



This is a repository copy of *Validation of the PHQ-9 in adults with dissociative seizures*.

White Rose Research Online URL for this paper:
<https://eprints.whiterose.ac.uk/173336/>

Version: Published Version

Article:

Lopez, M.B., Goldstein, L.H., Robinson, E.J. et al. (5 more authors) (2021) Validation of the PHQ-9 in adults with dissociative seizures. *Journal of Psychosomatic Research*, 146. 110487. ISSN 0022-3999

<https://doi.org/10.1016/j.jpsychores.2021.110487>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>



Validation of the PHQ-9 in adults with dissociative seizures

Maria Baldellou Lopez^a, Laura H. Goldstein^{a,*}, Emily J. Robinson^b, Silia Vitoratou^c,
Trudie Chalder^d, Alan Carson^e, Jon Stone^e, Markus Reuber^f

^a Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK

^b School of Population Health and Environmental Sciences, King's College London, UK

^c Psychometrics and Measurement Lab, Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK

^d Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK

^e Department of Clinical Neuroscience, Centre for Clinical Brain Sciences, University of Edinburgh, UK

^f Academic Neurology Unit, Royal Hallamshire Hospital, University of Sheffield, Sheffield, UK

ARTICLE INFO

Keywords:

Dissociative seizures

Depression

PHQ-9

Validation

Factor analysis

ABSTRACT

Background: The PHQ-9 is a self-administered depression screening instrument. Little is known about its utility and accuracy in detecting depression in adults with dissociative seizures (DS).

Objectives: Using the Mini – International Neuropsychiatric Interview as a reference, we evaluated the diagnostic accuracy of the PHQ-9 in adults with DS, and examined its convergent and discriminant validity and uniformity.

Methods: Our sample comprised 368 people with DS who completed the pre-randomisation assessment of the CODES trial. The uniformity of the PHQ-9 was determined using factor analysis for categorical data. Optimal cut-offs were determined using the area under the curve (AUC), Youden Index, and diagnostic odds ratio (DOR). Convergent and discriminant validity were assessed against pre-randomisation measures.

Results: Internal consistency of the PHQ-9 was high ($\alpha = 0.87$). While the diagnostic odds ratio suggested that a cut-off of ≥ 10 had the best predictive performance (DOR = 14.7), specificity at this cut off was only 0.49. AUC (0.74) and Youden Index (0.48) suggested a ≥ 13 cut-off would yield an optimal sensitivity (0.81) and specificity (0.67) balance. However, a cut-off score of ≥ 20 would be required to match specificity resulting from a cut-off of ≥ 13 in other medical conditions. We found good convergent and discriminant validity and one main factor for the PHQ-9.

Conclusions: In terms of internal consistency and structure, our findings were consistent with previous validation studies but indicated that a higher cut-off would be required to identify DS patients with depression with similar specificity achieved with PHQ-9 screening in different clinical and non-clinical populations.

1. Introduction

Dissociative seizures (DS) superficially resemble epileptic seizures but are not caused by the abnormal electrical discharges in the brain. DS are interpreted as an experiential and behavioural dissociative response to arousal and perceived as non-volitional [1]. Together with epilepsy and syncope, dissociative seizures account for 90% of clinical presentations with transient loss of consciousness. The prevalence of DS in the general population has been estimated as 50/100.000 [2]. Depression is one of the commonest comorbid disorders in patients with dissociative seizures (DS) [3–5]. Prevalence rates of depression in adults with DS range from 21% to 60% and are higher than in the general

population and patients with epilepsy [6]. Health-related quality of life (HRQoL) is lower in patients with DS compared to those with epilepsy [7] and closely correlated with depression [8]. A recent systematic review found that depression correlated more closely with HRQoL than with seizure-related factors, and lower depression scores were the only determinant of higher HRQoL [9], suggesting that successful treatment of depression may improve clinical outcomes in patients with DS [10].

Unanswered questions remain about how to optimally identify clinically relevant depression, and what contribution self-report tools can make to this. A recent meta-analysis compared studies based on self-report methods and clinical diagnoses and highlighted clear disparities in the reported prevalence of depression in adults with DS depending on

* Corresponding author at: Department of Psychology, PO77, Institute of Psychiatry, Psychology and Neuroscience, De Crespigny Park, London SE5 8AF, UK.

E-mail address: laura.goldstein@kcl.ac.uk (L.H. Goldstein).

<https://doi.org/10.1016/j.jpsychores.2021.110487>

Received 4 November 2020; Received in revised form 2 April 2021; Accepted 2 April 2021

Available online 8 April 2021

0022-3999/© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

the case ascertainment method [5]. There are a number of factors, which could adversely affect the accuracy of self-report measures of depression in this population; there is also the pragmatic issue of availability of expertise. For many patients, assessment by an experienced consultant neuropsychiatrist is simply not available and diagnostic support tools are a necessary part of service provision. Among other factors that may affect accuracy of assessment, psychological aspects of DS are often stigmatised [11,12], and patients may find it difficult to admit to negative emotions. Depressive symptoms may also present atypically in this population and, compared to patients with epilepsy, DS patients may be more likely to endorse somatic rather than affective symptoms of depression [5]. These characteristics and the high levels of somatic symptoms reported by many patients in addition to their DS could affect the specificity of screening tools for depression [13].

The Patient Health Questionnaire-9 (PHQ-9) [14] is a brief, self-administered tool that is widely used to screen for depressive symptoms in clinical and research settings. For instance, in the United Kingdom the PHQ-9 is not only employed by providers of “Improving Access to Psychological Therapies” (IAPT) to assess patients’ suitability for particular services but also as an outcome and performance measure of IAPT services. Although it has been widely used in clinical and non-clinical populations, current evidence regarding optimal PHQ-9 cut-offs for depression screening accuracy remains inconclusive. Due to the high specificity and sensitivity reported in the original validation study, a cut-off score of ≥ 10 has become the recommended threshold for the identification of individuals likely to have clinically significant depression [14]; however, inflexible adherence to a specific cut-off has been criticised by the same authors [15]. Two individual patient data meta-analyses across different patient populations have generated different values as optimal cut-offs: a PHQ-9 ≥ 10 [16], or PHQ-9 ≥ 14 [17].

To our knowledge, no study has yet sought to validate the use of the PHQ-9 and determine optimal cut-off scores in a DS population. Thus, its utility as a screening instrument and sensitive treatment outcome measure remains uncertain in this patient group. The primary aim of this study was, therefore, to measure the sensitivity and specificity of the PHQ-9 in detecting depression in patients with DS, using the Mini - International Neuropsychiatric Interview (M.I.N.I.) [18] as a “gold standard” diagnostic reference.

Previous studies indicate that the PHQ-9’s cognitive/affective (PHQ-9/CA), and somatic (PHQ-9/S) subscales may differentially contribute to depression screening accuracy in patients with epilepsy [19,20]. Therefore, secondary aims of our study included an examination of the uniformity of the PHQ-9 scale through exploratory and confirmatory factor analysis as well as an examination of the convergent and discriminant (construct) validity of the PHQ-9 in its use for people with DS.

2. Methods

2.1. Participants

The patients who contributed self-report and interview data for this analysis were participating in the CODES trial, a multi-centre randomised controlled trial (RCT) comparing the clinical and cost effectiveness of Cognitive Behavioural Therapy (CBT) plus standardized medical care (SMC) and SMC alone in adults with DS. Patient data were captured at the point of recruitment into the intervention phase of this study between January 2015 and May 2017 (see previous publications for details of the trial and recruitment procedure) [21–24]. Inclusion and exclusion criteria are given in Supplementary Material 1. Written informed consent was obtained from all participants at each phase of the CODES study and ethical approval was granted by the NRES London-Camberwell St Giles Ethics Committee (13/LO/1595).

2.2. Measures

The Mini - International Neuropsychiatric Interview (M.I.N.I.; version 6) [18] was used. It is a structured interview for psychiatric diagnoses, based on DSM-IV criteria and the International Classification of Diseases-10 (ICD-10) [25]. It comprises 17 modules, each with 8–10 questions, which measure the symptoms for common psychiatric disorders including major depression. Its reliability and validity are well established [26]. Interviewers were postgraduate research assistants with different training and professional backgrounds. All interviewers had previous direct experience with patients and received a full day training to administer the M.I.N.I. in compliance with the administration procedures laid out in the M.I.N.I. manual. A similar approach to validating the PHQ-9 against the M.I.N.I. has previously been undertaken in patients with epilepsy [20]. The M.I.N.I. was chosen for the main CODES trial and therefore for this study due to its ease of use and previously documented validity in comparison studies with more elaborate diagnostic instruments [27]. In view of our aim to collect a considerable amount of outcome data, we were keen to minimise the assessment load for study participants as much as possible. Although all patients were also assessed by psychiatrists, this assessment was not on the same occasion as the completion of M.I.N.I. and PHQ-9. Furthermore, the trial protocol for this pragmatic trial did not require psychiatrists to complete a structured diagnostic assessment.

The PHQ-9 [14] has been widely used to screen for major depression and grade depressive symptom severity. It includes questions about each of the nine DSM-IV criteria for depression: anhedonia, low mood, fatigue, poor appetite, poor concentration, low self-esteem, hyper or hypoactive behaviours, sleep disturbances and suicidality. Responses range from 0 (not at all) to 3 (nearly every day), with total scores ≥ 10 typically taken to indicate clinical depression. The PHQ-9 is perceived as reliable, with good sensitivity and specificity across a range of medical settings and populations [28], including patients with epilepsy [20,29]. It has been shown to have good construct validity and can be useful as a categorical or continuous measure of depression severity [14,30,31]. While there is evidence that the PHQ-2 (comprising only the two first items of PHQ-9) could be of similar value as the PHQ-9 as a screening tool for depression [32], we chose the PHQ-9 for this study because it is a more widely used instrument, captures a broader range of (potentially relevant) symptoms and contains item 9 which is appreciated by some services as an independent “risk” item (“Thoughts that you would be better off dead, or hurting yourself”).

The Generalised Anxiety Disorder Assessment-7 (GAD-7) [33] is a reliable and valid seven-item scale that can be used as a screening tool and severity measure for generalised anxiety disorder.

The Clinical Outcomes in Routine Evaluation-10 (CORE-10) [34] is a 10-item scale used to measure psychological distress and screen for mental illness. It has high internal reliability and validity [34].

A modified version of the Patient Health Questionnaire-15 (PHQ-15) [35] was used to identify somatic symptoms. This modification comprises 15 items referring to common somatic symptoms (excluding upper respiratory infections), 10 relating to neurological symptoms [36] and five referring to psychological symptoms derived from the Prime MD Questionnaire [37]. This is a dichotomous version that asked patients if they were ‘bothered a lot’ by each of the symptoms over the past month (0 = No, 1 = Yes).

The Beliefs About Emotions Scale (BES) [38] was used to measure beliefs about the unacceptability of experiencing or expressing negative emotions. The scale has good validity and internal consistency [38].

The Work and Social Adjustment Scale (WSAS) [39] assesses the ability to perform day-to-day tasks including work, home management, relationship interaction, and social and private leisure activities. High validity and test-retest reliability have been reported for the WSAS [39].

The Short Form 12-item Health Survey – version 2 (SF-12v2) [40] is a widely-used measure of health-related quality of life (HRQoL). It comprises eight subscales assessing mental and physical health domains and

gives rise to two summary scores: The Mental Component Summary (MCS) and Physical Component Summary (PCS). The SF-12v2 has good reliability and validity in measuring health status [41].

The European Quality of Life-5 Dimensions-5 Levels (EQ-5D-5L) [42] also measures HRQoL. It includes five dimensions (mobility, self-care, usual activities, pain/discomfort and anxiety/depression), each with five response levels ranging from no problems to extreme problems, and a visual analogue scale (EQ-5D-5L VAS). The EQ-5D-5L has adequate test-retest reliability and convergent / construct validity [42–44].

2.3. Statistical analyses

Before tackling the primary aim of this study, we explored the factor structure of the PHQ-9 in our DS population: to this end, the sample was randomly divided in two halves for exploratory and confirmatory factor analyses (EFA and CFA). In both cases, item factor analysis for categorical data was used, using the weighted least squares estimator (WLSMV [45]). Both split-half samples were sufficiently large ($N = 169$ and $N = 199$) for factor analyses since they captured over 100 responses with N s 18 or 22 times greater than the number of items [46].

Measures of fit that are reported include: relative chi-square (relative χ^2 : values close to 2 suggest close fit [47]), Root Mean Square Error of Approximation (RMSEA - values below 0.06 indicate close fit [48]), Comparative Fit Index (CFI: values above 0.95 indicate a close fit; [49]), Taylor-Lewis Index (TLI: values above 0.95 are required for close fit [48]), and Standardized Root Mean Residual (SRMR: values below 0.08 suggest a good fit [48]). Analyses were conducted in MPLUS [50].

The factor number in the EFA was explored using the Guttman-Kaiser criterion [51,52], parallel analysis [53] for categorical data using R package 'random.polychor.pa' for categorical data [54], a Cattell's scree plot [55] and goodness of fit indices.

In the CFA sample we fitted the unidimensional model, a two-factor model suggested by EFA and a bifactor model combining the two. Following the procedure described [56] we fitted a bifactor model to resolve the dimensionality problem in our data. A multiple indicators-multiple causes model (MIMIC; [57]) was fitted to the complete data to explore potential measurement non-invariance (biased measurements) with respect to age and gender (one adjusting for the other).

Table 1

Baseline clinical and demographic characteristics of the study sample.

		n (%) ^a (Overall N = 368)
Age (years)	Mean (SD) ^b	37.5 (14.3)
	Median (IQR) ^c [range]	35 (25, 48) [18, 78]
Age at first dissociative seizure (years)	Mean (SD)	30.9 (14.1) n = 365
	Median (IQR) [range]	29 (19, 42) [1, 76]
Duration of dissociative seizures disorder (years)	Mean (SD)	6.2 (8.8) n = 365
	Median (IQR) [range]	3 (1, 8) [0, 65]
Gender	Male	102 (27.7)
	Female	266 (72.3)
Ethnicity	White	330 (89.7)
	Mixed	17 (4.6)
	Black	6 (1.6)
	Asian	6 (1.6)
	Other	9 (2.4)
	Yes	195 (53.0)
Dissociative seizures diagnosed by video-EEG ^d	Yes (≥ 1)	247 (67.1)
Any previous DSM-IV ^e diagnoses from M.I.N.I. ^f	Yes (≥ 1)	255 (69.3)
Any current DSM-IV diagnoses from M.I.N.I.	Yes	193 (52.4)
Past diagnosis of Major Depressive Disorder from M.I.N.I.	Yes	114 (31.0)
Current diagnosis of Major Depressive Disorder from M.I.N.I.	Yes	

^a n = available sample size: where n < 368 this is due to missing data.

^b SD = standard deviation.

^c IQR = inter-quartile range.

^d EEG = electroencephalogram.

^e DSM-IV = Diagnostic and Statistical Manual of Mental Disorders, 4th Edition.

^f M.I.N.I. = Mini - International Neuropsychiatric Interview.

In order to address the primary aim of our study, the diagnostic accuracy of the PHQ-9 total score was assessed against Current Major Depressive Disorder diagnoses from the M.I.N.I.

The area under the receiver operating characteristics (ROC) curve (AUC) [58] was used to test the performance of the PHQ-9 at various cut-off points using Stata V16.0 (Stata, Texas).

There are many ways to determine optimal cut-off points. To help balance sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) we calculated the following summary statistics: PROC01, the distance between the point on the ROC curve and (0,1) that should be minimised; Youden index (sensitivity + specificity – 1), which should be as close to 1 as possible; and the diagnostic odds ratio (DOR), which should be maximised as it indicates the predictive performance of the PHQ-9 cut-off point. More specifically, the DOR calculates the odds of a patient categorised as 'depressed' in the M.I.N.I. as being classed 'depressed' on the PHQ-9, compared to a patient not found depressed.

Reliability of the PHQ-9 was assessed using Cronbach's alpha [59] and item-total correlations. Concurrent (convergent or discriminant) validity was assessed using parametric or non-parametric correlations as appropriate.

3. Results

3.1. Participants

Table 1 shows basic demographic information for our sample. Further information can be found elsewhere [22].

Table 2 presents scores on measures of psychological distress and somatic symptoms (PHQ-9, GAD-7, CORE-10, Modified PHQ-15), Beliefs about Emotions, psychosocial functioning (WSAS) and health-related quality of life (SF-12v2 and EQ-5D-5L VAS). Further classifications of responses on the EQ-5D-5L are summarised in Supplementary Material 2.

3.2. Dimensionality and internal consistency

Parallel analysis indicated that the 1-factor solution was suitable for

Table 2
Descriptive statistics for currently considered measures (maximum N = 368).

Measure	n ^a	Mean (SD) ^b [range]
PHQ-9 ^c (depression) (min-max 0–27) (Scores ≥10 typically taken to indicate clinical depression)	367	12.4 (6.6) [0, 27]
GAD-7 ^d (anxiety) (min-max 0–21) (≥5 = mild, ≥10 = moderate, ≥15 = severe)	368	9.8 (6.2) [0,21]
CORE-10 ^e (distress) (min-max 0–40) (≥11 = clinical distress)	368	18.2 (6.5) [4, 34]
Modified PHQ-15 ^f (somatic symptoms) (min-max 0–30) (Higher scores reflect more symptoms)	364	16.7 (6.5) [2,30]
Beliefs about Emotions (min-max 0–72) (Higher scores reflect stronger beliefs about the unacceptability of experiencing and expressing negative emotions)	367	41.9 (16.9) [0, 72]
WSAS ^g (Impact of DS on functioning) (min-max 0–40) (<10 subclinical, 10–20 significant impairment but less severe symptomatology, >20 moderate, severe or worse psychopathology)	366	22.7 (10.5) [0, 40]
SF-12v2 ^h Mental Component Summary score (0 = worst health 100 = best health)	366	37.8 (11.8) [13.4, 68.1]
SF-12v2 ^h Physical Component Summary score (0 = worst health 100 = best health)	366	39.7 (12.2) [13.4, 65.9]
EQ-5D-5L VAS ⁱ (Health today) (0 = worst health 100 = best health)	363	55.5 (23.0) [1,100]

^a n = available sample size.

^b SD = standard deviation.

^c PHQ-9 = Patient Health Questionnaire-9.

^d GAD-7 = Generalised Anxiety Disorder Assessment-7.

^e CORE-10 = Clinical Outcomes in Routine Evaluation-10.

^f PHQ-15 = Patient Health Questionnaire-15.

^g WSAS = Work and Social Adjustment Scale.

^h SF-12v2 = Short Form 12-item Health Survey - version 2.

ⁱ EQ-5D-5L VAS European Quality of Life-5 Dimensions-5 Levels Visual Analogue Scale.

Table 3
Goodness of fit indices per item factor analysis model.

Model	rel χ^2	RMSEA ^c (95% CI)	TLI ^d	CFI ^e	SRMR ^f
EFA ^a -1 factor	2.96	0.108 (0.081,0.136)	0.967	0.956	0.077
EFA ^a -2 factor	1.51	0.055 (0.000,0.094)	0.994	0.988	0.042
CFA ^b -1 factor	5.20	0.145 (0.122,0.169)	0.938	0.918	0.07
CFA ^b -2 factor	2.67	0.092 (0.066,0.118)	0.976	0.967	0.047
Bifactor model	1.45	0.035 (0.000,0.062)	0.998	0.995	0.018

^a EFA: exploratory factor analysis.

^b CFA confirmatory factor analysis.

^c RMSEA Root Mean Square Error of Approximation.

^d TLI Tucker-Lewis Index.

^e CFI Comparative fit index.

^f SRMR Standardized Root Mean Residual.

our data (see scree plot in Supplementary Material 3). One eigenvalue was larger than 1 (5.22), but the second was also close to 1 (0.96). Therefore, we fitted both the 1- and the 2-factor models. As presented in Table 3, a closer fit was achieved for the 2-factor model, but the fit of the 1-factor model was adequate. These results were replicated in CFA where both models fitted well.

However, as presented in Table 4, the item loadings on the uni-factor model were similar to the loadings of the items on the general factor of the bi-factor model. According to Reise et al. [56], this indicates that there is no loss of information if a 1-factor model is applied. Secondly, the loadings of the items on the general factor of the bi-factor model were substantially larger than their loadings on the specific factors they

were assigned to. According to Reise et al. [56] this indicates that specific factors reflected a general trait. Finally, the loadings of the items on their designated factors in the two-factor CFA model (where there was no general factor) were much larger than their loadings to their designated factors in the bifactor model (where there was a general factor). These results suggest that the general factor is the prominent factor and the specific factors essentially reflect the general factor. The PHQ-9 therefore emerges as a univariate tool in patients with DS.

Two significant direct effects were identified, namely the direct effect of age on item 7 (d.e. = 0.013, $p < 0.001$) and the direct effect of gender on item 5 (d.e. = 0.3, $p = 0.026$). As only one item per covariate was found non-invariant, with small effect sizes, we conclude that PHQ-9 is measurement invariant with respect to gender or age.

The internal consistency of the PHQ-9 was satisfactory ($\alpha = 0.87$), with no increase in alpha if any of the items were deleted and satisfactory individual item-total correlations ranging between 0.44 and 0.70.

3.3. Diagnostic accuracy

Using Current Major Depressive Disorder from the M.I.N.I. as the gold-standard, the overall area under the ROC curve (AUC) of the PHQ-9 in our sample was 0.80 (Fig. 1).

Table 5 indicates that, in our sample, using a PHQ-9 cut-off of ≥10 for detecting depression cases has the best predictive performance; i.e. depressed patients have the greatest odds (DOR = 14.7) of a positive result compared to non-depressed patients. This is consistent with the current clinical cut-off. However, this cut-off achieves high sensitivity at the cost of low specificity, whereby only half of the individuals above this threshold will have clinical depression.

Using the AUC, a more overall measure, a cut-off of ≥13 has the best performance, with an AUC of 0.74 and a Youden index of 0.48. This may be considered an optimal balance of sensitivity and specificity. The distance between the point on ROC curve and (0,1) is also better (smaller) if ≥13 is used as the cut-off compared to ≥10 (0.24 vs. 0.30), as well as the percentage of agreement (71.4% vs. 62.9%).

3.4. Clinical correlates

Table 6 shows evidence of convergent validity reflected by high correlations with the GAD-7 (0.779), CORE-10 (0.795) and the MCS score of the SF12v2 (−0.711). Discriminant validity is indicated by moderate to low correlations with Beliefs about Emotions, psychosocial functioning, health-related quality of life and the PCS score of the SF-12v2.

4. Discussion

4.1. Brief summary of diagnostic accuracy findings

To our knowledge, this is the first study assessing the validity of the PHQ-9 for depression-screening in adults with DS, comparing it to the M. I.N.I, a structured diagnostic tool. In our sample of DS patients, the PHQ-9 had very good internal consistency (Cronbach's $\alpha = 0.87$) and area under the curve (AUC = 0.80) values indicating high overall diagnostic accuracy. While the PHQ-9 should never, in isolation, be used as a diagnostic tool for clinical depression, these findings (and our confirmation of the questionnaire as a univariate instrument) suggest that it is a useful screening method for depressive symptoms allowing users to describe the likelihood of a diagnosis of clinical depression at particular cut-off scores. The 'optimal' PHQ-9 cut-off is likely to be highly dependent on the intention of a particular use. If employed as a clinical screening tool to identify depressed patients, a lower cut-off would ensure that no cases are missed, although expert assessments would reveal that many patients scoring above this threshold are not depressed. If used in research with the aim of selecting patients highly likely to be suffering from comorbid major depression, a higher cut-off

Table 4
Item factor analysis loadings per model.

	EFA ^a (N = 168)		CFA ^b (N = 199)					
<i>1-factor model</i>	Uni-factor		<i>1-factor model</i>	Uni-factor	<i>Bifactor model</i>	General factor		
	Item 1	0.83		Item 1		0.71	Item 1	0.71
	Item 2	0.82		Item 2		0.85	Item 2	0.65
	Item 3	0.70		Item 3		0.76	Item 3	0.78
	Item 4	0.75		Item 4		0.76	Item 4	0.80
	Item 5	0.71		Item 5		0.65	Item 5	0.71
	Item 6	0.86		Item 6		0.83	Item 6	0.69
	Item 7	0.66		Item 7		0.69	Item 7	0.71
	Item 8	0.66		Item 8		0.66	Item 8	0.73
<i>2-factors model</i>	Item 9	0.63	Item 9	0.72	Item 9	0.46		
	factor 1		<i>2-factors model</i>	factor 1		specific factor 1		
	Item 3	0.87		Item 3	0.80	Item 3	0.20	
	Item 4	0.80		Item 4	0.80	Item 4	0.25	
	Item 5	0.49 (0.31) ^c		Item 5	0.69	Item 5	0.00	
	Item 7	0.52		Item 7	0.73	Item 7	-0.13	
	Item 8	0.42 (0.31) ^c		Item 8	0.68	Item 8	-0.53	
	factor 2			factor 2		specific factor 2		
	Item 1	0.52 (0.4) ^c		Item 1	0.73	Item 1	0.31	
	Item 2	0.94		Item 2	0.87	Item 2	0.61	
Item 6	0.80	Item 6		0.85	Item 6	0.52		
Item 9	0.71	Item 9	0.72	Item 9	0.59			

^a EFA = exploratory factor analysis.
^b CFA = confirmatory factor analysis.
^c Cross loadings within parentheses.

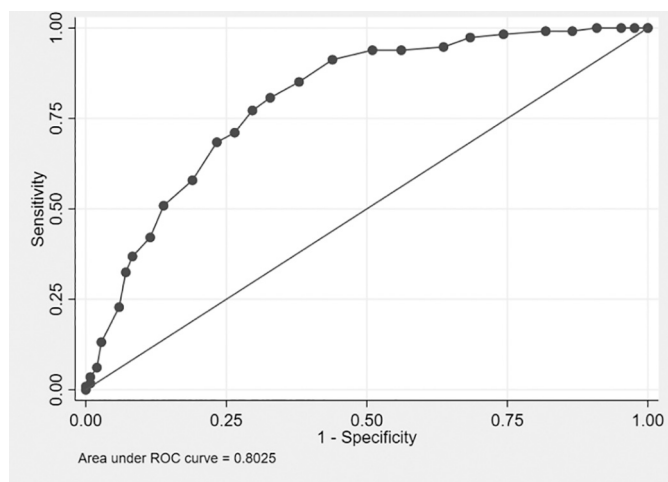


Fig. 1. Receiver Operating Characteristics (ROC) curve.

would be appropriate.

With this proviso, the present study found that using a PHQ-9 cut-off of ≥ 10 for detecting depression cases had the best predictive performance in terms of the diagnostic odds ratio (DOR = 14.7). This is consistent with the widely-used clinical cut-off [14]. However, a cut-off of ≥ 10 favoured sensitivity over specificity and greatly overestimated the presence of major depression. Using the AUC (a more global measure), we found that a cut-off of ≥ 13 had the best performance, with an AUC of 0.74 and a Youden index of 0.48, and provided an optimal balance of sensitivity (0.81) and specificity (0.67). At this score, over two-thirds of individuals with DS could be expected to have current major depression. However users will need to be aware that, at this cut off, one in five individuals with major depression is likely to be missed and one in three of those identified as depressed may not be found to be so on further examination. Nevertheless, the specificity of the instrument at a cut-off of ≥ 13 compares favourably with the poor specificity value of 0.49 at the ‘standard’ cut-off score of ≥ 10 . They should also be aware that the PHQ-9 identifies depressive symptoms, as indeed does the M.I.N.I. when used as per our protocol, rather than specifically diagnosing major depressive disorder. We would not, for example,

suggest commencing pharmacotherapy on this basis but one might suggest that these tools have a role to play in triaging referral to an appropriate clinician.

4.2. Findings in the context of previous literature

Although this study benefitted from a large sample size, it must be acknowledged that participants were captured at a particular time in their illness journey and may not have been typical of the DS population at large. Given that sensitivity and specificity will very depend very much on the prevalence of an observation, it is therefore important to note that, in terms of the distribution of PHQ-9 scores [5], but also in terms of age and gender distribution, comorbidity and social profile, our findings are consistent with those among other unselected DS cohorts [3,6]. Our findings show that a cut-off score of ≥ 10 overestimates depression diagnoses in patients with DS even more than in other medical conditions [16]. To match the level of specificity of the PHQ-9 in patients with other medical conditions at a cut-off of ≥ 13 , as indicated by Levis et al., [16], we would need a cut-of score of ≥ 20 in our dataset. Similar findings by the same authors [17] suggested that a higher cut-off score of ≥ 14 estimated prevalence closest to SCID-based diagnoses, and thus was the most accurate depression estimate.

A possible explanation for the much higher cut-offs needed to achieve high specificity of the PHQ-9 in our DS sample may be due to physical symptom confounding. The high scores on the somatic symptom scale (Modified PHQ-15) and the high level of disabilities reported on the PCS score by patients in our study are in keeping with evidence from other large studies of patients with DS demonstrating the high level of physical comorbidities and medication use in this patient group [60]. It is likely that patients diagnosed with DS who have physical comorbidity will need higher cut-off scores compared to patients screened for depression without such comorbidity. Notably, the symptoms addressed in the PHQ-9 (anhedonia, low mood, sleeping difficulties, tiredness, lack of appetite, low self-esteem, poor concentration and moving slowly or being fidgety) can be associated with physical comorbidity, disability or medication [61–63]. Support for this comes from previous findings indicating strong correlations between depression scores and somatic symptoms even where a different version of the PHQ-15 from that used here was employed and in different populations [64,65]. Indeed, we previously indicated [22] that our sample of DS patients most commonly

Table 5
Diagnostic accuracy of the PHQ-9 for Current Major Depressive Disorder (from the M.I.N.I.)

PHQ-9 score	Sensitivity	Specificity	PPV ^a	NPV ^b	% agree	LR + ^c	LR- ^d	PROC01 ^e	Youden index ^f	DOR ^g	AUC ^h
8	0.947	0.364	0.402	0.939	54.5	1.489	0.145	0.361	0.311	10.288	0.656
9	0.939	0.439	0.430	0.941	59.4	1.672	0.140	0.328	0.378	11.945	0.689
10	0.939	0.490	0.453	0.947	62.9	1.841	0.125	0.302	0.429	14.691	0.714
11	0.912	0.561	0.484	0.934	67.0	2.079	0.156	0.271	0.473	13.303	0.737
12	0.851	0.621	0.503	0.902	69.2	2.242	0.240	0.257	0.472	9.332	0.736
13	0.807	0.672	0.526	0.885	71.4	2.460	0.287	0.238	0.479	8.565	0.740
14	0.772	0.704	0.540	0.873	72.5	2.604	0.324	0.228	0.476	8.032	0.738
15	0.711	0.735	0.547	0.849	72.8	2.683	0.394	0.228	0.446	6.815	0.723
16	0.684	0.767	0.569	0.844	74.1	2.934	0.412	0.210	0.451	7.125	0.726
17	0.579	0.810	0.579	0.81	73.8	3.052	0.520	0.213	0.389	5.873	0.695
18	0.509	0.862	0.624	0.796	75.2	3.678	0.570	0.183	0.371	6.451	0.685
19	0.421	0.885	0.623	0.772	74.1	3.673	0.654	0.194	0.306	5.618	0.653
20	0.368	0.917	0.667	0.763	74.7	4.439	0.689	0.167	0.285	6.445	0.643

^a Positive predictive value.

^b Negative predictive value.

^c Positive likelihood ratio.

^d Negative likelihood ratio.

^e Distance between optimal point on ROC curve and (0,1).

^f Sensitivity + Specificity – 1.

^g Diagnostic odds ratio.

^h Area under the curve.

reported the following co-morbid symptoms on the Modified PHQ-15: tiredness and low energy, headaches, sleeping difficulties, memory and concentration problems, worry, pain in the extremities, dizziness, anxiety, depression in addition to other symptoms. Other factors, including stigmatisation of mental health complaints and response bias, appear prevalent in the DS population [5,11,12] and may explain the low combined sensitivity and specificity of the PHQ-9 in our sample compared to other medical conditions [16]. Indeed, discrepancies between self-report and clinically assessed depression have been previously reported in the DS population [5]. This may be due to DS patients having difficulty accepting the psychological nature of their diagnosis and disclosing negative emotions [11,12]. To a degree, emotional dysregulation and high levels of alexithymia, commonly reported in DS [66], may also compromise the sensitivity of self-report measures.

4.3. Implications of clinical correlates with the PHQ-9

Intercorrelations between the PHQ-9 and other baseline measures provided good evidence of construct validity when considering its use in patients with DS; high correlations were found between PHQ-9 and the GAD-7 and CORE-10 measures (0.779, 0.795 respectively) and with the MCS score of the SF-12v2 (–0.711). Of relevance here of course is the observation of wide overlap between depression, anxiety and somatic symptoms more generally in the population [61]. Discriminant validity was indicated by lower correlations between the PHQ-9 and Beliefs in Emotions, WSAS, EQ-5D-5L VAS, and the PCS score of the SF-12v2.

4.4. Factor structure

Factor analysis showed that for this population the PHQ-9 scale is unidimensional. To date, studies assessing the factor structure of the PHQ-9 yield inconclusive findings. While some studies have reported a two-factor structure [67] and found greater diagnostic accuracy using the PHQ9-CA subscale compared to the PHQ-9/S [20] (although the diagnostic accuracy of the PHQ9-CA was similar to the entire PHQ-9), several others suggest a unidimensional factor structure [68–70]. Since we have found one main factor, our results do not support considering a differential role of the cognitive/affective (PHQ-9/CA), and somatic (PHQ-9/S) subscales in diagnosing depression in this population.

4.5. Implications of a higher cut-off score

More generally, higher cut-offs ranging from 13 to 15 have been previously reported as ‘optimal’ across a range of conditions [17,20,67,71,72]. With respect to DS, the use of higher PHQ-9 cut-off scores may help reduce the overestimation of depression diagnoses in DS [5]. Determining the optimal diagnostic accuracy of the PHQ-9 will have important implications for depression screening in primary care services, which in turn, will influence management and treatment outcomes. Nonetheless, whether raising the PHQ-9 cut-off score to ≥ 13 will maximise accurate screening and mental health outcomes and minimise resource costs and adverse events compared to the traditional cut-off at ≥ 10 remains unclear.

4.6. Strengths and limitations

Our sample, derived from a high number of participating centres, allowed a detailed study of the validity of the PHQ-9 in a large, well-characterised cohort. We also used a cut-off score approach for the PHQ-9, which has been recently shown to be better than alternative methods, such as algorithms, at detecting clinical depression [73]. However, potential limitations of this study also need to be considered. Our eligibility criteria may have led to the exclusion of people with more marked psychopathology along with those who did not want to be randomised. Additionally, we validated self-report measures against the M.I.N.I. and these self-report measures can be susceptible to response bias. While the M.I.N.I. is a well-validated instrument for the diagnosis of depression which has been widely used to validate self-report measures, structured or semi-structures diagnostic instruments administered by experienced psychiatrists may have produced different results; however, the costs of doing so may have impacted on sample sizes [74]. For instance, previous studies indicate that, compared to the SCID, the M.I.N.I. may also overestimate the number of participants having clinical depression across a range of conditions [75,76], which might require the use of an even higher PHQ-9 cut-off score.

When interpreting our study, it is important to be aware that the PHQ-9 and the M.I.N.I. measure somewhat different constructs of depression. The PHQ-9 is a continuous measure of depressive symptomatology for which scores are given suggesting likelihood of a depressive disorder and consequent treatment implications. The M.I.N.I. by contrast offers a categorical definition of major depressive disorder. This is somewhat akin to measuring or reporting on blood pressure by

Table 6
Correlation matrix between PHQ-9 and other clinical measures.

	PHQ-9 ^a	GAD-7 ^b	CORE-10 ^c	Mod PHQ-15 ^d	Belief about Emotions	WSAS ^e	SF-12v2 ^f PCS ^g	SF-12v2 MCS ^h	EQ-5D-5L ⁱ mobility	EQ-5D-5L self care	EQ-5D-5L usual activities	EQ-5D-5L pain	EQ-5D-5L anxiety/depression	EQ-5D-5L VAS ^j health today
PHQ-9	–													
GAD-7	*0.779	–												
CORE-10	*0.795	*0.791	–											
Modified PHQ-15	*0.715	*0.604	*0.616	–										
Belief about Emotions	*0.428	*0.454	*0.421	*0.353	–									
WSAS	*0.485	*0.306	*0.367	*0.482	**0.147	–								
SF-12v2 PCS	*–0.330	**–0.103	**–0.148	*–0.424	–0.043	*–0.515	–							
SF-12v2 MCS	*–0.711	*–0.712	*–0.694	*–0.575	*–0.381	*–0.339	–0.075	–						
EQ-5D-5L mobility	*0.321	**0.141	*0.197	*0.375	0.093	*0.456	*–0.689	–0.070	–					
EQ-5D-5L self-care	*0.411	*0.245	*0.264	*0.444	**0.152	*0.507	*–0.609	*–0.219	*0.664	–				
EQ-5D-5L usual activities	*0.459	*0.300	*0.313	*0.424	*0.179	*0.529	*–0.609	*–0.230	*0.591	*0.668	–			
EQ-5D-5L pain	*0.465	*0.268	*0.340	*0.524	*0.173	*0.476	*–0.735	*–0.185	*0.667	*0.608	*0.573	–		
EQ-5D-5L anxiety/depression	*0.669	*0.675	*0.679	*0.540	*0.351	*0.365	*–0.206	*–0.635	*0.279	*0.308	*0.424	*0.352	–	
EQ-5D-5L VAS health today	*–0.563	*–0.398	*–0.454	*–0.509	*–0.213	*–0.460	*0.551	*0.352	*–0.469	*–0.455	*–0.536	*–0.534	*–0.478	–

* Significant at $p < 0.001$.

** Significant at $p < 0.05$.

^a PHQ-9 = Patient Health Questionnaire-9.

^b GAD-7 = Generalised Anxiety Disorder Assessment-7.

^c CORE-10 = Clinical Outcomes in Routine Evaluation-10.

^d Mod PHQ-15 = Modified Patient Health Questionnaire-15.

^e WSAS = Work and Social Adjustment Scale.

^f SF-12v2 = Short Form 12-item Health Survey - version 2.

^g PCS = Physical Component Summary.

^h MCS = Mental Component Summary.

ⁱ EQ-5D-5L European Quality of Life-5 Dimensions-5 Levels.

^j VAS = Visual Analogue Scale.

systolic/diastolic values or reporting a group of patients as suffering hypertension. There are advantages and disadvantages to both constructs. Kroenke and colleagues [14] in their original validation paper found some problems with mapping one construct onto the other and found a trade-off between sensitivity and specificity when mapping PHQ-9 scores to categorical diagnoses. Importantly, a recent meta-analysis demonstrated that both of these modes of assessment only showed low to moderate correlations with clinical mental health diagnosis (including depression) based on the assessment of an experienced clinician capable of taking account of a much wider range of personal and contextual circumstances [77].

In conclusion, our findings are not inconsistent with other PHQ-9 validation studies suggesting generalisability of our results and that use of the PHQ-9 in assessing depression in patients with DS should be considered, albeit with a higher than normally recommended cut-off score. While the PHQ-9 can be an effective screening instrument in patients with DS, a diagnosis of depression must involve assessment by an appropriately trained clinician.

Funding statement

This paper describes independent research funded by the National Institute for Health Research (Health Technology Assessment programme, 12/26/01, COgnitive behavioural therapy vs standardised medical care for adults with Dissociative non-Epileptic Seizures: A multicentre randomised controlled trial (CODES)). This paper also describes independent research part-funded (LHG, SV, TC) by the National Institute for Health Research (NIHR) Maudsley Biomedical Research Centre at the South London and Maudsley NHS Foundation Trust and King's College London and (EJR) by the NIHR Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. JS is supported by an NHS Scotland NHS Research Scotland (NRS) Career Fellowship and JS and AC also acknowledge the financial support of NRS through the Edinburgh Clinical Research Facility. MR benefitted from the support of the NIHR Sheffield Biomedical Research Centre (Translational Neuroscience). The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Declaration of Competing Interest

Markus Reuber is the paid Editor-in-Chief of Seizure-European Journal of Epilepsy and receives authorship fees from Oxford University Press in relation to a number of books about dissociative seizures. Jon Stone reports independent expert testimony work for personal injury and medical negligence claims, royalties from UpToDate for articles on functional neurological disorder and runs a free non-profit self-help website, www.neurosymptoms.org. Alan Carson reports being a paid editor of the Journal of Neurology, Neurosurgery and Psychiatry, and is the director of a research programme on functional neurological disorders; he gives independent testimony in Court on a range of neuropsychiatric topics (50% pursuer, 50% defender). The remaining authors have no conflict of interest to declare.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jpsychores.2021.110487>.

References

- [1] R.J. Brown, M. Reuber, Towards an integrative theory of psychogenic non-epileptic seizures (PNES), *Clin. Psychol. Rev.* 47 (2016) 55–70.
- [2] K. Kanemoto, W.C. LaFrance, R. Duncan, D. Gigineishvili, S.-P. Park, Y. Tadokoro, et al., PNES around the world: Where we are now and how we can close the diagnosis and treatment gaps—an ILAE PNES Task Force report, *Epilepsia Open* 2 (3) (2017 Jun 23) 307–316 (Erratum in: *Epilepsia Open*. 2019 Jan 07;4(1):219).

- [3] R.J. Brown, M. Reuber, Psychological and psychiatric aspects of psychogenic non-epileptic seizures (PNES): a systematic review, *Clin. Psychol. Rev.* 45 (2016) 157–182.
- [4] A.M. Kanner, S.C. Schachter, J.J. Barry, D.C. Hesdorffer, M. Mula, M. Trimble, et al., Depression and epilepsy, pain and psychogenic non-epileptic seizures: clinical and therapeutic perspectives, *Epilepsy Behav.* 24 (2) (2012) 169–181.
- [5] S. Walsh, L. Levita, M. Reuber, Comorbid depression and associated factors in PNES versus epilepsy: systematic review and meta-analysis, *Seizure-Eur J Epilep.* 60 (2018) 44–56.
- [6] W. Diprose, F. Sundram, D.B. Menkes, Psychiatric comorbidity in psychogenic nonepileptic seizures compared with epilepsy, *Epilepsy Behav.* 56 (2016) 123–130.
- [7] J.P. Szaflarski, M. Szaflarski, Seizure disorders, depression, and health-related quality of life, *Epilepsy Behav.* 5 (1) (2004) 50–57.
- [8] G.H. Rawlings, I. Brown, M. Reuber, Predictors of health-related quality of life in patients with epilepsy and psychogenic nonepileptic seizures, *Epilepsy Behav.* 68 (2017) 153–158.
- [9] B. Jones, M. Reuber, P. Norman, Correlates of health-related quality of life in adults with psychogenic nonepileptic seizures: a systematic review, *Epilepsia* 57 (2) (2016) 171–181.
- [10] W.C. LaFrance, G.I. Keitner, G.D. Papandonatos, A.S. Blum, J.T. Machan, C. E. Ryan, et al., Pilot pharmacologic randomized controlled trial for psychogenic nonepileptic seizures, *Neurology* 75 (13) (2010) 1166.
- [11] G.H. Rawlings, M. Reuber, What patients say about living with psychogenic nonepileptic seizures: a systematic synthesis of qualitative studies, *Seizure* 41 (2016) 100–111.
- [12] C. Robson, L. Myers, C. Pretorius, O.S. Lian, M. Reuber, Health related quality of life of people with non-epileptic seizures: the role of socio-demographic characteristics and stigma, *Seizure* 55 (2018) 93–99.
- [13] M. Reuber, House AO, R. Pukrop, J. Bauer, C.E. Elger, Somatization, dissociation and general psychopathology in patients with psychogenic non-epileptic seizures, *Epilepsy Res.* 57 (2–3) (2003) 159–167.
- [14] K. Kroenke, R.L. Spitzer, J.B. Williams, The PHQ-9: validity of a brief depression severity measure, *J. Gen. Intern. Med.* 16 (9) (2001) 606–613.
- [15] K. Kroenke, R.L. Spitzer, J.B. Williams, B. Löwe, The patient health questionnaire somatic, anxiety, and depressive symptom scales: a systematic review, *Gen. Hosp. Psychiatry* 32 (4) (2010) 345–359.
- [16] B. Levis, A. Benedetti, B.D. Thombs, Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis, *BMJ* 365 (2019) 1476.
- [17] B. Levis, A. Benedetti, J.P.A. Ioannidis, Y. Sun, Z. Negeri, C. He, et al., Patient Health Questionnaire-9 scores do not accurately estimate depression prevalence: individual participant data meta-analysis, *J. Clin. Epidemiol.* 122 (2020) 115–128.
- [18] D.V. Sheehan, Y. Lecrubier, K.H. Sheehan, P. Amorim, J. Janavs, E. Weiller, et al., The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10, *J. Clin. Psychiatry* 59 (Suppl. 20) (1998) 22–33.
- [19] A.J. Mitchell, N. Ioannou, J.M. Rampling, A. Sajid, T.J. von Oertzen, H.R. Cock, et al., Which symptoms are indicative of depression in epilepsy settings? An analysis of the diagnostic significance of somatic and non-somatic symptoms, *J. Affect. Disord.* 150 (3) (2013) 861–867.
- [20] J.S. Rathore, L.E. Jehi, Y. Fan, S.I. Patel, N. Foldvary-Schaefer, M.J. Ramirez, et al., Validation of the Patient Health Questionnaire-9 (PHQ-9) for depression screening in adults with epilepsy, *Epilepsy Behav.* 37 (2014) 215–220.
- [21] L.H. Goldstein, J.D. Mellers, S. Landau, A. Carson, N. Medford, et al., COgnitive behavioural therapy vs standardised medical care for adults with Dissociative non-Epileptic Seizures (CODES): a multicentre randomised controlled trial protocol, *BMC Neurol.* 15 (2015) 98.
- [22] L.H. Goldstein, E.J. Robinson, J.D.C. Mellers, J. Stone, A. Carson, T. Chalder, et al., Psychological and demographic characteristics of 368 patients with dissociative seizures: data from the CODES cohort, *Psychol. Med.* (2020 May 11) 1–13.
- [23] L.H. Goldstein, E.J. Robinson, J.D.C. Mellers, J. Stone, A. Carson, M. Reuber, et al., Cognitive behavioural therapy for adults with dissociative seizures (CODES): a pragmatic, multicentre, randomised controlled trial, *Lancet Psychiatry* 7 (6) (2020) 491–505.
- [24] E.J. Robinson, L.H. Goldstein, P. McCrone, I. Perdue, T. Chalder, J.D.C. Mellers, et al., COgnitive behavioural therapy versus standardised medical care for adults with Dissociative non-Epileptic Seizures (CODES): statistical and economic analysis plan for a randomised controlled trial, *Trials*. 18 (1) (2017) 258.
- [25] World Health Organisation, The ICD-10 Classification of Mental and Behavioural Disorders. Clinical Description and Diagnostic Guidelines, World Health Organisation, Geneva, 1992.
- [26] Y. Lecrubier, D.V. Sheehan, E. Weiller, P. Amorim, I. Bonora, K. Harnett Sheehan, et al., The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI, *Eur. Psychiat.* 12 (5) (1997) 224–231.
- [27] D.V. Sheehan, Y. Lecrubier, K. Harnett Sheehan, J. Janavs, E. Weiller, A. Keskiner, et al., The validity of the Mini International Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability, *Eur. Psychiatry* 12 (5) (1997) 232–241.
- [28] S. Gilbody, D. Richards, S. Brealey, C. Hewitt, Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis, *J. Gen. Intern. Med.* 22 (11) (2007) 1596–1602.
- [29] K.M. Fiest, S.B. Patten, S. Wiebe, A.G. Bulloch, C.J. Maxwell, N. Jetté, Validating screening tools for depression in epilepsy, *Epilepsia* 55 (10) (2014) 1642–1650.
- [30] A.S. Moriarty, S. Gilbody, D. McMillan, L. Manea, Screening and case finding for major depressive disorder using the Patient Health Questionnaire (PHQ-9): a meta-analysis, *Gen. Hosp. Psychiatry* 37 (6) (2015) 567–576.

- [31] D. McMillan, S. Gilbody, D. Richards, Defining successful treatment outcome in depression using the PHQ-9: a comparison of methods, *J. Affect. Disord.* 127 (1–3) (2010) 122–129.
- [32] B. Levis, Y. Sun, C. He, Y. Wu, A. Krishnan, P.M. Bhandari, et al., Accuracy of the PHQ-2 alone and in combination with the PHQ-9 for screening to detect major depression: systematic review and meta-analysis, *Jama.* 323 (22) (2020) 2290–2300.
- [33] R.L. Spitzer, K. Kroenke, J.B. Williams, B. Löwe, A brief measure for assessing generalized anxiety disorder: the GAD-7, *Arch. Intern. Med.* 166 (10) (2006) 1092–1097.
- [34] J. Connell, M. Barkham, CORE-10 User Manual, version 1.1, CORE System Trust & CORE Information Management System Ltd, 2007.
- [35] A.J. Carson, J. Stone, C.H. Hansen, R. Duncan, J. Cavanagh, K. Matthews, et al., Somatic symptom count scores do not identify patients with symptoms unexplained by disease: a prospective cohort study of neurology outpatients, *J. Neurol. Neurosurg. Psychiatry* 86 (3) (2015) 295–301.
- [36] T. Lempert, M. Dieterich, D. Huppert, T. Brandt, Psychogenic disorders in neurology: frequency and clinical spectrum, *Acta Neurol. Scand.* 82 (5) (1990) 335–340.
- [37] R.L. Spitzer, J.B.W. Williams, K. Kroenke, M. Linzer, F.V. De Gruy, Hahn S.R. III, et al., Utility of a new procedure for diagnosing mental disorders in primary care: the PRIME-MD 1000 study, *JAMA* 272 (22) (1994) 1749–1756.
- [38] K.A. Rimes, T. Chalder, The beliefs about emotions scale: validity, reliability and sensitivity to change, *J. Psychosom. Res.* 68 (3) (2010) 285–292.
- [39] J.C. Mundt, I.M. Marks, M.K. Shear, J.H. Greist, The work and social adjustment scale: a simple measure of impairment in functioning, *Br. J. Psychiatry* 180 (2002) 461–464.
- [40] J. Ware Jr., M. Kosinski, S.D. Keller, A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity, *Med. Care* 34 (3) (1996) 220–233.
- [41] N.C. Cheak-Zamora, K.W. Wyrwich, T.D. McBride, Reliability and validity of the SF-12v2 in the medical expenditure panel survey, *Qual. Life Res.* 18 (6) (2009) 727–735.
- [42] M. Herdman, C. Gudex, A. Lloyd, M. Janssen, P. Kind, D. Parkin, et al., Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L), *Qual. Life Res.* 20 (10) (2011) 1727–1736.
- [43] B.L. Conner-Spady, D.A. Marshall, E. Bohm, M.J. Dunbar, L. Loucks, A. Al Khudairy, et al., Reliability and validity of the EQ-5D-5L compared to the EQ-5D-3L in patients with osteoarthritis referred for hip and knee replacement, *Qual. Life Res.* 24 (7) (2015) 1775–1784.
- [44] M.F. Janssen, E. Birnie, J.A. Haagsma, G.J. Bonsel, Comparing the standard EQ-5D three-level system with a five-level version, *Value Health* 11 (2) (2008) 275–284.
- [45] B. Muthén, S. du Toit, D. Spisic, Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes, *Psychometrika* 75 (1) (1997) 40–45.
- [46] T. Kyriazos, Applied psychometrics: sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general, *Psychology* 9 (2018) 2207–2230.
- [47] J.W. Hoelter, The analysis of covariance structures: goodness-of-fit indices, *Sociol. Methods Res.* 11 (3) (1983) 325–344.
- [48] L-t Hu, P.M. Bentler, Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives, *Struct. Equ. Model. Multidiscip. J.* 6 (1) (1999) 1–55.
- [49] P.M. Bentler, Comparative fit indexes in structural models, *Psychol. Bull.* 107 (2) (1990) 238–246.
- [50] L.K. Muthén, B.O. Muthén, *Mplus User's Guide*, 6th ed., Muthén & Muthén, Los Angeles, CA, 1998–2013.
- [51] L. Guttman, Some necessary conditions for common-factor analysis, *Psychometrika* 19 (2) (1954) 149–161.
- [52] H.F. Kaiser, The application of electronic computers to factor analysis, *Educ. Psychol. Meas.* 20 (1960) 141–151.
- [53] J.L. Horn, A rationale and test for the number of factors in factor analysis, *Psychometrika* 30 (2) (1965) 179–185.
- [54] Presaghi F, Desimoni M. *random.polychor.pa: A Parallel Analysis With Polychoric Correlation Matrices. R package version 1.1.4-3.* 2019.
- [55] R.B. Cattell, The scree test for the number of factors, *Multivariate Behav. Res.* 1 (2) (1966) 245–276.
- [56] S.P. Reise, J. Morizot, R.D. Hays, The role of the bifactor model in resolving dimensionality issues in health outcomes measures, *Qual. Life Res.* 16 (Suppl. 1) (2007) 19–31.
- [57] B. Muthén, A structural probit model with latent variables, *J. Am. Stat. Assoc.* 74 (368) (1979) 807–811.
- [58] J.N. Mandrekar, Receiver operating characteristic curve in diagnostic test assessment, *J. Thorac. Oncol.* 5 (9) (2010) 1315–1316.
- [59] L.J. Cronbach, Coefficient alpha and the internal structure of tests, *Psychometrika* 16 (3) (1951) 297–334.
- [60] R. Dixit, A. Popescu, A. Bagić, G. Ghearing, R. Hendrickson, Medical comorbidities in patients with psychogenic nonepileptic spells (PNES) referred for video-EEG monitoring, *Epilepsy Behav.* 28 (2) (2013) 137–140.
- [61] B. Löwe, R.L. Spitzer, J.B. Williams, M. Mussell, D. Schellberg, K. Kroenke, Depression, anxiety and somatization in primary care: syndrome overlap and functional impairment, *Gen. Hosp. Psychiatry* 30 (3) (2008) 191–199.
- [62] S. Moussavi, S. Chatterji, E. Verdes, A. Tandon, V. Patel, B. Ustun, Depression, chronic diseases, and decrements in health: results from the world health surveys, *Lancet* 370 (9590) (2007) 851–858.
- [63] J.R. Read, L. Sharpe, M. Modini, B.F. Dear, Multimorbidity and depression: a systematic review and meta-analysis, *J. Affect Disord.* 221 (2017) 36–46.
- [64] R.-D. Kocalevent, A. Hinz, E. Brähler, Standardization of a screening instrument (PHQ-15) for somatization syndromes in the general population, *BMC Psychiatry* 13 (1) (2013) 91.
- [65] R. Leonhart, L. de Vroege, L. Zhang, Y. Liu, Z. Dong, R. Schaefer, et al., Comparison of the factor structure of the Patient Health Questionnaire for Somatic Symptoms (PHQ-15) in Germany, the Netherlands, and China. A Transcultural Structural Equation Modeling (SEM) Study, *Front Psychiatry* 9 (240) (2018).
- [66] B. Novakova, S. Howlett, R. Baker, M. Reuber, Emotion processing and psychogenic non-epileptic seizures: a cross-sectional comparison of patients and healthy controls, *Seizure* 29 (2015) 4–10.
- [67] C. Beard, K.-J. Hsu, L.S. Rifkin, A.B. Busch, T. Björgvinsson, Validation of the PHQ-9 in a psychiatric sample, *J. Affect. Disord.* 193 (2016) 267–273.
- [68] B.T. Keum, M.J. Miller, K.K. Inkelas, Testing the factor structure and measurement invariance of the PHQ-9 across racially diverse U.S. college students, *Psychol. Assess.* 30 (8) (2018) 1096–1106.
- [69] Kocalevent RD, Hinz A, Brähler E. Standardization of the depression screener patient health questionnaire (PHQ-9) in the general population. *Gen. Hosp. Psychiatry* 2013b;35(5):551–5.
- [70] T.A. Ryan, A. Bailey, P. Fearon, J. King, Factorial invariance of the patient health questionnaire and generalized anxiety disorder questionnaire, *Br. J. Clin. Psychol.* 52 (4) (2013) 438–449.
- [71] T. Inoue, T. Tanaka, S. Nakagawa, Y. Nakato, R. Kameyama, S. Boku, et al., Utility and limitations of PHQ-9 in a clinic specializing in psychiatric care, *BMC Psychiatry* 12 (2012) 73.
- [72] L. Manea, S. Gilbody, D. McMillan, Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis, *Can. Med. Assoc. J.* 184 (3) (2012) 191.
- [73] C. He, B. Levis, K.E. Riehm, N. Saadat, A.W. Levis, M. Azar, 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.* 89 (1) (2020) 25–37.
- [74] B. Levis, D. McMillan, Y. Sun, C. He, D.B. Rice, A. Krishnan, et al., Comparison of major depression diagnostic classification probability using the SCID, CIDI, and MINI diagnostic interviews among women in pregnancy or postpartum: an individual participant data meta-analysis, *Int. J. Methods Psychiatr. Res.* 28 (4) (2019), e1803.
- [75] Y. Wu, B. Levis, J.P.A. Ioannidis, A. Benedetti, B.D. Thombs, Probability of major depression classification based on the SCID, CIDI, and MINI diagnostic interviews: a synthesis of three individual participant data meta-analyses, *Psychother. Psychosom.* (2020) 1–13.
- [76] Y. Wu, B. Levis, Y. Sun, A. Krishnan, C. He, K.E. Riehm, et al., Probability of major depression diagnostic classification based on the SCID, CIDI and MINI diagnostic interviews controlling for hospital anxiety and depression scale - depression subscale scores: an individual participant data meta-analysis of 73 primary studies, *J. Psychosom. Res.* 129 (2020) 109892.
- [77] D.C. Rettew, A.D. Lynch, T.M. Achenbach, L. Dumenci, M.Y. Ivanova, Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews, *Int. J. Methods Psychiatr. Res.* 18 (3) (2009) 169–184.