

## Standardization of depression measurement: a common metric was developed for 11 self-report depression measures

Inka Wahl<sup>a,\*</sup>, Bernd Löwe<sup>a</sup>, Jakob Bue Bjorner<sup>b,c</sup>, Felix Fischer<sup>d,e</sup>, Gernot Langs<sup>f</sup>, Ulrich Voderholzer<sup>g,h</sup>, Stephen A. Aita<sup>i</sup>, Niels Bergemann<sup>j</sup>, Elmar Brähler<sup>k</sup>, Matthias Rose<sup>a,e,l</sup>

<sup>a</sup>Department of Psychosomatic Medicine and Psychotherapy, University Medical Center Hamburg-Eppendorf & Schön Klinik Hamburg Eilbek, Martinistr. 52, 20246 Hamburg, Germany & Dehnhaiide 120, 22081 Hamburg, Germany

<sup>b</sup>Department of Public Health, Section of Social Medicine, University of Copenhagen, Gothersgade 160, 1014 Copenhagen, Denmark

<sup>c</sup>QualityMetric, 24 Allbion Road, Lincoln 02865, RI, USA

<sup>d</sup>Institute for Social Medicine, Epidemiology and Health Economics, Charité—Universitätsmedizin Berlin, Luisenstr. 57, 10117 Berlin, Germany

<sup>e</sup>Department of Psychosomatic Medicine, Charité—Universitätsmedizin Berlin, Luisenstr. 13a, 10117 Berlin, Germany

<sup>f</sup>Schön Klinik Bad Bramstedt, Birkenweg 10, 24576 Bad Bramstedt, Germany

<sup>g</sup>Schön Klinik Roseneck, Am Roseneck 6, 83209 Prien am Chiemsee, Germany

<sup>h</sup>Clinic for Psychiatry and Psychotherapy, University Hospital of Freiburg, Hauptstr. 5, 79104 Freiburg, Germany

<sup>i</sup>Schön Klinik Starnberger See, Münchner Straße 23-29, 82335 Berg, Germany

<sup>j</sup>Saxon Hospital for Psychiatry and Neurology, Bahnhofstr. 1, 08228 Rodewisch, Germany

<sup>k</sup>Department of Medical Psychology and Medical Sociology, Leipzig University, Ph.-Rosenthal-Str. 55, 04103 Leipzig, Germany

<sup>l</sup>Department of Quantitative Health Sciences, UMASS Medical School, 368 Plantation Street, The Albert Sherman Center, Worcester, MA 01605, USA

Accepted 16 April 2013

### Abstract

**Objectives:** To provide a standardized metric for the assessment of depression severity to enable comparability among results of established depression measures.

**Study Design and Setting:** A common metric for 11 depression questionnaires was developed applying item response theory (IRT) methods. Data of 33,844 adults were used for secondary analysis including routine assessments of 23,817 in- and outpatients with mental and/or medical conditions (46% with depressive disorders) and a general population sample of 10,027 randomly selected participants from three representative German household surveys.

**Results:** A standardized metric for depression severity was defined by 143 items, and scores were normed to a general population mean of 50 (standard deviation = 10) for easy interpretability. It covers the entire range of depression severity assessed by established instruments. The metric allows comparisons among included measures. Large differences were found in their measurement precision and range, providing a rationale for instrument selection. Published scale-specific threshold scores of depression severity showed remarkable consistencies across different questionnaires.

**Conclusion:** An IRT-based instrument-independent metric for depression severity enables direct comparisons among established measures. The “common ruler” simplifies the interpretation of depression assessment by identifying key thresholds for clinical and epidemiologic decision making and facilitates integrative psychometric research across studies, including meta-analysis. © 2014 Elsevier Inc. All rights reserved.

**Keywords:** Depression; Patient-reported outcomes; Item response theory; Item bank; Linking; Health outcome assessment

Conflict of interest: All authors have no financial conflicts of interest to declare.

**Funding:** This research did not receive external funding. The Charité—Universitätsmedizin Berlin, five hospitals of the Schön Kliniken Company (Bad Arolsen, Bad Bramstedt, Hamburg Eilbek, Roseneck, and Starnberger See), the Heidelberg University Hospital, 12 family practices in Heidelberg, and the Leipzig University supported the study by providing the anonymous data for secondary analysis.

\* Corresponding author. Department of Psychosomatic Medicine and Psychotherapy, University Medical Center Hamburg-Eppendorf, Martinistr. 52, 025, 20246 Hamburg, Germany. Tel.: +49-40-7410-54257; fax: +49-40-7410-54975.

### 1. Introduction

Depressive disorders are severe and widespread diseases, imposing a significant burden for the individual and the society [1,2]. Reliable tools for depression measurement are essential for case recognition [3–5], treatment monitoring [6,7], and clinical research in general [8–14]. Today, a plethora of carefully developed and well-established self-report instruments for the assessment of depressive symptoms exist. However, scores of these instruments are not directly comparable. The heterogeneity of scale-specific

### What is new?

#### Key findings

- In this study, a common metric was developed for the first time for a large number of established depression measures using item response theory methods.

#### What this adds to what was known?

- To date, the variety of different scales for the assessment of certain patient-reported outcomes (PROs) seriously impairs research and communication among clinicians. Thus, standardization of PRO measurement is urgently needed.
- The new standardized metric for depression severity provides easy comparability of scores, measurement range, and precision among the different scales.

#### What is the implication and what should change now?

- The results offer a conjoint definition and understanding of the latent depression construct as defined by the items from a variety of established depression questionnaires.
- The outlined standardization approach calibrating different depression measures to a common latent metric can be applied to the assessment of other PROs as well.

metrics seriously impairs comparability across study results and complicates communication among researchers and clinicians. Pooling study results from different depression measures in quantitative reviews or meta-analyses is difficult and may even lead to biased results [15,16]. To avoid this bias, some meta-analyses limit the selection of studies to those that use the same instrument(s) [6,7]. However, such restrictions lead to a significant loss of information.

It is recognized that results for biomedical parameters need to be comparable across laboratory methods and facilities [17], and in our opinion, this is equally important for the measurement of patient-reported outcomes (PROs) [18,19].

This issue has been identified earlier [15,16], but only the recent increases in computational power have enabled the introduction of new psychometric methods in this field of health care [20–25]. The most frequently discussed solution [26–28] to achieve a standardized metric for PROs is offered by the item response theory (IRT) [29–33]. Items of different established depression questionnaires can be included in one “item bank” to provide one common metric [34–36]. Some depression item banks have already been developed [37–42], but to our knowledge, no study so far

has attempted to establish a comprehensive metric to achieve comparability for a larger number of existing depression measures.

In this study, we aim to provide such a metric for some of the most established depression measures. This metric should allow the comparison of results from different instruments on one common “ruler,” like using different thermometers to measure temperature on a meaningful anchored metric.

## 2. Methods

### 2.1. Sample

The study is based on secondary data analysis. The total study sample contains data from seven clinical and three general population samples. The clinical samples were consecutively drawn within clinical routines or cross-sectional studies [37,43,44], including in- and outpatients with mental and/or medical conditions being treated in 7 hospitals and 12 family practices across Germany. Clinical diagnoses according to ICD-10 criteria were given by health-care providers. The representative population samples were randomly selected by a national polling company (USUMA) in the years 2006, 2007, and 2010 as face-to-face household surveys, as previously described [45]. Ethical approval was not required as decided by the Ethics Committee of the Medical Registering Authority Hamburg, Germany.

### 2.2. Instruments

Eleven depression measures were included in our analyses: the *Beck Depression Inventory* (BDI) [46,47] and its second edition (BDI-II) [48,49], two subscales of the Mood Adjective Check List (MACL, in German BSF) [50,51] (anxious depression and elevated mood), the depression scale of the *Brief Severity Index* (BSI) [52,53], the *Center for Epidemiologic Studies Depression Scale* (CES-D, in German ADS) [54,55], the depression scale of the *Hospital Anxiety and Depression Scale* (HADS) [56,57], the depression scale of the *ICD-10-Symptom-Rating* (ISR) [58], the *Mental Health Index* (MHI-5) as one scale of the SF-36 Health Survey [59,60], the depression scales of the Patient Health Questionnaire (PHQ), that is, the PHQ-9 [61,62] and the PHQ-2 [63,64], the WHO-Five Well-being Index (WHO-5) [65,66], and a new translation of the Patient-Reported Outcomes Measurement Information System (PROMIS) depression item bank [41,67]. Scale scores were computed according to the authors’ suggestions (sum, mean, 0–100, etc). Most instruments include only morbidity-phrased items (BDI, BDI-II, PHQ-9/-2, PROMIS depression item bank, and depression scales of BSI, ISR, and MACL), some only positively worded items (MACL elevated mood scale and WHO-5), and some both types of item phrasing (CES-D, HADS depression scale, and MHI-5). Participants answered between one and five instruments.

Download English Version:

<https://daneshyari.com/en/article/10513742>

Download Persian Version:

<https://daneshyari.com/article/10513742>

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