VALUE IN HEALTH 15 (2012) 1141–1147

Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Unidimensional 12-Item Zarit Caregiver Burden Interview for the Assessment of Dementia Caregivers' Burden Obtained by Item Response Theory

Javier Ballesteros, MD, PhD[1,2,*], *Borja Santos, BSc Maths*[2], *Eduardo González-Fraile, MSc Psychol*[1,3], *Paula Muñoz-Hermoso, BSc Psychol*[3,4], *Ana I. Domínguez-Panchón, MSc Psychol*[3,5], *Manuel Martín-Carrasco, MD, PhD*[2,3]

[1]*Department of Neuroscience – Psychiatry, Medical School, University of the Basque Country, UPV/EHU, Leioa, Spain;* [2]*CIBERSAM, Madrid, Spain;* [3]*Instituto de Investigaciones Psiquiátricas – IIP, Bilbao, Spain;* [4]*Clínica Psiquiátrica Padre Menni, Pamplona, Spain;* [5]*Hospital Aita-Menni, Arrasate-Mondragón, Spain*

ABSTRACT

Objectives: To develop a one-dimensional version of the 22-item Zarit Caregiver Burden Interview (ZBI) by applying item response theory approaches. **Methods:** The answers to the 22-item ZBI of 241 caregivers participating in a clinical trial were analyzed 1) with a Mokken nonparametric item response theory analysis to ascertain the dimensional structure underlying the scale and obtain a one-dimensional reduced version, 2) with the Samejima's graded response model to assess the item characteristics of the reduced version, and 3) with confirmatory factor analysis to confirm the unidimensionality of the reduced ZBI version and assess the item loadings to the burden latent variable. **Results:** Mokken analysis resulted in a major one-dimensional scale comprising 12 items directly related with burden. All items showed scalability indices over 0.30. The scalability for the overall scale was 0.44 defining a medium scale according to Mokken's criteria. An unconstrained Samejima's graded response model showed appropriate fit, and most items of the reduced 12-item ZBI presented pertinent difficulty and discrimination parameters. The results of the 12-item ZBI confirmatory factor analysis fitted to a one-dimensional latent structure for burden (comparative fit index = 0.975; root-mean-square error of approximation = 0.067; weighted root mean square residual = 0.677). All factor lodgings were above 0.40 with items 9 (strained by the relative) and 22 (overall feeling of burden) presenting the highest loadings. **Conclusions:** The reduced 12-item ZBI fits a one-dimensional latent variable of burden. Further psychometric studies, focusing on its equivalence for different populations, sensitivity to change, and minimal important difference are warranted.

Keywords: confirmatory factor analysis, item response theory, Mokken analysis, Samejima's graded response model, Zarit Caregiver Burden Interview.

Introduction

Originally, the Zarit Caregiver Burden Interview (ZBI) was developed in 1980 as a 29-item self-report scale aimed to assess the subjective burden experienced by an informal (not paid) caregiver, usually a relative [1]. Some years later, shorter versions of 20 and 22 items were released [2,3], and in 1991, the standard version of 22 items with two factorial subscales— Personal Strain and Role Strain—was produced [4]. While in the first version items were scored on a four-point ordinal Likert-type scale, a five-point ordinal scale (0: never; 1: rarely; 2: sometimes, 3: quite frequently; and 4: nearly always) recording the feeling the caregiver has on the corresponding statement is currently used. Despite the two factorial subscales mentioned above, the ZBI has been almost universally used as a one-dimensional measure given the high correlation between factors originally described. Later on, several attempts to find latent dimensions of the ZBI have been successfully made [5–7]. It has been used primarily, but not exclusively, among caregivers of patients diagnosed with dementia [8–10], and caregivers of patients presenting with other pathologies or settings including palliative care, heart failure, brain injury, or schizophrenia [11–14] have also been assessed. Nowadays, the ZBI is believed to be the most commonly used measure of caregivers' burden [15].

Because of its multidimensionality, the ZBI total score discloses several underlying latent constructs with two to five factors as previously reported [7,16]. The multidimensional structure of the ZBI implies that a clinical interpretation based on its total score could not be as informative as it should be because of mixing different latent constructs in a unique observed score. Consequently, if the 22-item ZBI is used as primary outcome in trials designed to evaluate the efficacy of interventions to improve caregiver burden, and an absolute change since baseline is reported for its total score, it could be

* *Address Correspondence to:* Javier Ballesteros, Department of Neuroscience – Psychiatry, Medical School, University of the Basque Country, UPV/EHU, Barrio Sarriena S/N, E-48640 Leioa, Spain.

E-mail: javier.ballesteros@ehu.es.

1098-3015 - see front matter

http://dx.doi.org/10.1016/j.jval.2012.07.005

unclear whether some dimensions have changed differentially over the intervention process, and this is the information that could be relevant to properly address the trial aims.

Factor analysis studies of the ZBI have been mostly done to verify or explore the underlying structure of the original ZBI. Nonetheless, others were intended to develop briefer or screening versions of the ZBI, and several reduced versions of the canonical 22-item ZBI with a variable number of items have been reported in the literature [17–19]. None of those versions, however, were developed or assessed by using psychometric approaches rooted on item response theory (IRT). IRT is less concerned than classical test theory with reliability of total scores. On the contrary, it is more concerned with the analysis of the responses elicited to individual items of the evaluated scale. Benefits of IRT include comprehensive analysis and reduction of measurement error, meaningful scaling of latent variables, objective calibration and equating, evaluation of test and item bias, greater accuracy in the assessment of change due to therapeutic interventions, and evaluation of model and person fit [20]. IRT models use item endorsement frequencies as outcomes to estimate parameters that characterize the properties of an item and are increasingly used to improve the accuracy of classical psychometric tools [21,22] or to develop shortened versions [23]. Our aim in this study was to obtain a reduced and one-dimensional version of the 22-item ZBI by using both nonparametric and parametric IRT analyses of the baseline measurements of caregivers of dementia patients recruited for a randomized clinical trial designed to assess the efficacy of a psychoeducational program on the caregivers' burden (EDUCA-2 trial; ISRCTN14411440).

Methods

Study Design and Population

This study includes a validation sample of 241 caregivers of patients with dementia recruited within a multicenter randomized clinical trial (20 centers across Spain and Portugal). To be included, a caregiver should be informally caring (not paid for) for a patient with dementia (*Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision*, criteria). The patient should be treated as an outpatient in memory clinics or psychogeriatric day centers at the research sites, to present impairment of at least two instrumental activities, or one activity of daily life. The caregiver (adult male or female) should care for the patient at least 4 hours daily. The recruitment for the trial began in September 2010 and lasted till October 2010.

Data Collection Procedures and Measures

The EDUCA-2 trial included three evaluations: at baseline (visit 1), after finishing the trial intervention (visit 2 at 4 months since inception), and finally at 4 months since finishing the trial intervention (visit 3 at 8 months since inception). This article reports results of the 22-item ZBI as obtained at trial inception (prerandomization period). The classical psychometric properties of the Spanish and Portuguese versions of the ZBI used in this study have been published elsewhere [24–26] and showed appropriate internal consistency (Cronbach's α values of 0.92 and 0.88, respectively) and test-retest reliability (intraclass correlation coefficient [ICC] of 0.93 for the Portuguese ZBI).

Statistical Analysis

The analyses to obtain a reduced and one-dimensional ZBI version were conducted according to a three-step approach.

First, the 22-item responses were analyzed by a nonparametric IRT analysis (Mokken analysis) to elucidate the latent constructs and likely subscales underlying the association matrix of observed responses [27,28]. A secondary aim was to check in advance the psychometric assumptions associated with the parametric IRT model chosen for the second analytical step: the Samejima's graded response model (GRM) [29]. Two Mokken models were fitted to the data—the monotone homogeneity model and the double monotonicity model. The former aims to test whether a scale total score is a valid tool for ordering and classifying subjects according to the degree of the construct exhibited. The later is more restrictive because it also aims to identify whether an order exists among the items to rate the corresponding construct that is independent of the selected sample. All items linked to underlying Mokken scales were retained if the scales had at least three items attached to them. The Mokken constructed scales were interpreted according to customary rules of thumb: to be considered as relevant, all items should have a scalability coefficient (H_i) ≥ 0.30, and also the total scale should have a scalability (H) of ≥ 0.30. Mokken [27] suggested the following thresholds to interpret scalability coefficients for a measurement scale: weak scale for $0.3 \leq H < 0.4$, medium scale for $0.4 \leq H < 0.5$, and strong scale for $H \geq 0.5$.

In the second step, we used the Samejima's GRM as the parametric IRT to obtain estimates of the relationship among the latent construct and the item characteristics. Specifically, we estimated the item response characteristic curve parameters (ICC) and item information. If items behave adequately, the ICCs should present an ordered shape discriminating among the category thresholds. Even if overlapping, each category within an item should present a distinct probability of being selected more than any other category for a specific difficulty. We adjusted two GRMs, one assuming equal discrimination among items (restricted model) and other relaxing such assumption (unrestricted model). Because the restricted GRM is nested within the unrestricted model, we selected the most parsimonious model according to the likelihood ratio test.

The third and final step was to assess the unidimensionality of the reduced scale so far obtained by confirmatory factor analysis (CFA). CFA was carried out by using robust weighted least squares on the sample variance-covariance matrix of polychoric correlations among the reduced ZBI items. Goodness of fit for the CFA was evaluated by using the comparative fit index (CFI), the root-mean-square error of approximation (RMSEA), and the weighted root mean square residual. A value of CFI > 0.95 was considered as acceptable model fit, an RMSEA value of < 0.08 was considered to reflect an adequate fit to the model, and a value of < 0.05 was considered as good fit. A weighted root mean square residual value of less than 1 is customarily interpreted as a good value, but its behavior as a goodness-of-fit index is not as well studied as are the CFI and RMSEA indexes. Finally, the reliability of the final reduced scale was evaluated according to both Mokken and Cronbach estimates. The statistical packages R v2.13.1 (nonparametric and parametric IRT with libraries "mokken" and "ltm," respectively) [30,31] and Mplus v5 (CFA) were used to carry on the analyses.

Results

Table 1 shows the main characteristics of the dyads of caregivers and patients with dementia. As seen, patients were equally distributed in the sample regarding cognitive severity. All caregivers shared home with the patient, were informal (not paid for caring for the patient), and patient's relatives (spouses 50.4%, sons/daughters 44.3%, brothers/sisters and nephews 5.3%). They were mostly females, with a caring time well beyond 8 h/d.