correlations, the scales may lack a coherent biologic conception despite the impressive mathematical associations.

Predictive Validity

The concept of predictive validity refers to the ability of a rating scale to predict response to treatment and patient outcomes [100].

Predictive validity has been considered an important issue in clinimetrics [5]. For instance, Topp et al. [64] supported the predictive validity of the 5-item version of the Well-Being Index (WHO-5), one of the most widely used PROMs for the assessment of psychological or subjective well-being. They showed that patients with cardiovascular disorders reporting a score < 50 on the WHO-5 had significantly higher mortality rates than those scoring ≥ 50 [64].

Incremental Validity

The concept of "incremental validity" was originally introduced by Sechrest [101] and was applied to clinimetrics by Fava and associates [4]. It refers to the distinctive contribution or increase in predictive ability associated with the inclusion of a particular instrument in the clinical decision process [4]. Regression analyses, particularly hierarchical approaches, can be used to test the incremental validity of PROMs [69], but other statistical methods (e.g., Bayesian statistics, multiple or partial correlations, and discriminant analyses) are also available [101, 102].

In the clinimetric approach, the choice of PROMs is dictated by the clinimetric principle of incremental validity: each distinct aspect of the clinical assessment should deliver a distinctive increase in information in order to qualify for inclusion [7].

The clinimetric concept of incremental validity also applies to the selection of items: each item should provide distinctive clinical information to be included [69].

Violation of the clinimetric principle of incremental validity may lead to conflicting results. According to the psychometric model, selection of assessment instruments includes several highly redundant measures, which are often included in a psychometric battery of tests with the misguided assumption that nothing will be missed [7].

Concurrent Validity

The term "concurrent validity" refers to the evaluation of the extent to which the rating scale significantly correlates with another related and previously validated assessment instrument [103]. Such a property is not considered to be as important as other clinimetric features. In the psychometric model, a high correlation is often regarded as evidence that two rating scales actually measure the same clinical factor. However, a high correlation does not indicate similar clinical validity: two rating scales may have a common content, that ensures a positive association, but they may display differential sensitivity and validity [53]. From a clinimetric perspective, concurrent validity of PROMs should be evaluated at baseline and at follow-up, considering the severity and staging of the clinical condition under assessment [80, 104].

Clinical Utility

The concept of clinical utility refers to the degree of influence that PROMs have not only on the clinical process of assessment but also on treatment planning [105-107]. Such a concept includes the clinimetric evaluation of the extent to which PROMs contribute to detect, predict, and

monitor symptom changes and treatment outcomes [105-107]. The following properties should be considered when evaluating the clinical utility of PROMs.

Sensibility

The concept of sensibility was originally introduced by Feinstein and his research group [3, 108, 109] and refers to the assessment of the ease with which PROMs can be used in clinical research and practice. Sensibility cannot be evaluated quantitatively using statistical analyses, but should instead include the assessment of "the specific purpose and clinical setting for which the index will be used" [109, p. 418].

Another key characteristic of sensibility refers to the ability of PROMs to facilitate clinician-patient interaction and collaboration [4, 108]. This clinimetric principle therefore has a number of important advantages not only for clinicians, as they will use PROMs quick and easy to administer and score, but also for patients who are actively involved in the clinical process of assessment [12, 100].

Format

To examine the structure of PROMs, several aspects should be considered, such as length of assessment, wording of items, and response format or calibration of items.

The use of a short scale with a limited number of items is indeed of great importance in clinical research and practice [9, 28, 59, 62-64, 110, 111].

Transferability

Bech et al. [112] provided a comprehensive definition of the concept of "transferability", that refers to whether a clinimetric index continues to measure the same clinical dimension across different groups of patients (e.g., men versus women, young versus old patients), or within the same

group of patients when the index is used for repeated measures over a treatment course. In other words, transferability consists in the assessment of "how universally useful a scale is" [112, p. 49].

Clinimetric Requirements

The following methodological recommendations apply to PROMs according to a clinimetric approach:

1. If the aim of the investigation is to discriminate between an active treatment and placebo, to differentiate between different groups of patients, or to capture changes in clinical trials, PROMs that were found to entail the clinimetric property of sensitivity should be used. The clinimetric properties of clinical and predictive validity should also be considered. Exploration of construct validity is not mandatory; it may be helpful for a better understanding of the properties (i.e., construct validity) of an index, particularly when PROMs are used to assess the burden or severity of the clinical condition under evaluation.

2. The principle of incremental validity, to avoid the use of multiple redundant instruments, should dictate the choice of PROMs to be included in an assessment battery.

Table 1 summarizes the main CLIPROM criteria that have been outlined in the present paper. They are shown in comparison with commonly used psychometric criteria. CLIPROM criteria apply not only to clinimetric indices such as the PsychoSocial Index [62, 63] and the Mental Pain Questionnaire [28], but also to the evaluation of existing PROMs originally developed within the psychometric model [65-68, 72]. Indeed, it is important to note that also when developed according to psychometric criteria, PROMs, rating scales and indices should not be discarded, since they may display important clinimetric properties. PROMs such as Kellner's Symptom Questionnaire [72] and the Hopkins Symptom Checklist [65-68], which were originally developed using psychometric principles, have been found to entail the clinimetric properties of sensitivity and clinical validity.

Conclusions

Psychometric criteria are often inadequate in the setting of clinical assessment because of their quest for homogeneity of components and lack of attention to clinical utility and sensitivity in the clinical environment [113].

In addition to the CLIPROM criteria, the main methodological recommendations that have been reported in the PROMIS and COSMIN initiatives [13, 15, 43, 48, 49] are summarized in Table 1. Simple reference to psychometric recommendations is no longer advised, given that PROMs are instruments to be used to assess clinical issues. Psychometric theory stifled the field and led to the routine use of scales widely accepted yet with a history of poor performance [113]. The guidelines developed by the Food and Drug Administration Center for Drug Evaluation and Research (CDER) in cooperation with the Center for Biologics Evaluation and Research (CBER) and the Center for Devices and Radiological Health (CDRH) [114] still focus on psychometric criteria and need to include the clinimetric principles of sensitivity and validity for PROMS. The choice of the most adequate PROMs should therefore depend on clinimetric criteria rather than on psychometric principles that are likely to clash with the complex clinical reality. The present paper underpins the importance of the clinimetric methodology in choosing the appropriate PROMs. Investigators and clinicians may weigh strengths and weaknesses of each assessment tool and select PROMs that best fit their needs. CLIPROM criteria challenge the traditional views of how PROMs should be developed and guide the construction of new indices and the validation process of existing PROMs to be employed in clinical settings.

Conflict of Interest Statement

Dr. Charlson reports the following grants: PCORI (2017 Cycle 3), NHLBI (T32) and NIMHD (T37) outside of the submitted work and Cornell University has filed a patent for the use of the enhanced comorbidity index to predict future costs. Dr. Concato reports that this article reflects the view of the authors and should not be construed to represent FDA's views or policies. Dr.

Furukawa reports grants and personal fees from Mitsubishi-Tanabe, personal fees from MSD, grants and personal fees from Shionogi, outside the submitted work. Dr. Furukawa has also a patent 2018-177688 concerning smartphone CBT apps pending, and intellectual properties for Kokoro-app licensed to Mitsubishi-Tanabe. Dr. Horwitz is senior advisor to McKinsey, and also member of the RWJF National Advisory Committee, and of Scripps Translational Research Institute, Scientific Advisory Committee. Dr. Nierenberg is in the Scientific Advisory Board of Alkermes, Jazz Pharma, Sage Pharma, Otsuka, and Neuronetics. Dr. Nierenberg reports consultancies with Acadia Pharm, Esai, Myriad, Ginger, Protogenics, and Merck. Dr. Nierenberg has also received honoraria from Sunovion and Neurostar. All other Authors have no conflicts of interest to declare.

Funding Sources

There are no funding sources to declare.

Author Contributions

D.C., C.P. and G.A.F. wrote the first draft. All other Authors revised and added contributions to the paper. All Authors reviewed and approved the final version of the paper.

References

1. Feinstein AR. T. Duckett Jones memorial lecture. The Jones criteria and the challenge of clinimetrics. Circulation. 1982 Jul;66(1):1–5.
2. Feinstein AR. An additional science for clinical medicine: IV. The development of clinimetrics. Ann Intern Med. 1983 Dec;99(6):843–8.
3. Feinstein AR. Clinimetrics. New Haven: Yale University Press; 1987.
4. Fava GA, Tomba E, Sonino N. Clinimetrics: the science of clinical measurements. Int J Clin Pract. 2012 Jan;66(1):11-5.
5. Wright JG, Feinsten AR. A comparative contrast of clinimetric and psychometric methods for constructing indexes and rating scales. J Clin Epidemiol. 1992 Nov;45(11):1201-18.
6. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th ed. Arlington (VA): American Psychiatric Association; 2013.
7. Fava GA, Rafanelli C, Tomba E. The clinical process in psychiatry: a clinimetric approach. J Clin Psychiatry. 2012 Feb;73(2):177-84.
8. Rush AJ, First MB, Blacker D. Handbook of psychiatric measures. Second edition. Washington: American Psychiatric Publishing; 2008.
9. Bech P. Clinical Psychometrics. Oxford: Wiley-Blackwell; 2012.
10. Guidi J, Brakemeier EL, Bockting CL, Cosci F, Cuijpers P, Jarrett RB, et al. Methodological recommendations for trials of psychological interventions. Psychother Psychosom. 2018 Sep;87(5):276-84.
11. Hunsley J, Mash EJ. A guide to assessments that work. Second edition. Oxford: Oxford University Press; 2018.
12. Bech P. Measurement-based care in mental disorders. New York: Springer-Verlag; 2016.
13. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The patient-reported outcomes measurement information system (PROMIS): progress of an NIH roadmap cooperative group during its first two years. Med Care. 2007 May;45(5 Suppl 1):S3-11.
14. Greenhalgh J. The applications of PROs in clinical practice: what are they, do they work, and why? Qual Life Res. 2009 Feb;18(1):115-23.
15. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. The patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. J Clin Epidemiol. 2010 Nov;63(11):1179-94.
16. Gershon RC, Rothrock N, Hanrahan R, Bass M, Cella D. The use of PROMIS and assessment center to deliver patient-reported outcome measures in clinical research. J Appl Meas. 2010;11(3):304–14.
17. Deshpande PR, Rajan S, Sudeepthi BL, Nazir CA. Patient-reported outcomes: a new era in clinical research. Perspect Clin Res 2011 Oct;2(4):137-44.

18. Rothrock NE, Kaiser KA, Cella D. Developing a valid patient-reported outcome measure. Clin Pharmacol Ther. 2011 Nov;90(5):737-42.

19. Magasi S, Ryan G, Revicki D, Lenderking W, Hays RD, Brod M, et al. Content validity of patient-reported outcome measures: perspectives from a PROMIS meeting. Qual Life Res. 2012 Jun;21(5):739-46.

20. Bevans M, Ross A, Cella D. Patient-reported outcomes measurement information system (PROMIS): efficient, standardized tools to measure self-reported health and quality of life. Nurs Outlook. 2014 Sep-Oct;62(5):339-45.

21. Coulter A. Measuring what matters to patients. BMJ. 2017 Feb;356:j816.

22. Greenhalgh J, Gooding K, Gibbons E, Dalkin S, Wright J, Valderas J, et al. How do patient reported outcome measures (PROMs) support clinician-patient communication and patient care? A realist synthesis. J Patient Rep Outcomes. 2018 Sep;2:42.

23. Calvert M, Kyte D, Price G, Valderas JM, Hjollund NH. Maximising the impact of patient reported outcome assessment for patients and society. BMJ. 2019 Jan;364:k5267.

24. Carlton J, Peasgood T, Khan S, Barber R, Bostock J, Keetharuth AD. An emerging framework for fully incorporating public involvement (PI) into patient-reported outcome measures (PROMs). J Patient Rep Outcomes. 2020 Dec;4(1):4.

25. Ciani O, Federici CB. Value lies in the eye of the patients: the why, what, and how of patient-reported outcomes measures. Clin Ther. 2020 Jan;42(1):25-33.

26. Jacobson RP, Kang D, Houck J. Can patient-reported outcomes measurement information system® (PROMIS) measures accurately enhance understanding of acceptable symptoms and functioning in primary care? J Patient Rep Outcomes. 2020 May;4(1):39.

27. Stover AM, Haverman L, van Oers HA, Greenhalgh J, Potter CM. Using an implementation science approach to implement and evaluate patient-reported outcome measures (PROM) initiatives in routine care settings. Qual Life Res. 2020 Jul; 1-19. Available from: https://doi.org/10.1007/s11136-020-02564-9.

28. Fava GA, Tomba E, Brakemeier EL, Carrozzino D, Cosci F, Eöry A, et al. Mental Pain as a Transdiagnostic Patient-Reported Outcome Measure. Psychother Psychosom. 2019 Nov;88(6):341-9.

29. Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: a structured review. J Eval Clin Pract. 2006 Oct;12(5):559-68.

30. Fries JF, Bruce B, Cella D. The promise of PROMIS: using item response theory to improve assessment of patient-reported outcomes. Clin Exp Rheumatol. 2005 Sep-Oct;23(5 Suppl 39):S53-7.

31. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the patient-reported outcomes measurement information system (PROMIS). Med Care. 2007 May;45(5 Suppl 1):S22-31.

32. Chan AW, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jerić K, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. Ann Intern Med. 2013 Feb;158(3):200-7.

33. Vodicka E, Kim K, Devine EB, Gnanasakthy A, Scoggins JF, Patrick DL. Inclusion of patient-reported outcome measures in registered clinical trials: evidence from ClinicalTrials.gov (2007-2013). Contemp Clin Trials. 2015 Jul;43:1-9.

34. Calvert M, Kyte D, Mercieca-Bebber R, Slade A, Chan AW, King MT, et al. (2018). Guidelines for inclusion of patient-reported outcomes in clinical trial protocols: the SPIRIT-PRO extension. JAMA. 2018 Feb;319(5):483-94.

35. de Jong MJ, Huibregtse R, Masclee AA, Jonkers DM, Pierik MJ. Patient-reported outcome measures for use in clinical trials and clinical practice in inflammatory bowel diseases: a systematic review. Clin Gastroenterol Hepatol. 2018 May;16(5):648-63.

36. Mehran R, Baber U, Dangas G. Guidelines for patient-reported outcomes in clinical trial protocols. JAMA. 2018 Feb;319(5):450-1.

37. Ishaque S, Karnon J, Chen G, Nair R, Salter AB. A systematic review of randomised controlled trials evaluating the use of patient-reported outcome measures (PROMs). Qual Life Res. 2019 Mar;28(3):567-92.

38. McGee RG. How to include patient-reported outcome measures in clinical trials. Curr Osteopor Rep. 2020 Aug;18(5):480-5.

39. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res. 2002 May;11(3):193-205.

40. Chassany O, Sagnier P, Marquis P, Fullerton S, Aaronson NK, for the European Regulatory Issues on Quality of Life Assessment Group. Patient-reported outcomes: the example of health-related quality of life - a European guidance document for the improved integration of health-related quality of life assessment in the drug regulatory process. Drug Inf J. 2002 Jan;36(1):209-38.

41. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. Protocol of the COSMIN study: COnsensus-based Standards for the selection of health Measurement INstruments. BMC Med Res Methodol. 2006 Jan;6(1):2.

42. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007 Jan;60(1):34-42.

43. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international delphi study. Qual Life Res. 2010 May;19(4):539-49.

44. Porter ME, Larsson S, Lee TH. Standardizing patient outcomes measurement. N Engl J Med. 2016 Feb;374(6):504-6.

45. Bombardier C, Tugwell P. Methodological considerations in functional assessment. J Rheumatol Suppl. 1987 Aug;14(Suppl 15):6-10.

46. McDowell I, Jenkinson C. Development standards for health measures. J Health Serv Res Policy. 1996 Oct;1(4):238-46.

47. Andresen EM, Meyers AR. Health-related quality of life outcomes measures. Arch Phys Med Rehabil. 2000 Dec;81(12 Suppl 2):S30-45.

48. Mokkink LB, Prinsen CA, Bouter LM, de Vet HCW, Terwee CB. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. Braz J Phys Ther. 2016 Mar-Apr;20(2):105-13.

49. Mokkink LB, de Vet HCW, Prinsen CA, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res. 2018 May;27(5):1171-9.

50. Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD, Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group. What is sufficient evidence for the reliability and validity of patient‐reported outcome measures? Value Health. 2007 Nov-Dec;10(Suppl 2):S94-105.

51. Kelley TA. International consortium for health outcomes measurement (ICHOM). Trials 2015 Nov;16(Suppl 3):O4.

52. Francis DO, McPheeters ML, Noud M, Penson DF, Feurer ID. Checklist to operationalize measurement characteristics of patient-reported outcome measures. Syst Rev. 2016 Aug;5(1):129.

53. Fava GA, Ruini C, Rafanelli C. Psychometric theory is an obstacle to the progress of clinical research. Psychother Psychosom. 2004 May-Jun;73(3):145–8.

54. Bech P. Modern psychometrics in clinimetrics: impact on clinical trials of antidepressants. Psychother Psychosom. 2004 May-Jun;73(3):134-8.

55. Tomba E, Bech P. Clinimetrics and clinical psychometrics: macro-and micro-analysis. Psychother Psychosom. 2012 Oct;81(6):333-43.

56. Feinstein AR. Multi-item “instruments” vs Virginia Apgar's principles of clinimetrics. Arch Intern Med. 1999 Jan;159(2):125-8.

57. Guidi J, Fava GA, Bech P, Paykel E. The Clinical Interview for Depression: a comprehensive review of studies and clinimetric properties. Psychother Psychosom. 2011;80(1):10-27.

58. Cosci F, Fava GA. Staging of mental disorders: systematic review. Psychother Psychosom. 2013 Nov;82(1):20-34.

59. Fava GA, Bech P. The concept of euthymia. Psychother Psychosom. 2016 Jan;85(1):1–5.

60. Guidi J, Tomba E, Cosci F, Park SK, Fava GA. The role of staging in planning psychotherapeutic interventions in depression. J Clin Psychiatry. 2017 Apr;78(4):456-63.

61. Fava GA, Guidi J. The pursuit of euthymia. World Psychiatry. 2020 Feb;19(1):40-50.

62. Sonino N, Fava GA. A simple instrument for assessing stress in clinical practice. Postgrad Med J. 1998 Jul;74(873):408-10.

63. Piolanti A, Offidani E, Guidi J, Gostoli S, Fava G, A, Sonino N. Use of the PsychoSocial Index: a sensitive tool in research and practice. Psychother Psychosom. 2016 Oct;85(6):337-45.

64. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 well-being index: a systematic review of the literature. Psychother Psychosom. 2015 Apr;84(3):167-76.

65. Carrozzino D, Vassend O, Bjørndal F, Pignolo C, Olsen LR, Bech P. A clinimetric analysis of the Hopkins Symptom Checklist (SCL-90-R) in general population studies (Denmark, Norway, and Italy). Nord J Psychiatry. 2016 Jul;70(5):374-9.

66. Arrindell WA, Urbán R, Carrozzino D, Bech P, Demetrovics Z, Roozen HG. SCL-90-R emotional distress ratings in substance use and impulse control disorders: one-factor, oblique first-order, higher-order, and bi-factor models compared. Psychiatry Res. 2017 Sep;255:173-85.

67. Carrozzino D, Morberg BM, Siri C, Pezzoli G, Bech P. Evaluating psychiatric symptoms in Parkinson's disease by a clinimetric analysis of the Hopkins Symptom Checklist (SCL-90-R). Prog Neuropsychopharmacol Biol Psychiatry. 2018 Feb;81:131-7.

68. Carrozzino D, Siri C, Bech P. The prevalence of psychological distress in Parkinson's disease patients: the brief symptom inventory (BSI-18) versus the Hopkins symptom checklist (SCL-90-R). Prog Neuropsychopharmacol Biol Psychiatry. 2019 Jan;88:96-101.

69. Carrozzino D, Svicher A, Patierno C, Berrocal C, Cosci F. The euthymia scale: a clinimetric analysis. Psychother Psychosom. 2019 Apr;88(2):119-21.

70. De Las Cuevas C, de Leon J. A clinimetric approach for improving the measurement of pharmacophobia with replication in two other samples. Psychother Psychosom. 2019 Apr;88(2):116-8.

71. Guidi J, Piolanti A, Gostoli S, Schamong I, Brakemeier EL. Mental pain and euthymia as transdiagnostic clinimetric indices in primary care. Psychother Psychosom. 2019 Aug;88(4):252-3.

72. Benasi G, Fava GA, Rafanelli C. Kellner's Symptom Questionnaire, a highly sensitive patient-reported outcome measure: systematic review of clinimetric properties. Psychother Psychosom. 2020 Mar;89(2):74-89.

73. Carrozzino D, Christensen KS, Cosci F. Construct and criterion validity of patient-reported outcomes (PROs) for depression: a clinimetric comparison. J Affect Disord. 2021 Mar;283:30-5.

74. Carrozzino D, Christensen KS, Mansueto G, Cosci F. Construct validity of the smoker complaint scale: a clinimetric analysis using item response theory (IRT) models. Addict Behav. 2021 Jun;117:106849.

75. Spearman C. General intelligence: objectively determined and measured. Am J Psychol. 1904 Apr;15(2):201-92.

76. Spearman C. The abilities of man: their nature and measurement. New York: Macmillan; 1927.

77. Stallings WM, Gillmore GM. A note on “accuracy” and “precision”. J Educ Meas. 1971;8(2):127-9.

78. Prinsen CA, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018 May;27(5):1147-57.

79. Hunsley J, Allan T. Psychometrics and psychological assessment. In: Sellbom M, Suhr J, editors. The Cambridge handbook of clinical assessment and diagnosis. Cambridge: Cambridge University Press; 2019. p. 9-24.

80. Carrozzino D, Patierno C, Fava GA, Guidi J. (2020) The Hamilton rating scales for depression: a critical review of clinimetric properties of different versions. Psychother Psychosom. 2020 May;89(3):133-50.

81. Streiner DL. Starting at the beginning: an introduction to coefficient alpha and internal consistency. J Pers Assess. 2003 Feb;80(1):99-103.

82. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951 Sep;16:297-334.

83. Fava GA, Carrozzino D, Lindberg L, Tomba E. The clinimetric approach to psychological assessment: a tribute to Per Bech, MD (1942-2018). Psychother Psychosom. 2018 Nov;87(6):321-6.

84. Carrozzino D. Clinimetric approach to rating scales for the assessment of apathy in Parkinson's disease: a systematic review. Prog Neuropsychopharmacol Biol Psychiatry. 2019 Aug;94:109641.

85. Nierenberg AA, Ostacher MJ, Borrelli DJ, Iosifescu DV, Perlis RH, Desrosiers A, et al. The integration of measurement and management for the treatment of bipolar disorder: a STEP-BD model of collaborative care in psychiatry. J Clin Psychiatry. 2006;67(Suppl 11):3-7.

86. Kellner R. Part 2. Improvement criteria in drug trials with neurotic patients. Psychol Med. 1972 Feb;2(1):73-80.

87. Fava M, Evins AE, Dorer DJ, Schoenfeld DA. The problem of the placebo response in clinical trials for psychiatric disorders: culprits, possible remedies, and a novel study design approach. Psychother Psychosom. 2003 May-Jun;72(3):115–27.

88. Fava GA, Guidi J, Rafanelli C, Rickels K. The clinical inadequacy of the placebo model and the development of an alternative conceptual framework. Psychother Psychosom. 2017 Nov;86(6):332-40.

89. Fava GA, Tomba E, Bech P. Clinical pharmacopsychology: conceptual foundations and emerging tasks. Psychother Psychosom. 2017 May;86(3):134-40.

90. Bech P, Jørgensen B, Jeppesen K, Poulsen DL, Vanggaard T. Personality in depression: concordance between clinical assessment and questionnaires. Acta Psychiatr Scand. 1986 Sep;74(3):263-8.

91. He C, Levis B, Riehm KE, Saadat N, Levis AW, Azar M, et al. The accuracy of the Patient Health Questionnaire-9 algorithm for screening to detect major depression: an individual participant data meta-analysis. Psychother Psychosom. 2020 Jan;89(1):25-37.
92. Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychol Bull. 1955 Jul;52(4):281–302.
93. Strauss ME, Smith GT. Construct validity: advances in theory and methodology. Annu Rev Clin Psychol. 2009 Apr;5:1-25.
94. Delis DC, Jacobson M, Bondi MW, Hamilton JM, Salmon DP. The myth of testing construct validity using factor analysis or correlations with normal or mixed clinical populations: lessons from memory assessment. J Int Neuropsychol Soc. 2003 Sep;9(6):936-46.
95. Mokken RJ. A theory and procedure of scale analysis. Berlin: Mouton; 1971.
96. Rasch G. Probabilistic models for some intelligence and attainment tests. Expanded edition. Chicago: The University of Chicago Press; 1980.
97. Gummel H, Wildner GP. Relation of biological validity of malignant tumors to the time of survival of the tumor bearers. Chirurg. 1958 Jan;29(1):14-8.
98. Bech P. Measurement issues. In: D'Haenen H, den Boer JA, Willner P, editors. Biological Psychiatry. New York: Wiley; 2002. p. 25-36.
99. Østergaard SD, Bech P, Trivedi MH, Wisniewski SR, Rush AJ, Fava M. Brief, unidimensional melancholia rating scales are highly sensitive to the effect of citalopram and may have biological validity: implications for the research domain criteria (RDoC). J Affect Disord. 2014 Jul;163:18-24.
100. Bech P. Clinimetric dilemmas in outcome scales for mental disorders. Psychother Psychosom. 2016 Oct;85(6):323-6.
101. Sechrest L. Incremental validity: a recommendation. Educ Psychol Meas. 1963 Apr;23(1):153-8.
102. Hunsley J, Meyer, GJ. The incremental validity of psychological testing and assessment: conceptual, methodological, and statistical issues. Psychol Assess. 2003 Dec;15(4):446–55.
103. Bagby RM, Taylor GJ, Parker JD. The twenty-item Toronto Alexithymia Scale-II. Convergent, discriminant, and concurrent validity. J Psychosom Res. 1994 Jan;38(1):33-40.
104. Fava GA, Kellner R. Staging: a neglected dimension in psychiatric classification. Acta Psychiatr Scand. 1993 Apr;87(4):225-30.
105. Bech P. The instrumental use of rating scales for depression. Pharmacopsychiatry. 1984 Jan;17(1):22-8.
106. Applegate WB. Use of assessment instruments in clinical settings. J Am Geriatr Soc. 1987 Jan;35(1):45-50.
107. Foster SL, Cone JD. Validity issues in clinical assessment. Psychol Assess. 1995;7(3):248–60.

108. Feinstein AR, Wells CK, Joyce CM, Josephy BR. The evaluation of sensibility and the role of patient collaboration in clinimetric indexes. Trans Assoc Am Physicians. 1985 Jan;98:146-9.

109. Feinstein AR, Josephy BR, Wells CK. Scientific and clinical problems in indexes of functional disability. Ann Intern Med. 1986 Sep;105(3):413-20.

110. Bech P. Rating scales for psychopathology, health status and quality of life. Berlin: Springer; 1993.

111. Sonino N, Peruzzi P. A psychoneuroendocrinology service. Psychother Psychosom. 2009 Sep;78(6):346-51.

112. Bech P, Allerup P, Rosenberg R. The Marke-Nyman temperament scale. Evaluation of transferability using the Rasch item analysis. Acta Psychiatr Scand. 1978 Jan;57(1):49-58.

113. Fava GA, Belaise C. A discussion on the role of clinimetrics and the misleading effects of psychometric theory. J Clin Epidemiol. 2005 Aug;58(8):753-6.

114. U.S. Department of Health and Human Services, Food and Drug Administration (FDA), Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Center for Devices and Radiological Health (CDRH). Guidance for Industry. Patient-Reported Outcome Measures: use in medical product development. Silver Spring: FDA; 2009. Available from: https://www.fda.gov/media/77832/download.

Table 1. Methodological approaches for the assessment of Patient-Reported Outcome Measures (PROMs)

Measurement property	CLIPROM criteria	PROMIS initiative	COSMIN checklist
Reliability	**Internal reliability or consistency** Homogeneity of components and unidimensionality are not required.	**Internal reliability** Homogeneous set of items.	**Internal consistency** Homogeneous and inter-correlated items.
Sensitivity	**Sensitivity** Ability to discriminate between different groups of patients, to differentiate patients from healthy controls, to detect changes in clinical trials, to discriminate between wanted and unwanted effects of treatments, and to differentiate active treatment from placebo.	**Responsiveness** The ability of an instrument to detect meaningful changes.	**Responsiveness** The ability of an instrument to detect change over time.
Validity	**Clinical validity** The ability to discriminate between subjects with or without a condition. **Construct validity** Evidence that items provide distinctive clinical information and belong to underlying clinical dimensions. **Predictive validity** The ability to predict response to treatment and clinical outcomes. **Concurrent validity** Significant correlations with another related and previously validated assessment instrument.	**Content validity** Evidence of moderate to strong correlations between the rating scale under evaluation and commonly used and accepted PROMs. **Convergent validity** Evidence of moderate to high correlations with existing measures covering the same concept. **Divergent validity** Evidence of low correlations with measures that assess different dimensions. **Construct validity** Psychometric evidence of the unidimensionality of the rating scale under examination. **Criterion validity** The degree to which an instrument agrees with an external standard measure.	**Content validity** Evaluation of the degree to which the content of an instrument is an adequate reflection of the construct to be measured. **Construct validity** The degree to which the scores of an instrument are consistent with hypotheses (e.g., relationships to scores of other instruments). This property also includes the evaluation of structural validity (i.e., the degree to which the scores of an instrument are an adequate reflection of the dimensionality of the construct to be measured). **Criterion validity** The degree to which the scores of an instrument are an adequate reflection of a "gold standard" measures.
Clinical utility	**Sensibility** Evaluation of the ease to use in clinical research and practice. **Format** Length of the index, wording and calibration of items.	**Interpretability** Items should be presented in a plain, easy-to-read format.	**Interpretability** Evaluation of the degree to which one can assign qualitative meaning to quantitative scores. **Feasibility** Evaluation of factors (e.g., completion time, length of the instrument, ease of administration) contributing to the ease of use in its intended setting.

CLIPROM, Clinimetric Patient-Reported Outcome Measures; PROMIS, Patient-Reported Outcomes Measurement Information System; COSMIN, COnsensus-based Standards for the selection of health Measurement INstruments.