Cross-Cultural Psychometric Properties of the Hamilton Depression Rating Scale

Propriétés psychométriques transculturelles de l'échelle de dépression de Hamilton

Erik Vindbjerg, MSc[1,2] ⓘ, Guido Makransky, PhD[2], Erik Lykke Mortensen, MSc[3], and Jessica Carlsson, MD, PhD[1]

Abstract

Objective: The Hamilton Depression Rating Scale (HDRS) is considered the gold standard measure of depression. The factor structure of the HDRS is generally unstable, but 4 to 8 items appear to form a general depression factor. As transcultural studies of the HDRS have received little attention, and as most of the studies have taken a data-driven approach with a tendency to yield fragmented results, it is not clear if an HDRS general depression factor can also be found in non-Western populations. This is an important issue in deciding on the appropriateness of the scale as a gold standard in transcultural psychiatry.

Method: A systematic review was carried out to compare previously reported factor structures of the HDRS in non-Western cultures. Overlapping clusters across studies were identified and subsequently tested with confirmatory factor analysis (CFA) of responses from an independent sample.

Results: Fourteen relevant studies were identified, 12 of which were obtained. A general depression factor was identified, consisting of the following symptoms: depressed mood, guilt, loss of interests, retardation, suicide, and psychological anxiety. The subsequent CFA analysis supported the fit of this model.

Conclusions: This study indicates that a general depression cluster is manifest in responses to the HDRS across cultures. While psychometric properties of the full-length HDRS are still debated, the general depression cluster appears pertinent to the assessment of depression across cultures. We recommend that cross-cultural clinicians and researchers focus on the use of unidimensional depression scales, which are in agreement with this cluster.

Abrégé

Objectif : L'échelle de dépression de Hamilton (HDRS) est considérée comme étant la référence absolue pour mesurer la dépression. La structure factorielle de la HDRS est généralement instable, mais de 4 à 8 items semblent former un facteur de dépression générale. Comme les études transculturelles de la HDRS ont suscité peu d'attention, et comme la plupart des études ont adopté une approche axée sur les données avec une tendance à produire des résultats fragmentés, il n'est pas déterminé si un facteur de dépression générale de la HDRS peut aussi se trouver dans les populations non occidentales. C'est un critère important pour décider de la pertinence de l'échelle comme référence absolue en psychiatrie transculturelle.

Méthode : Une revue systématique a été menée pour comparer les structures factorielles de la HDRS précédemment étudiées dans les cultures non occidentales. Des groupes se chevauchant parmi les études ont été identifiés et subséquemment testés par une analyse factorielle confirmatoire (AFC) des réponses d'un échantillon indépendant.

[1] Competence Centre for Transcultural Psychiatry, Mental Health Centre Ballerup, Copenhagen, Denmark
[2] Department of Psychology, University of Southern Denmark, Odense, Denmark
[3] Department of Public Health and Center for Healthy Aging, University of Copenhagen, Denmark

Corresponding Author:
Erik Vindbjerg, Competence Centre for Transcultural Psychiatry, Capital Region of Denmark, Maglevænget 2, bygn. 14, 2750 Ballerup, Denmark.
Email: erik.vindbjerg@regionh.dk

Résultats : Quatorze études pertinentes ont été identifiées, dont 12 ont été obtenues. Un facteur de dépression générale a été identifié, qui consistait en des symptômes tels *humeur dépressive, culpabilité, perte d'intérêts, arriération, suicide,* et *anxiété psychologique.* L'AFC subséquente a appuyé la concordance de ce modèle.

Conclusions : Cette étude indique qu'un groupe de dépression générale est manifeste dans les réponses à la HDRS dans les différentes cultures. Bien que les propriétés psychométriques de la HDRS intégrale fassent encore l'objet de débats, le groupe de dépression générale semble pertinent pour l'évaluation de la dépression dans les différentes cultures. Nous recommandons que les cliniciens et chercheurs transculturels mettent l'accent sur l'utilisation des échelles de dépression unidimensionnelles, qui concordent avec ce groupe.

Since its introduction in the early 1960s, the Hamilton Depression Rating Scale (HDRS) has been a favoured scale in the evaluation of depression treatment. It is the most widely used observer-rated depression scale all over the world and is considered a gold standard in depression studies.[1-3] As this position has consolidated, however, criticism of the scale has also accumulated. Most notably, the scale has been criticized for poor content validity and a multidimensional structure that varies across studies.[1,2,4]

Since 1979, the only review of the validity of the HDRS was published in 2004 by Bagby and colleagues.[2] Here the authors concluded that while criteria are generally met for convergent, discriminant, and predictive validity, the scale lacks factorial and content validity.[2,5] Although the factor structure as a whole was considered unstable, the review found some support for a general depression factor consisting of depressed mood and suicide along with either guilt or psychic anxiety.[2] Other suggestions for a general depression factor tend to revolve around the same 4 symptoms, as illustrated in Table 1. As an example, Gibbons et al.[1] suggested an 8-item subscale with the additional inclusion of loss of interests, agitation, loss of libido, and somatic anxiety. The only subscale that has been repeatedly tested and suggested as a formal measure in itself, however, is the 6-item melancholia subscale defined by Bech and colleagues.[4]

Only a few cross-cultural studies are represented in the review by Bagby and colleagues.[2] This may partly reflect a lack of studies in this area by 2004, but also that existing publications were not always sufficiently indexed to appear in the MEDLINE search conducted by Bagby et al. Thus, it is not clear whether the findings in the review apply to non-Western populations. This has important implications: if the core concept of depression is not evident in Hamilton ratings of non-Western patients, this casts doubt on the validity of the scale in these populations and may indicate the need for a more appropriate gold standard for measuring depression in transcultural psychiatry. On the other hand, if a general depression construct of the HDRS is supported in non-Western samples, this supports the HDRS as a basis for further evaluation and potential refinement.

The aim of the present study is to evaluate if a general depression subscale of the HDRS is supported in non-Western populations. We approach this through 2 specific objectives, each assigned a separate section of this article: the first objective is to outline potential cross-cultural subscales of the HDRS in non-Western populations by conducting a review of HDRS studies in non-Western populations. Studies are included, which report on the factor structure of the scale. The second objective is to test and evaluate the resulting subscales using data from a sample of refugees from across the Middle East and former Yugoslavia. This offers an independent evaluation and potential support of the results obtained in the review.

Methods: Part A

Literature Search

The objective of the literature search was to identify studies that assess the factorial structure of the HDRS in non-Western populations. In the present context, we defined non-Western cultures as cultures originating from outside Europe and the United States. This is both a narrow and somewhat vague definition but sufficient in the sense that all the identified studies conducted outside geographic Europe and the United States were based on clearly non-Western samples. Also, we found no refugee or immigrant studies from Europe or the United States.

We made no restrictions on the HDRS version used in the identified studies. While a number of alternative scales have been derived from the original 17-item HDRS, the only one of these commonly used is the 21-item version. This version retains the original 17-item scale, both in content and format, and adds 4 symptoms to indicate subtypes of depression.

A search was conducted in both PubMed and PsycINFO with the keywords "Hamilton depression" AND "Crosscultural," as well as "Hamilton depression" AND "structure." This yielded 6 studies matching the above criteria. To further expand the search, all references in these studies were reviewed. For each newly identified and obtained study, this procedure was repeated until all references were exhausted of studies meeting the criteria. This procedure revealed 7 additional studies. Finally, 1 study was located by reviewing cross-cultural studies of depression in general (i.e., not specifically assessed by the HDRS).[6] Of the

Table 1. General Depression Models.

	Bech Melancholia Subscale (Ham-D6)	Maier and Phillip Severity Subscale	Bagby et al.[a]	Gibbons Global Depression Severity Scale
Depressed mood	x	x	x	x
Retardation	x	x		
Loss of interests	x	x		x
Guilt	x	x	x	x
Suicide			x	x
Anxiety, psychic	x	x	x	x
Agitation		x		x
Loss of libido				x
Anxiety, somatic				x
Somatic, general	x			

[a]Bagby et al.[2] suggested 2 overlapping 3-item clusters. The model included here combines these models into 4 items.

identified studies, 11 were obtainable online through PubMed, and 1 was obtained through direct contact with the authors.[7] For the remaining 2 studies, we were unable to get a response from the authors.[8,9] It should be noted that one of the included studies did not specify location or nationality of its sample.[7] As the article is published in the *Arab Journal of Psychiatry* and the first author is based in Jordan, we assume the sample to be from the Middle East.

Reporting Factorial Overlap across Studies

To report factorial overlap, we counted the number of studies in which a given pair of items loaded on the same factor. When an item loaded on more than 1 factor, only the highest loading was considered. The frequencies of each item pair were ordered in a matrix, allowing for a visual inspection of the most frequently co-occurring pairs or clusters of items. One study[10] did not report a factor structure but provided a correlation matrix, which we were able to subject to principal component analysis. For this, we used varimax rotation and Kaiser's normalisation, while the number of extracted components was determined with Cattell's scree test.[11] Three studies tested the 6-item melancholia subscale and did not report on the factor structure of remaining items.[12-14] To prevent potential overrepresentation of this predefined model, we allowed these 3 studies to have only 1 combined count in the matrix.

Results: Part A

Factorial Overlap across Previous Studies

The resulting list of studies is presented in Table 2, and an overview of methods and results of each study is presented in Table 3.[6,7,10,12-20] Four studies included Western samples for

comparison. The study by Furukawa and colleagues[15] qualified for this review due to its Japanese sample but also contains a US and a European sample. The reported factor solution reflects the best fit across all these samples rather than the Japanese sample specifically. Fleck and colleagues[16] included a French sample, and Binitie[18] included a British sample. As those populations were analysed independently, however, we could disregard them in our review. Last, the study by Bech and colleagues[21] includes numerous Western nationalities but does not present a factor solution for the HDRS. Instead, the focus is on testing a predefined 6 items model, which was originally defined in a Danish context.

The co-occurrence matrix is presented in Table 4. Visual inspection revealed 2 evident clusters. The most discrete cluster consists of the 3 insomnia items, with each pair (H3 and H4, H3 and H5, and H4 and H5) occurring in 8 studies. All remaining combinations of items involving insomnia were rare. Second, the items of mood, loss of interests, and psychomotor retardation co-occurred in 6 to 8 studies. Frequently co-occurring with this cluster was guilt and, to a lesser degree, suicide and psychological anxiety. A less evident, third cluster was indicated by a number of somatic symptoms: somatic anxiety, general somatic, hypochondria, and, to a lesser extent, gastrointestinal symptoms and weight loss. Contrary to the review by Bagby and colleagues,[2] we found no anxiety cluster.

Methods: Part B

Testing the Identified Models with an Independent Sample of Refugees

Given that a relatively large number of cross-cultural HDRS-17 responses were available to the authors ($n = 494$), it was possible to test the identified models with an independent sample. The objective was not to explore an optimal model for this particular population, thereby potentially arriving at an alternative to existing models. Rather, the objective was to test whether the already defined models, developed using data from other populations, would also provide a plausible model in our sample. To test this, we used confirmatory factor analysis (CFA).

For comparison, we also tested 2 alternative versions of the general depression subscale as defined by Bech et al.,[4] Gibbons et al.,[1] and a third version deduced from the review by Bagby et al.[2] The items of each model are defined in Table 1. In the case of Bagby et al., we constructed one 4-item factor from 2 alternative 3-item factors. This was necessary as 3 items do not suffice for a testable model.

Sample

All data for the CFA were collected at the Competence Centre for Transcultural Psychiatry (CTP) from 2008 to 2012.[22] All respondents were refugees in Denmark undergoing psychiatric treatment and participating in randomised trials at the CTP. All respondents fulfilled criteria for post-traumatic stress disorder (PTSD) according to the

Table 2. Overview of Studies Reporting on the Psychometric Properties of the HDRS in Non-Western Populations.

	Location (Sample Size)	Characteristics of Respondents
Furukawa et al.[15]	United States, Caucasians (*n* = 2828)	Patients diagnosed with major depression, single or recurrent, and undergoing treatment with antidepressants.
	Europe, Caucasians (*n* = 1120)	Exclusion criteria: severe physical illness, suicide risk, incomplete response.
	Japan (*n* = 1237)	
Fleck et al.[16]	Brazil (*n* = 70)	Inpatients diagnosed with major depression.
Daradkeh et al.[7]	Middle East (*n* = 73)	Inpatients diagnosed with depression. The nationality of the population of the study is not specified but assumed to be from the Middle East.
Akdemir et al.[6]	Ankara, Turkey (*n* = 94)	Sixty-two female and 32 male patients with any *DSM-III-R* depressive mood disorder (major depression, dysthymia, and depressive period of bipolar depression and major depression and dysthymia) diagnosed by the Structured Clinical Interview (SCID) for *DSM-III-R*. Forty-four healthy individuals served as a control group, but their responses were not included in the factor analysis.
Zheng et al.[17]	China (*n* = 329)	A total of 126 inpatients and 203 outpatients. Sampled in 24 different locations in China, both urban and rural. Mean GAF score: 43.6.
		Criteria for inclusion: 1) currently meeting criteria of one of the *DSM-III* depression-related diagnoses, 2) being at least 16 years of age, and 3) having at least 7 years of education.
Binitie[18]	Benin (unknown)	All patients diagnosed with an affective disorder.
Hamdi et al.[19]	United Arab Emirates (*n* = 100)	A total of 67 inpatients and 33 outpatients, all fulfilling the diagnostic criteria for "depressive disorders" (in accordance with the World Health Organization Diagnostic Criteria for Research). The sample consisted of 53 nationals and 47 expatriates. The expatriates were mostly from other Arab countries, with 9 individuals from the Asian subcontinent.
Lotrakul et al.[20]	Thailand (*n* = 50)	Fifty depressive patients, both inpatients and outpatients.
Jang et al.[10]	Korea (*n* = 1183)	Both in- and outpatients diagnosed with major depressive disorder (psychotic or nonpsychotic), dysthymia, or depressive disorder not otherwise specified according to *DSM-IV* criteria. Exclusion criteria: diagnosis of schizophrenia, organic mental disorder, seizure disorder, eating disorder, brief psychotic disorder, or schizoaffective disorder, or presence of clinically significant nephrological, haematological, cardiovascular, respiratory, cerebrovascular, or endocrinological disease.
Bech et al.[12]	Belgium, Brazil, Canada, Colombia, Denmark, England, France, Italy, Mexico, Spain, Sweden, the United States, Germany/ Austria (*n* = 1128)	All patients fulfilled *DSM-III* criteria for panic disorders with the following modifications: only 3 symptoms were required in an attack, rather than 4.
		Exclusion criteria: melancholia (endogenous depression), psychotic depression, any history of bipolar disorder, a current episode of depression that did not appear after the current episode of panic.
		All patients took part in a randomised trial, receiving imipramine or placebo. A third group receiving alprazolam was not included in this article. All 3 groups totalled 1128 patients at baseline. We assume this is the number of respondents used to analyse the HDRS.
Lee et al.[13]	China (*n* = 214)	Psychiatric outpatients; 44.4% were in a *DSM-IV* current major depressive episode.
Bachner[14]	Israel (*n* = 125)	Bedouin Arabs who were primary carers for cancer patients.

DSM, *Diagnostic and Statistical Manual of Mental Disorders;* GAF, Global Assessment of Functioning; HDRS, Hamilton Depression Rating Scale.

International Classification of Diseases, Tenth Revision (*ICD-10*) (F43.1), and 97% fulfilled criteria for major depression. Excluded from the trials were patients with any F1 or F2 *ICD-10* diagnosis. The sample consisted of a total of 494 subjects, of whom 293 (59%) were male and 201 (41%) female. The average time since arrival in Denmark was 14.8 years, and the average time since first trauma was generally high, with more than 20 years for 51% of the patients. The predominant country of origin was Iraq (37%), followed by the former Yugoslavia (14%), Iran (13%), Lebanon (12%), and Afghanistan (11%).

Analysis

CFA of our own data set was conducted using Mplus 6.1.[23] The following fit indices are reported: χ^2, the root mean square error of approximation (RMSEA), the comparative fit index (CFI), and the Tucker-Lewis index (TLI). In evaluating relative fit indices, we follow criteria suggested by Hu and Bentler,[24] where CFI and TLI $\geq$0.90 and RMSEA $\leq$0.06 indicate acceptable fit.

Only the general depression cluster was subjected to CFA as remaining clusters were considered too narrow (insomnia) or too vague (somatic) for replication with independent data to yield meaningful results. The general depression cluster was tested in both its narrower, most coherent 4-item composition and as a 6-item cluster incorporating the items of suicide and psychological anxiety.

Results: Part B

Fit statistics for the tested CFA models are listed in Table 5.

Table 3. Overview of Methods and Results.

Study (HDRS Version)	Mean (SD) Score	Internal Reliability α	Applied Factor Analytic Method(s)	Results: No. of Identified Factors/ Percentage of Variance Accounted for
Furukawa et al.[15] (HDRS-17)	Japan: 24.1 (5.2) Europe: 22.5 (4.7) United States: 22.8 (3.5)	—	Simultaneous component analysis (PCA). Multigroup confirmatory factor analysis.	5 factors/48.8% (Japan) 47.5% (Europe), 41.1% (United States)
Daradkeh et al.[7] (HDRS-17)	27.5 (4.8)	.6	EFA	5 factors/61.0%
Hamdi et al.[19] (HDRS-21)	—	.51	PCA, varimax rotation, and Kaiser normalization + cluster analysis	7 factors/61.2%
Binitie[18] (HDRS-21)	—	—	Unspecified	4 factors/not specified
Zheng et al.[17] (CHDS-17[a])	—	.714	PCA, varimax rotation	5 factors/52.4%
Akdemir et al.[6] (HDRS-17)	Male: 21.28 (8.22) Female: 24.16 (5.93)	.75	PCA, varimax rotation	5 factors/54.9% 6 factors/61.2%
Fleck et al.[16] (HDRS-17)	28.7 (6.1)	—	PCA, varimax rotation, and either Kaiser's or Lebart's criterion	Brazil sample: 4 factors/55% 6 factors/68%
Jang et al.[10] (HDRS-17)	Male: 19.76 (6.32) Female: 19.82 (5.96)		Pearson correlation matrix. Subjected to PCA only in this review. Two clusters obtained based on visual inspection of scree plot. Rotated with varimax and Kaiser's normalisation.	2 factors/25%
Lotrakul et al.[20] (HDRS-17)	19.36 (6.84)		EFA with varimax rotation and Kaiser's criterion	6 factors/68.8%
Bech et al.[12] (HamD6)	—	—	Rasch analysis	Melancholia subscale confirmed
Lee et al.[13] (HamD6)			Mokken scale analysis	Psychic anxiety violated invariant item ordering, and general somatic had minor violations of nonintersection. Otherwise, all 6 items met criteria.
Bachner[14] (HamD6)	7.14 (5.17)		Single-factor CFA	Melancholia subscale confirmed

CFA, confirmatory factor analysis; EFA, exploratory factor analysis; HDRS, Hamilton Depression Rating Scale; PCA, principal component analysis.
[a]The CHDS-17 is a Chinese literal translation of the HDRS-17, but clinicians are reported to often use culturally appropriate idioms.[14]

The Gibbons 8-item model and Bech's 6-item melancholia model both displayed sufficient CFI and TLI values, but both models also displayed an excessive RMSEA value of 0.068. Remaining models (i.e., the 4-item model deduced from the Bagby et al.[2] review and the 6- and 4-item models of the current review) displayed satisfactory CFI, TLI, and RMSEA values.

Discussion

The first objective of this study was to review studies reporting on the factorial structure of the HDRS in non-Western populations and to identify overlaps in the reported factor structures. The main finding was a 4- to 6-item general depression cluster. That a 3-item insomnia cluster was also found is arguably trivial, given the obvious content overlap of these items and the fact that they make up 18% of the total number of items in the scale. Similarly, it is not surprising that a number of somatic items load together, although this pattern was weak.

The general depression cluster consisted of depressed mood, guilt, retardation, loss of interest, and, to a lesser extent, suicide and psychological anxiety. This constellation was supported by the subsequent CFA analysis, displaying adequate fit with responses from the independent sample of refugees. The 4-item model of the review by Bagby et al.[2] also received support, while remaining models displayed excessive RMSEA values. No model should be disregarded based on these results, as they may reflect sample idiosyncrasies, but 2 overall conclusions may be drawn. First, the combined results of the review and the CFA indicate that a general depression subscale of the HDRS may be suitable for cross-cultural ratings of depression. Second, while the optimal configuration of such a subscale is not obvious, the assessment of retardation and loss of interest appears as pertinent as the 4 items of the Bagby et al.[2] review.

Bech's melancholia subscale offers the best agreement with our results, and short of deriving new subscales or adapting other scales, this subscale would appear to offer a pragmatic translation of our results into psychiatric

Table 4. Co-occurrence Matrix.

	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14	H15	H16	H17
Mood, H1																	
Guilt, H2	4a																
Suicide, H3	4a	3a															
Sleep, early, H4	0	0	0														
Sleep, middle, H5	0	0	0	8b													
Sleep, late, H6	0	0	0	8b	8b												
Interests, H7	7a	5a	2a	0	0	0											
Retardation, H8	6a	5a	1a	0	0	0	8a										
Agitation, H9	2	1	3	0	0	1	0	0									
Anxiety, psych., H10	4a	4a	2a	2	1	1	3a	2a	2								
Anxiety, somatic, H11	0	1	1	0	0	0	0	0	0	1							
Gastrointestinal, H12	2	0	1	0	0	0	1	1	1	2	2						
Somatic, general, H13	2	2	1	0	1	0	2	2	0	2	4c	3					
Loss of libido, H14	1	0	2	1	1	1	0	0	0	1	1	1	1				
Hypochondria, H15	0	0	0	0	0	0	0	0	0	2	4c	1	3c	1			
Loss of insight, H16	1	1	2	0	0	0	1	2	2	0	0	2	1	0	0		
Weight loss, H17	2	0	1	0	0	1	0	0	3	2	1	4	1	0	1	0	

Note: Each cell represents the number of studies in which a pair of items displays their highest loading on the same factor. The total number of studies is 8. Letters indicate the 3 clusters: a) general depression, b) insomnia, and c) somatic.

Table 5. Fit Statistics for the Tested Models.

Model	χ^2 (df)	P	RMSEA (90% CI)	CFI	TLI
Bagby et al.,[2] 4 items[a]	4.482 (2)	0.132	0.046 (0.000-0.110)	0.996	0.987
Melancholia subscale, 6 items	29.779 (9)	<0.001	0.068 (0.042-0.096)	0.972	0.953
Gibbons et al.,[1] 8 items	65.130 (20)	<0.001	0.068 (0.050-0.086)	0.963	0.949
Cluster a, 6 items[b]	21.372 (9)	0.011	0.053 (0.024-0.082)	0.986	0.976
Cluster a, 4 items[b]	2.782 (2)	0.249	0.028 (0.000-0.098)	0.998	0.994

CFI, comparative fit index; RMSEA, root mean square error of approximation; SRMR, standardized root mean square residual; TLI, Tucker-Lewis index.
$n = 494$.
[a]This model is a compound of the two 3-item factors reported by Bagby et al.[2]
[b]The general depression cluster, as reflecting consensus across the reviewed studies of non-Western respondents and as reported in Table 4. The 4-item model includes only the most overlapping symptoms: mood, guilt, interests, and retardation.

measurement in non-Western settings. If following arguments to retire the HDRS,[2] we would encourage psychometric studies of the Bech-Rafaelsen Melancholia Scale (MES).[25] The 11 items of the MES cover all of the content of the 6 items derived in this review, cover more systems of the official diagnostic systems than the HDRS, and are constructed with unidimensionality in mind.

Methodological limitations of the included studies may help explain why a more consistent overlap of symptoms was not found in this review. Most of the studies rely on exploratory factor analysis (EFA), which poses a number of challenges to direct comparison between studies. EFA mainly seeks to explain as much variance as possible in the individual data set rather than to test hypotheses across cultures. If a wider cross-cultural factor structure does not provide the best fit in the individual study, it is not reported. This problem is aggravated when the rationale for choosing among the various EFA procedures, such as estimators and types of rotation, is not declared. In some of the reviewed studies, reported factor solutions pose very little face

validity, and it is rarely clear whether this is due to a deliberate omission of other methods (e.g., oblique rotation) or if the results indeed represent the best face validity of a number of liberally explored alternatives. As the most prominent example of this, Zheng and colleagues[17] present their 3 first factors as an "anxiety/somatization/weight loss factor," an "agitation/insight factor," and a "depressed mood/suicide/genital-symptoms factor." Our concerns echo those of Henson and Roberts,[26] who in a review of 60 EFA-based studies found that methodological information was mostly too limited to allow external verification. They suggest that many of the researchers may have relied on the default options in the chosen statistical package. Another challenge of EFA is that of sample dependent symptoms. Factors such as depressive mood and loss of interests are likely to be sensitive across a wide range of depression severity, whereas items such as genital dysfunction and loss of insight may mostly be sensitive to more severe cases.[27] As a hypothetical example, loss of insight may be highly related to the core construct of depression

yet display a weak correlation on the HDRS because the sample is not severely depressed or because the response options are not defined in a way that makes them sensitive to more subtle differences in insight.

The reliance on factor analysis, whether EFA or CFA, springs from the inherent multidimensionality of the HDRS. Given the fragmented results presented in our as well as previous reviews, it may be time to give up the pursuit of a shared, multidimensional HDRS profile and rather pursue a unidimensional scale. In this context, the predominance of the HDRS still offers a pool of standard items already available across cultures, from which selections can be tested as unidimensional constructs. Departing from the nondescript multidimensional framework of the HDRS to pursue a unidimensional scale opens up for a much more rigorous psychometric validation. Of particular relevance to the field of transcultural psychiatry, approaches within the item response theory framework[28] allow for a detailed analysis of cultural and gender-based differences in the responses to individual items.

Conclusion

The present review and its subsequent CFA indicate that a general depression factor, consisting of depressed mood, guilt, suicide, loss of interests, retardation, and psychological anxiety, as assessed by the Hamilton depression interview, can be found across cultures in non-Western populations. Future research may help to more precisely delimit which additional items contribute most to such a model. This would entail putting the general depression model at the heart of the analysis, rather than focusing on all 17 to 21 items.

Data Access

Data for this study were obtained from 4 intervention studies at the CTP, based on approval by each of the 4 sponsors, the Danish Data Protection Agency, and the Danish National Committee on Research Ethics. The authors have no right to share the data.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethics and Consent to Participate

All participants filling out the questionnaires have given informed consent. The data collection has been approved by the Danish National Committee on Research Ethics (project IDs H-D-2009-038, H-4-2011-020, H-3-2012-020, and H-3-2013-080).

Funding

ORCID iD

Erik Vindbjerg http://orcid.org/0000-0002-8096-0663

References

1. Gibbons RD, Clark DC, Kupfer DJ. Exactly what does the Hamilton Depression Rating Scale measure? J Psychiatr Res. 1993;27(3):259-273.
2. Bagby RM, Ryder AG, Schuller DR, et al. Reviews and overviews the Hamilton Depression Rating Scale: has the gold standard become a lead weight? Am J Psychiatry. 2004; 161(12):2163-2177.
3. Stefanis CN, Stefanis NC. Diagnosis of depressive disorders: a review. In: Majodina M, Sartorius N, eds. Depressive disorders WPA series evidence and experience in psychiatry. New York: John Wiley; 1999:1-51.
4. Bech P, Allerup P, Gram LF, et al. The Hamilton Depression Scale. Acta Psychiatr Scand. 1981;63(3):290-299.
5. Bagby RM, Schuller DR, Ryder AG, et al. Dr. Bagby and colleagues reply. Am J Psychiatry. 2005;162:2397-2398.
6. Akdemir A, Türkçapar MH, Örsel SD, et al. Reliability and validity of the Turkish version of the Hamilton Depression Rating Scale. Compr Psychiatry. 2001;42(2):161-165.
7. Daradkeh T, Abou-Saleh M, Karim L. The factorial structure of the 17-item Hamilton Depression Rating Scale. Arab J Psychiatry. 1997;8:6-12.
8. Kim KI. Clinical study of depressive symptomatology. (I) Adjustment of Hamilton's Rating Scale for Depression. Neuropsychiatry (Seoul). 1977;16:36-45.
9. Amin Y, Daradkeh T, Hamdi E, Abou-Saleh M. The structure of the Hamilton Depression Rating Scale (HDRS-17): a confirmatory factor analysis. Arab J Psychiatry. 1999;10:16-23.
10. Jang S, Park Y, Jae Y, et al. The symptom frequency characteristics of the Hamilton Depression Rating Scale and possible symptom clusters of depressive disorders in Korea: the Crescend Study. Psychiatry Investig. 2011;8(4):312-319.
11. Cattell RB. The scree test for the number of factors. Multivariate Behav Res. 1966;1(2):245-276.
12. Bech P, Allerup P, Maier W, et al. The Hamilton scales and the Hopkins Symptom Checklist (SCL-90): a cross-national validity study in patients with panic disorders. Br J Psychiatry. 1992;160:206-211.
13. Lee CP, Liu CY, Hung CI. Psychometric evaluation of a 6-item Chinese version of the Hamilton Depression Rating Scale: Mokken scaling and item analysis. Asia-Pacific Psychiatry. 2017;9(3):1-3.
14. Bachner YG. Psychometric properties of responses to an Arabic version of the Hamilton Depression Rating Scale (HAM-D6). J Am Psychiatr Nurses Assoc. 2016;22(1):27-30.
15. Furukawa TA, Streiner DL, Azuma H, et al. Cross-cultural equivalence in depression assessment: Japan–Europe–North American study. Acta Psychiatr Scand. 2005;112(4):279-285.
16. Fleck MPDA, Chaves MLF, Poirier-Littré MF, et al. Depression in France and Brazil: factorial structure of the 17-item Hamilton Depression Scale in inpatients. J Nerv Ment Dis. 2004;192(2):103-110.

17. Zheng YP, Zhao JP, Phillips M, et al. Validity and reliability of the Chinese Hamilton Depression Rating Scale. Br J Psychiatry. 1988;152:660-664.

18. Binitie A. A factor-analytical study of depression across cultures (African and European). Br J Psychiatry. 1975;127:559-563.

19. Hamdi E, Amin Y, Abou-Saleh MT. Performance of the Hamilton Depression Rating Scale in depressed patients in the United Arab Emirates. Acta Psychiatr Scand. 1997;96(6):416-423.

20. Lotrakul M, Sukanich P, Chakkrit Sukying M, et al. The reliability and validity of Thai version of Hamilton Rating Scale for Depression. J Psychiatr Assoc Thail. 1996;41:235-246.

21. Bech P, Gram LF, Dein E, et al. Quantitative rating of depressive states. Acta Psychiatr Scand. 1975;51(3):161-170.

22. Carlsson J, Sonne C, Silove D. From pioneers to scientists: challenges in establishing evidence-gathering models in torture and trauma mental health services for refugees. J Nerv Ment Dis. 2014;202(9):630-637.

23. Muthén LK, Muthén BO. Mplus (Version 7). Los Angeles; 2012.

24. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model A Multidiscip J. 1999;6(1):1-55.

25. Bech P. The Bech-Rafaelsen Melancholia Scale (MES) in clinical trials of therapies in depressive disorders: a 20-year review of its use as outcome measure. Acta Psychiatr Scand. 2002; 106(4):252-264.

26. Henson RK, Roberts JK. Use of exploratory factor analysis in published research: common errors and some comment on improved practice. Educ Psychol Meas. 2006;66:393-416.

27. Santor DA, Coyne JC. Examining symptom expression as a function of symptom severity: item performance on the Hamilton Rating Scale for Depression. Psychol Assess. 2001;13(1):127-139.

28. De Ayala RJ. The theory and practice of item response theory. New York: Guilford; 2009.