Journal of Affective Disorders 388 (2025) 119744

Contents lists available at ScienceDirect

Journal of Affective Disorders

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

Review article

Internal consistency, sensitivity, specificity and predictive value of the short version of the Zarit Burden Interview-7 scores: systematic review and generalization meta-analysis

Júlia Gisbert-Pérez [a], Elena Cejalvo [b], Manuel Martí-Vilar [b], Laura Badenes-Ribera [a,*]

[a] Department of Methodology of Behavioral Sciences, Faculty of Psychology and Speech Therapy, Universitat de València, Spain
[b] Department of Basic Psychology, Faculty of Psychology and Speech Therapy, Universitat de València, Spain

ARTICLE INFO

Keywords:
ZBI-7
Reliability
Validity
meta-analysis

ABSTRACT

Objectives: The *Zarit Burden Interview* (ZBI) is one of the most widely used instruments for assessing caregiver burden. One of the most used short scales of the ZBI is the 7-item version. The aim of the study was to conduct a systematic review and generalization meta-analysis of the reliability and validity of the scores of the short version ZBI-7 items to determine whether they provide average values that justify their application and, secondly, to identify the characteristics of the studies that could explain the heterogeneity among the estimates.
Methods: Forty-eight articles were identified that applied the short version of the ZBI-7, of which only 14 reported a reliability coefficient and 7 reported sensitivity, specificity and predictive values. A random effects model was applied in the statistical analyses.
Results: The average Cronbach's alpha coefficient was 0.866 (95 % CI [0.835, 0.891], k = 15), with high heterogeneity (I^2 = 86.4 %). Mean sensitivity was 98.6 % (95 % CI [70.0 %–95.3 %], k = 3), mean specificity was 87.4 % (95 % CI [93.6 %–99.7 %], k = 3), mean positive predictive value was 93.7 % (95 % CI [81.3 %–98.1 %], k = 3), mean negative predictive value was 97.3 % (95 % CI [87.5, 99.4], k = 3).
Conclusions: According to psychometric theory, the ZBI-7 can be considered reliable and valid for exploratory purposes in both research and clinical practice.

1. Introduction

In contemporary society, the aging population and the increasing prevalence of chronic diseases have become major concerns (Atella et al., 2019). As a result, healthcare is no longer limited to hospital settings. It is also provided at different levels, including in family settings (Kokorelias et al., 2019). Informal caregivers have become essential in supporting people with chronic conditions or palliative care (Chong et al., 2022). Such carers assist with both personal and instrumental activities of daily living. However, the demands of caregiving can lead to caregiver burden, which can negatively affect physical and emotional well-being (Chang et al., 2009).

Caregiver burden is defined as "the level of multifaceted strain perceived by the caregiver as a result of caring for a family member and/or loved one over time" (Liu et al., 2020). This burden includes an objective dimension, which refers to tangible changes such as the number of caregiving hours or specific tasks performed, and a subjective dimension, which refers to the emotional distress experienced in the caregiving role (Fekete et al., 2017). The assessment of caregiver burden is essential for public health and clinical intervention, as its negative consequences potentially compromise the quality of care provided, the quality of life, and the physical and psychological health of the caregiver (Liu et al., 2020). Furthermore, assessing levels of caregiver burden can foster the development of prevention strategies and interventions to reduce caregivers' psychosocial distress (Tyo and McCurry, 2020).

There are multiple instruments for assessing caregiver burden. A systematic review by Tu et al. (2022) examined caregiver burden questionnaires in dementia settings and identified 35 self-reported instruments, 15 of which were analyzed. The authors highlighted four instruments for assessing caregiver burden due to their adequate psychometric properties and analysis of objective and subjective dimensions of burden: The Screen for Caregiver Burden (SCB; Vitaliano et al., 1991), the Caregiver Burden Interview (CBI; Marvardi et al., 2005), the Burden Scale for Family Caregivers (BSFC; Gräßel et al.,

* Corresponding author at: Facultat de Psicologia i Logopèdia (Universitat de València), Avda. Blasco Ibáñez, 21, València 46010, Spain.
E-mail address: laura.badenes@uv.es (L. Badenes-Ribera).

https://doi.org/10.1016/j.jad.2025.119744
Received 12 September 2024; Received in revised form 21 May 2025; Accepted 19 June 2025
Available online 22 June 2025

2003), and the Zarit Burden Interview (ZBI; Zarit et al., 1980). The SCB and BSFC are dementia-specific, whereas the CBI and ZBI offer more general approaches. Tu et al. (2022) concluded that the ZBI was the most suitable instrument for assessing caregiver burden in informal caregivers due to its multiple versions, translations, and multidimensional burden analysis.

The ZBI (Zarit et al., 1980) is one of the most widely used measurement instruments to assess caregiver burden. In its original version, the questionnaire consisted of 28 items (Zarit et al., 1980) but was later reduced to 22 items (Zarit et al., 1985). The items are answered on a five-point Likert-type response scale, where 0 = "never" and 4 = "almost always", whereby higher scores reflect higher caregiving burden. The 22-item Zarit scale, originally developed for use with informal caregivers of people with Alzheimer's disease living in communities, has been administered to a wide range of patient populations, and formal caregivers, and has been translated into numerous languages. In addition, to facilitate faster administration in clinical settings, several short versions of the ZBI scale have been developed (Zarit et al., 1985), ranging from single-item to 18-item. One of the most widely used short-scales is the 7-item version developed by Gort et al. (2005).

A reliability generalization meta-analysis of the ZBI-22 items has suggested that the scale is reliable in the various contexts it has been used (Bachner and O'Rourke, 2007). The average internal consistency estimate measured by Cronbach's alpha coefficient was $\alpha = 0.86$, with a range between $\alpha = 0.62$ and $\alpha = 0.95$, suggesting considerable variability. Cronbach's alpha coefficient values above 0.69 are generally considered adequate (Nunnally and Bernstein, 1994). However, five studies using some short version of the ZBI reported reliability estimates below this threshold (Bachner and O'Rourke, 2007). Based on reliability generalization meta-analysis of caregiver burden instruments, variables that could explain reliability variability between studies include cultural and linguistic aspects and differences in characteristics of caregivers and cared-for persons in the samples (Bachner and O'Rourke, 2007; Cejalvo et al., 2025). Likewise, differences in study characteristics and conditions may also explain differences in reliability between studies (Gronlund and Linn, 1990; Sánchez-Meca et al., 2012).

1.1. Aim of the study

Accurately assessing caregiver burden is crucial for identifying those at risk of psychological disorders and evaluating treatment effectiveness. Thus, understanding the reliability and validity of the scores of measurement instruments is essential. Meta-analysis can enhance our understanding of test scores and their factors across different contexts and populations. Within meta-analysis, reliability generalization meta-analysis (RG MA) is a method used within meta-analyses to statistically integrate reliability estimates from various test applications (Badenes-Ribera et al., 2020). This approach enables researchers to calculate an average reliability score based on individual reliability estimates. Additionally, RG MA assesses the variability among the reliability estimates analyzed, which originate from studies and samples with different characteristics. RG MA provides valuable evidence regarding which measurement scales are more reliable and the specific conditions under which they perform best. This information aids in selecting the most suitable instruments for measuring a particular construct.

This study used a generalization meta-analysis approach to examine the psychometric properties of the short version of the ZBI-7 in terms of reliability and validity of scores with the following objectives (Badenes-Ribera et al., 2020): (1) to estimate the average reliability of the scores of the short version of the ZBI-7; (2) to estimate the variability between reliability estimates; (3) to estimate the average validity in terms of sensitivity, specificity and predictive value of the short version of the ZBI-7; (4) to estimate the variability between sensitivity, specificity and predictive value estimates; (5) to identify substantive and methodological characteristics of the studies that can be statistically associated with reliability coefficients and/or validity estimates. These results will guide researchers and health professionals on the contexts and circumstances in which ZBI-7 scores (Gort et al., 2005) demonstrate better psychometric properties for assessing caregiver burden.

2. Materials and methods

The study followed PRISMA guidelines for systematic reviews and meta-analyses (Page et al., 2021), as well as REGEMA guidelines for reliability generalization meta-analyses (Sánchez-Meca et al., 2021).

2.1. Eligibility criteria

For inclusion in the systematic review, studies had to meet the following criteria: (1) be an original, quantitative study published in a peer-reviewed journal, (2) administer the ZBI-7 with a 5-point response format, (3) use any type of target population. No limits were set on study dates and language, geographic location, or age of subjects. Exclusion criteria were as follows: (1) single case studies and (2) use of other versions of the Zarit.

For inclusion in the meta-analysis of reliability and validity generalizability studies, in addition to the above criteria, studies had to meet the following criteria: (1) report some reliability (e.g. Cronbach's alpha, McDonald's Omega, test-retest, etc.) or validity (sensitivity, specificity or predictive values) coefficient calculated on the specific study sample; (2) or report the necessary data that would allow such estimates to be calculated.

2.2. Search strategy

Online searches were conducted in February 2024 in the Web of Science, Pubmed, Scopus, and PsycInfo databases. The following terms were used: "Zarit 7", "ZBI 7", "Zarit-7" and "ZBI-7". In addition, Google Scholar articles citing the Gort et al. (2005) study were reviewed, and the references of the identified studies were analyzed to determine whether any met the inclusion criteria.

2.3. Study selection

After removing duplicate entries using RefWorks, two reviewers independently conducted a blind peer review based on title, abstract, and full text following eligibility criteria. A third reviewer was consulted for discrepancies. For longitudinal studies or studies with multiple measurements on the same subjects, the first study reporting reliability or validity estimates was chosen, unless an exceptional circumstance dictated otherwise (e.g., no reliability estimates were provided in the first study). Studies reporting reliability or validity estimates separately for each sample were considered independent samples.

2.4. Data extraction

Two independent reviewers systematically applied a protocol to extract the following characteristics of the studies: (1) authors, (2) year of publication, (3) first author affiliation, (4) study focus (psychometric or applied), (4) funding, (5) study language, (6) country of origin, (7) study design I (cross-sectional or longitudinal), (8) study design II (experimental or non-experimental), (9) sampling method (probability or non-probability), (10) sample size, (11) population background, (12) type of caregiver (formal or informal), (13) characteristics of the cared-for person, (14) mean age of caregiver, (15) standard deviation of caregiver age, (16) gender (percentage of females), (17) percentage of Caucasians, (18) educational level, (19) mean scale scores, (20) standard deviation of scale scores, (21) sensitivity, (22) specificity, (23) positive predictive value, (24) negative predictive value, (25) reported reliability estimates.

2.5. Statistical analysis

Statistical analyses were performed with the metafor package in R (Viechtbauer, 2010), and all were interpreted assuming a significance level of 5 % ($p \leq .050$) using 2-tailed tests. A reliability generalization meta-analysis was performed on Cronbach's alpha coefficients obtained for the total scores of the short version ZBI-7 scores (Gort et al., 2005). To normalize their distributions and stabilize their variances, the reported Cronbach's alpha coefficients were transformed by applying the formula proposed by Bonett (2002). To enhance the interpretation of the meta-analysis results, the average and individual reliability estimates, along with their confidence intervals, were converted to Cronbach's alpha coefficient metric.

A meta-analysis was also conducted to assess the validity of ZBI-7 scores compared to the original version of the Zarit, in terms of sensitivity, specificity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV) as measures of agreement between the two versions. These metrics are useful for evaluating how well a diagnostic instrument can accurately predict the presence (PPV) or absence (NPV) of a condition. According to standard meta-analytic methods (Borenstein et al., 2021), the concordance measures (proportions) of the individual studies were transformed to the logit metric. The average logit and individual logit, along with their confidence intervals, were then converted back to the proportions metric to make the meta-analysis results easier to interpret.

In all the meta-analyses conducted, a random-effects model was used. This model was chosen because of the significant variability observed in the characteristics of the studies analyzed. Additionally, it is generally recommended to generalize results to future studies that may not share similar characteristics (Schmidt et al., 2008). Estimates of reliability, sensitivity, specificity, and predictive values were weighted by the inverse of the variance (Borenstein et al., 2021), and the restricted maximum likelihood method was used to estimate the between-study variance (τ^2). Average Cronbach's alpha coefficient, average sensitivity, average specificity, and average predictive values with their 95 % confidence intervals (95 % CI) were calculated using the method recommended by Hartung and Knapp (2001).

To assess the robustness of the findings and to determine whether the application of Bonett's transformation substantially affected the meta-analytic estimates, an additional meta-analysis was conducted using untransformed Cronbach's alpha coefficients. Furthermore, outlier and influence diagnostics were performed to evaluate the potential impact of extreme or influential values on the meta-analytic results (Viechtbauer and Cheung, 2010; Viechtbauer, 2021). First, studentized residuals were computed for each Cronbach's alpha coefficient, sensitivity, specificity, and predictive value. Estimates with studentized residuals exceeding ±1.96 were considered potential outliers. Subsequently, case-deletion diagnostics were conducted to examine the influence of each estimate on both the overall and individual parameter estimates. Lastly, the potential impact of publication bias was evaluated using the Egger's test (Egger et al., 1997), and the trim-and-fill method (Duval and Tweedie, 2000).

To assess heterogeneity, we calculated Cochran's Q statistic and the I^2 index. A Q-statistic with $p \leq .05$ indicated heterogeneity among reliability estimates. The degree of this heterogeneity was interpreted using the I^2 index, which can be classified as follows (Higgins and Thompson, 2002; Higgins et al., 2003; Huedo-Medina et al., 2006): no heterogeneity (<25 %), low heterogeneity (25–49.9 %), moderate heterogeneity (50–74.9 %), and high heterogeneity (>75 %).

Given the heterogeneity observed among Cronbach's alpha coefficients, moderator analyses were conducted to identify study characteristics associated with reliability estimates. Simple meta-regressions were applied for quantitative moderator variables and weighted ANOVAs for categorical moderator variables; in both cases, the improved method proposed by Knapp and Hartung (2003) was applied, and a mixed-effects model was assumed (López-López et al., 2013; Viechtbauer et al., 2015). In addition, the R^2 index was calculated to estimate the proportion of variance explained by each moderator variable (López-López et al., 2014; Raudenbush, 2009).

3. Results

3.1. Study selection process

Fig. 1 shows the study selection process. Out of the 233 references initially identified, they were imported into RefWorks to be screened for duplicates and to review titles. After reviewing the abstracts, 83 articles were found to be eligible. Among these, 12 duplicates were identified and removed. Out of the remaining 71 articles, 48 used the ZBI-7 (Gort et al., 2005). Within these 48 articles, 18 were included in the meta-analysis because they reported one or more estimated reliability coefficients (N = 14) and/or provided estimates of sensitivity and specificity (n = 7). The other 30 articles did not report these estimates.

3.2. Characteristics of the selected studies

The total number of participants in the sample was N = 9907 (minimum = 13, maximum = 1618), with a mean of 194 participants per sample (SD = 249; see Table A1). The mean age of the participants was 57.58 years (SD = 5.74) and the mean percentage of females was 72.64 % (SD = 9.68 %). Only one study reported the percentage of Caucasians in their sample (75.6 %; Starr et al., 2022). The mean of the total scores of the ZBI-7 was 13.36 (Median = 13, SD = 7.03). Of the 48 articles, considering the 51 independent samples, 37 (63.6 %) appeared in articles written in English and the remaining 14 samples (27.4 %) were published in Spanish. Regarding the sample characteristics, 94.1 % of the samples consisted of informal caregivers (n = 48), while the remaining 3 samples included both formal and informal caregivers. The cared-for persons were mostly patients with oncological diseases (n = 10, 19.6 %), in palliative care (n = 8, 15.7 %), with dementia (n = 8, 15.7 %), dependent persons (n = 3, 5.9 %), with schizophrenia (n = 2, 3.9 %) or the older population in general (n = 2, 3.9 %), among others. In addition, several samples had combinations of the above or other difficulties (n = 6, 11.7 %). The studies were conducted in 14 different countries: Spain (26 samples; 51 %), France (3 samples; 5.9 %), Germany (3 samples; 5.9 %), South Korea (3 samples; 5.9 %), UK (3 samples; 5.9 %), Brazil (2 samples; 3.9 %), China (2 samples; 3.9), Colombia (2 samples; 3.9 %) and USA (2 samples; 3.9 %). Chile, Mexico, Portugal, Sweden, and Taiwan contributed one sample (2 %, respectively). 88.2 % (n = 45) of the studies used convenience samples, and most of them had cross-sectional research designs (76.5 %, n = 39). Also, the focus of the studies was mostly applied (68.6 %, n = 35), with a non-experimental design (86.3 %, n = 44).

3.3. Internal consistency

In the study, out of 48 articles that used ZBI-7, only 14 reported a reliability coefficient (29.2 %). Most of them (n = 13, 27.1 %) used Cronbach's alpha coefficient for measuring internal consistency. Two articles used McDonald's Omega reliability coefficient, and one study reported a test-retest reliability estimate (Breinbauer et al., 2009, r = 0.93, time interval = 4 months). As a result, the meta-analysis focused only on Cronbach's alpha coefficient for the total score of the short version ZBI-7 based on 15 independent samples.

A total of 1993 caregivers (N = 1933) were included in this meta-analysis with a mean of 129 caregivers per sample (Median = 130; SD = 68.59, Min = 32, Max = 270) recruited using a non-probability sampling method. The mean age of the caregivers was 58.04 years (SD = 8.09, Min = 38, Max = 69.60). The mean percentage of women was 74.4 % (Min = 61.6 %, Max = 91.5 %). The studies used a cross-sectional design and were conducted on three continents: Europe, 7 studies (53.8 %); South America, 5 studies (38.5 %) and Asia, 1 study

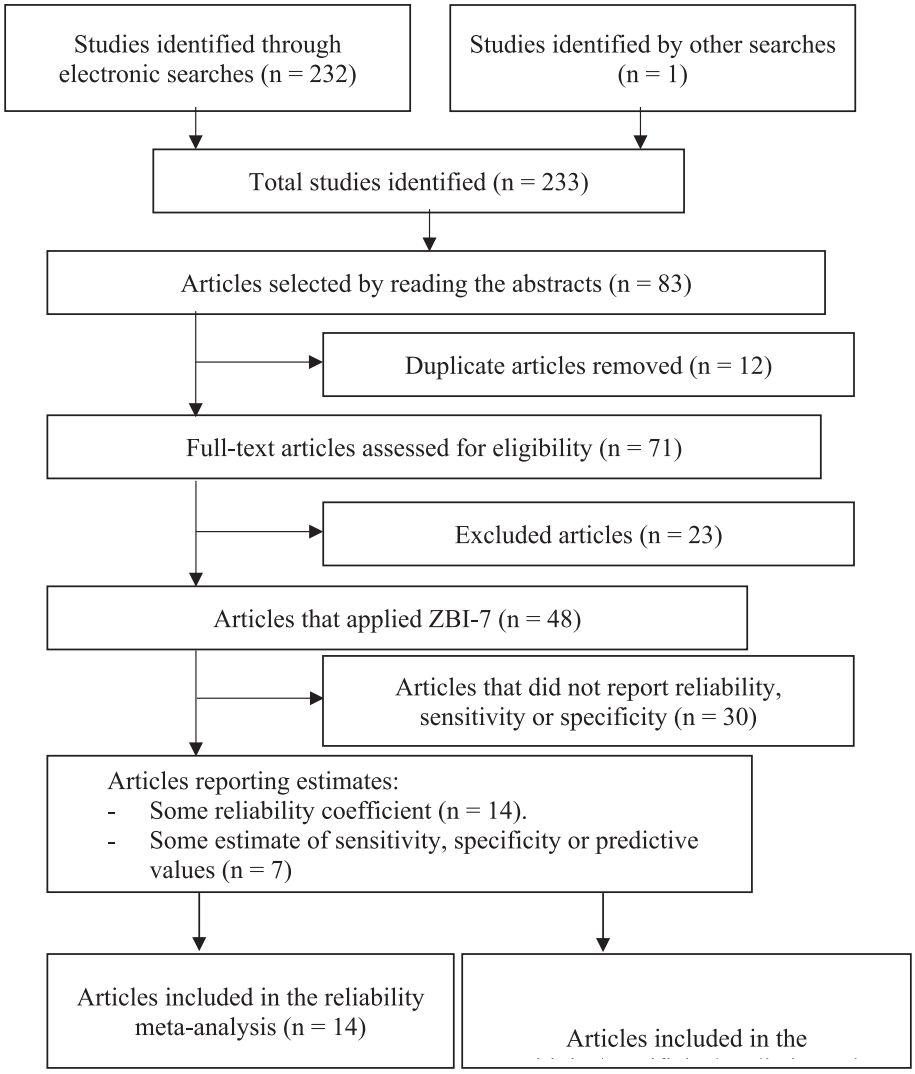

Fig. 1. REGEMA flowchart of the study selection process.

(7.7 %). Furthermore, of the 13 studies, 11 (73.3 %) were written in English and the remaining 4 studies (26.7 %) were written in Spanish. Finally, only one study reported a mean score on ZBI-7 (M = 8, Navarro Jiménez et al., 2022), and a mean standard deviation of the total scores of 6.77 (SD = 0.59 Min. = 5.60, Max. = 7.50).

The fifteen reported Cronbach's alpha coefficients ranged from 0.788 to 0.944 with an average Cronbach's alpha coefficient of 0.866 (95 % CI [0.835, 0.891]) (see Fig. 2). High heterogeneity was observed among the reported Cronbach's alpha coefficients ($Q(14) = 116.821$, $p < .0001$; $I^2 = 86.4$ %).

To assess the impact of using Bonett-transformed coefficients, a separate meta-analysis was conducted using untransformed Cronbach's alpha coefficients. The results were consistent, with the untransformed analysis yielding an average Cronbach's alpha of 0.863 (95 % CI [0.838, 0.888]). These findings suggest that the use of Bonett's transformation did not substantially affect the overall estimate.

Additionally, outlier and influence diagnostic procedures were conducted to assess the potential impact of extreme or influential cases on the meta-analytic estimate. One potential outlier was identified: the study by Hernández-Padilla et al. (2019). Diagnostics for influential cases indicated that this study exerted an influence on the overall meta-analytic estimates. Subsequently, case-deletion analyses (sensitivity analyses) were performed to examine the impact of including or excluding individual Cronbach's alpha coefficients on the average reliability estimate. When each coefficient was removed one at a time, the average Cronbach's alpha coefficient for the ZBI-7 total score ranged from 0.856 (95 % CI [0.831, 0.878]) to 0.869 (95 % CI [0.839, 0.995]) when each Cronbach's alpha coefficient was removed one by one. These results indicate that the reliability estimate was robust across studies.

Finally, Egger's test and the trim-and-fill method were used to assess the potential impact of publication bias on the results of the meta-analysis. Egger's test did not show statistical significance ($t(13) = -1.000$, $b = 2.999$, 95 % CI [1.636, 2.961], $p = .335$), suggesting that publication bias did not significantly impact the results of this meta-analysis. Additionally, the trim-and-fill procedure imputed only three missing reliability estimates to achieve funnel plot symmetry. The adjusted average Cronbach's alpha coefficient for the ZBI-7 total score was 0.877, which is nearly identical to the original unadjusted estimate. This negligible difference reinforces the conclusion that publication bias did not meaningfully influence the meta-analytic findings.

3.4. Sensitivity, specificity, positive predictive value and negative predictive value

Out of the 48 articles that utilized the abbreviated version of the ZBI-7 (Gort et al., 2005), 7 of them investigated the agreement between the scores of the abbreviated ZBI-7 and the original Zarit scale. Among these 7 articles, only 3 provided the raw data of the correlation matrix

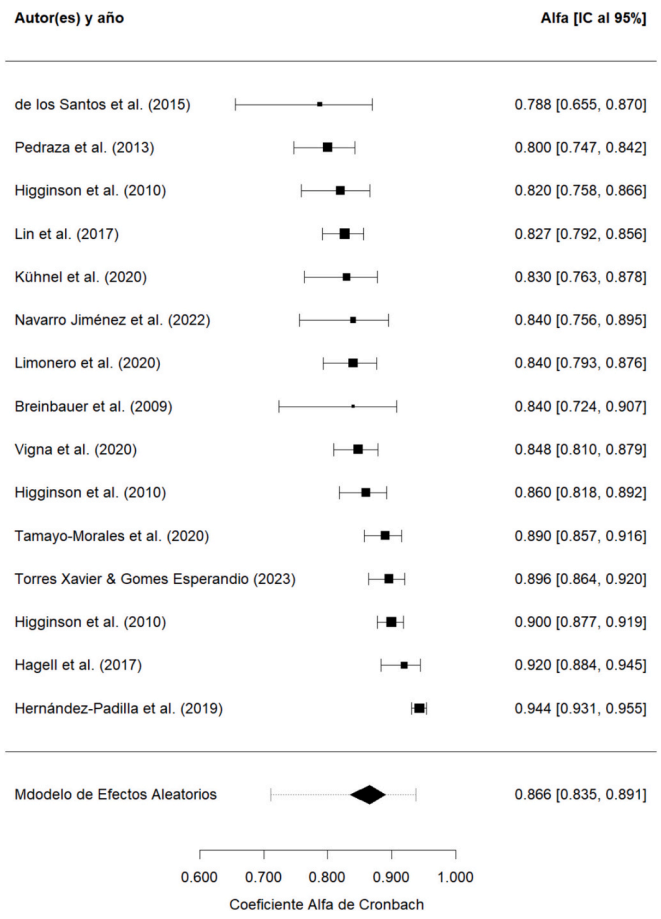

Fig. 2. Meta-analysis of reliability generalization on Cronbach's Alpha coefficients.

(Breinbauer et al., 2009; Gort et al., 2005; Regueiro Martínez et al., 2007), making it possible to conduct a meta-analysis on the data presented in these 3 studies. A total of 156 caregivers participated in these meta-analyses, with an average of 52 caregivers per sample (Median = 61; SD = 17.35, Min = 32, Max = 63). The average age of caregivers was 56.80 years (SD = 1.70, Min = 56.6, Max = 58). The average percentage of female caregivers was 86.5 % (Min. = 83.6 %, Max. = 89.3 %). Two studies were carried out in Spain, and the other study was conducted in Chile. All the studies were written in Spanish.

Among the 7 studies reporting sensitivity values, sensitivity ranged from 37 % (Vélez Lopera et al., 2012) to 100 % (Breinbauer et al., 2009; Gort et al., 2005; Regueiro Martínez et al., 2007). The meta-analysis on sensitivity yielded an average sensitivity of 98.6 % (95 % CI [70.0 %, 95.3 %], N = 156, k = 3), with no heterogeneity between sensitivity values ($Q(2) = 0.073$, $p = .64$, $I^2 = 0$ %). Non potential outliers were identified, as no studentized residuals exceeded ±1.96, and diagnostics for influential cases indicated that no individual studies exerted undue influence. Subsequently, case-deletion analyses (sensitivity analyses) were conducted to examine the strength of these results. When each sensitivity value of the short version of the ZBI-7 was removed one by one and the analyses were rerun, the average sensitivity estimates ranging from 98.5 % (95 % CI [90.3 %, 99.8 %]) to 98.8 % (95 % CI [92.2 %, 99.8 %]), indicating a stable average sensitivity estimate.

In the 7 studies reporting specificity values, specificity ranged from 78 % (Breinbauer et al., 2009) to 100 % (Gort et al., 2005; Vélez Lopera et al., 2012). The meta-analysis yielded an average specificity of 87.4 % (CI 95 % [93.6 %, 99.7 %], N = 156, k = 3), with no heterogeneity observed among the sensitivity values ($Q(2) = 2.993$, $p = .224$, $I^2 = 29.8$ %). No potential outliers were identified, as no studentized residuals exceeded ±1.96. However, diagnostics for influential cases indicated that the study by Breinbauer et al. (2009) exerted some influence on the meta-analytic estimates. Subsequently, case-deletion analyses (sensitivity analyses) were conducted to assess the impact of including or excluding individual specificity values on the average estimate. The results showed that the average specificity ranged from 83.7 % (95 % CI [66.47 %, 93.0 %]) to 92.8 % (95 % CI [77.9 %, 97.9 %]), indicating that the average estimate of diagnostic accuracy remained relatively robust, despite the influence of a single study.

In studies reporting the positive predictive value of the short version of the ZBI-7, it ranged from 87 % (Breinbauer et al., 2009) to 100 % (Vélez Lopera et al., 2012). The meta-analysis yielded an average positive predictive value of 93.7 % (95 % CI [81.3 %, 98.1 %], N = 156, k = 3), with low heterogeneity among the positive predictive values ($Q(2) = 3.793$, $p = .150$, $I^2 = 46.5$ %). Again, no potential outliers were identified, as no studentized residuals exceeded ±1.96. However, diagnostics for influential cases indicated that the study by Breinbauer et al. (2009) exerted some influence on the meta-analytic estimates. Case-deletion analyses (sensitivity analyses) were conducted to assess the impact of including or excluding individual positive predictive values on the average estimate. The results showed that the average positive predictive value ranged from 91.4 % (95 % CI [77.3 %, 97.0 %]) to 96.6 % (95 % CI [88.8 %, 99.0 %], indicating that the average estimate remained relatively robust despite the influence of a single study.

In studies reporting the negative predictive value of the short version of the ZBI-7, it ranged from 92 % (Vélez Lopera et al., 2012) to 100 % (Breinbauer et al., 2009; Regueiro Martínez et al., 2007). The meta-analysis yielded an average negative predictive value of 97.3 % (95 % CI [87.5, 99.4], N = 156, k = 3), with no heterogeneity among negative predictive values ($Q(2) = 0.028$, $p = .986$, $I^2 = 0$ %). Non potential outliers were identified, as no studentized residuals exceeded ±1.96, and diagnostics for influential cases indicated that no individual studies exerted undue influence. Subsequently, case-deletion analyses (sensitivity analyses) were conducted to examine the robustness of these results. When each negative predictive value of the short version of the ZBI-7 was removed one by one and the analyses were rerun, the average positive predictive value ranged from 97.1 % (95 % CI [82.2 %, 99.6 %]) to 97.5 % (95 % CI [84.3 %, 99.6 %], indicating a relatively robust average estimate.

Table 1
Results of simple meta-regressions applied on Cronbach's alpha coefficients for the ZBI-7 total score, taking quantitative moderator variables as predictors.

				IC 95 %					
Predictor variables	k	b_j	EE	LL	UL	F	p	Q_E	R^2
N	15	0.001	0.002	−0.002	0.004	0.683	0.423	115,418***	0.00
% female	14	0.014	0.011	−0.010	0.038	1724	0.214	75,104***	0.00
Age average (years)	13	0.015	0.014	−0.016	0.045	1133	0.310	108,594***	0.00
Year of publication	15	0.017	0.021	−0.029	0.062	0.621	0.448	113,215***	0.00

Note. N = Sample size. k = Number of independent samples. b_j = Regression coefficient for each predictor. EE = Standard Error for b_j. LL and UL = Lower and upper bound of the 95 % confidence interval for b_j. F = Knapp-Hartung statistic for the statistical significance of the predictor (the degrees of freedom of this statistic are 1 for the numerator and k-2 for the denominator). p = probability level for the F statistic. Q_E = statistic to assess model specification errors. R^2 = proportion of variance explained by the predictor. ***$p < .0001$.

Table 2
Results of weighted ANOVAs applied on Cronbach's alpha coefficients for the ZBI-7 total score, taking the categorical moderator variables as predictors.

Variable	*k*	N	α_+	IC al 95 % LL	IC al 95 % UL	ANOVA results
Study focus						$F(1,13) = 0.064$ $p = .805$ $R^2 = 0.00$ $Q_W(13) = 116{,}757$, $p < .0001$
Applied	5	529	0.861	0.797	0.905	
Psychometric	10	1.404	0.868	0.828	0.898	
Research Design (I)						$F(1,13) = 0.16$. $p = .696$ $R^2 = 0.00$ $Q_W(13) = 116{,}25$. $p < .0001$
Cross-sectional	14	1.901	0.869	0.834	0.893	
Longitudinal	1	32	0.840	0.580	0.939	
Research Design (II)						$F(1,13) = 0.309$, $p = .588$ $R^2 = 0.00$ $Q_W(13) = 115{,}207$, $p < .0001$
Experimental	1	130	0.890	0.753	0.951	
Non-experimental	14	1.803	0.864	0.830	0.891	
Funding						$F(1,13) = 1698$, $p = .215$ $R^2 = 0.036$ $Q_W(13) = 109{,}948$, $p < .0001$
No/Non reported	4	302	0.832	0.744	0.890	
Yes	11	1.631	0.874	0.842	0.900	
Study location						$F(2,12) = 1216$ $p = .330$ $R^2 = 0.024$ $Q_W(12) = 88{,}959$, $p < .0001$
South America	5	541	0.841	0.771	0.890	
Asia	1	270	0.827	0.633	0.919	
Europe	11	1.122	0.881	0.844	0.908	
Study language						$F(1,13) = 3639$ $p = .079$ $R^2 = 0.165$ $Q_W(13) = 97{,}098$, $p < .0001$
English	11	1.647	0.878	0.848	0.901	
Spanish	4	286	0.817	0.727	0.877	
Target population: caregivers of...						$F(3,11) = 0.341$ $p = .796$ $R^2 = 0.00$ $Q_W(11) = 107{,}352$, $p < .0001$
Dependent patients	4	453	0.863	0.785	0.912	
Oncology patients	2	157	0.830	0.676	0.910	
Palliative care patients	5	641	0.866	0.801	0.910	
Patients with brain damage (Alzheimer's, Parkinson's, acquired)	4	682	0.880	0.816	0.922	

Note. ANOVA = Analysis of variance. *k* = number of independent samples. N = sample size. α_+ = average alpha coefficient. LL and UL = lower and upper bound of the 95 % Confidence Interval for α_+. *F* = Knapp-Hartung statistic for assessing the statistical significance of the moderator variable. Q_W = statistic to assess model specification errors. R^2 = proportion of variance explained by the moderator variable.

Finally, the potential impact of publication bias on the results of the meta-analyses performed was assessed using Egger's test and the trim-and-fill method. Egger's test did not reach statistical significance in any of the cases examined (Sensitivity: $Z = -0.271$, $b = 0.397$, $p = .787$; Specificity: $Z = 1.640$, $b = -0.237$, $p = .101$; PPV: $Z = 1.8154$, $b = 0.397$, $p = .070$; NPV: $Z = -0.168$, $b = 72.545$, $p = .866$), suggesting no strong evidence of publication bias. Similarly, the trim-and-fill method did not impute any missing studies for the meta-analyses of sensitivity or negative predictive value, supporting the robustness of these results. However, two studies were imputed in the meta-analyses of specificity and positive predictive value. The adjusted average specificity decreased slightly to 77.8 %, and the adjusted average positive predictive value to 86.7 %, indicating a modest potential impact. Nonetheless, although slight adjustments were observed in specificity and positive predictive value after imputation, the direction and general pattern of results remained stable. This suggests that, while a minor publication bias cannot be entirely ruled out, it is unlikely to have substantially influenced the main conclusions of the present meta-analysis.

3.5. Moderator analysis

Moderator analyses could only be performed on Cronbach's alpha coefficient. Specifically, simple meta-regressions were performed for quantitative variables and ANOVAs for categorical variables. In both cases, a mixed effects model was assumed. In both simple meta-regression analyses (see Table 1) and weighted ANOVAs applied to Cronbach's alpha coefficients for total scores on ZBI-7 (see Table 2), none of the moderating variables examined showed a statistical relationship with Cronbach's alpha coefficient.

4. Discussion

This study aims to review the reliability and validity of the ZBI-7 total score and identify factors that affect these estimates.

In studies involving the ZBI-7, approximately 70.8 % of the studies did not report the reliability of the scores calculated from their study samples. Instead, they relied on findings from previous studies. This suggests that using a measurement instrument without verifying the score's accuracy in the study sample is common (e.g., Badenes-Ribera et al., 2023; Rubio-Aparicio et al., 2020). It is important to note that the reliability of scores can change with each test application, so researchers should assess the accuracy of the score for their specific sample.

In the studies that provided reliability estimates, Cronbach's alpha coefficient was the most frequently reported (13 studies), followed by McDonald's Omega coefficient (only two studies) and the test-retest coefficient (only one study), which is consistent with previous research (e.g., Badenes-Ribera et al., 2023). The average reliability coefficient of the total scores of the ZBI-7 assessed by Cronbach's alpha coefficient was 0.866. According to psychometric theory (Nunnally and Bernstein, 1994), Cronbach's alpha estimates above 0.70 are considered acceptable for exploratory research, with estimates above 0.80 recommended for general research and above 0.90 for clinical practice. Additionally, according to guidelines proposed by Cicchetti (1994) for assessing the clinical relevance of Cronbach's Alpha coefficients of test scores, coefficients below 0.70 are considered unacceptable, the range of 0.70 to 0.80 is considered fair, 0.80 to 0.90 is considered good, while values above 0.90 are considered excellent. According to these guidelines, the average internal consistency reliability of the ZBI-7 total score measured as Cronbach's alpha coefficient could be considered good for both research and clinical decision-making. This contrasts with the estimates found by Bachner and O'Rourke (2007), who suggested that the short ZBI versions might have reliability estimates below 0.69.

Regarding the validity of the 7-item ZBI scores, the average sensitivity was 98.6 % and the average negative predictive value was 97.3 %, which allows capturing almost all caregivers with caregiver strain with only 7 questions. In addition, this short version of the ZBI had a high average specificity and high average positive predictive value (87.4 % and 93.7 %, respectively), suggesting that few caregivers without overburden will be classified as having overburden. The evidence shows that the ZBI-7 has a good detection performance with a low risk of false positives and negatives.

Finally, Cronbach's alpha coefficients showed considerable variability across studies, which may not be explained by sampling error alone, suggesting that the reliability of the ZBI-7 item scores may be

influenced by factors such as the context in which the ZBI-7 have been applied, the characteristics of the study and the composition of the samples used. Given this wide heterogeneity among reliability estimates, moderator analyses were conducted to identify potential variables that could explain such variability among the reported reliability coefficients. However, none of the moderator variables analyzed showed a statistical relationship with the variability exhibited for the Cronbach's alpha coefficients, suggesting that factors such as age, gender, target population, geographical location and some study characteristics did not affect the reliability estimates of the ZBI-7 score. Consequently, Cronbach's alpha coefficients did not vary systematically with age, gender, target population, geographical location, or some study characteristics. This implies that other unexamined or unreported variables in the studies contribute to the high heterogeneity observed. Methodological or artificial heterogeneity may have played a role, such as differences in data quality and bias in the studies (Langan, 2021).

4.1. Limitations and future directions

The current study has some limitations. First, some of the meta-analyses conducted could only be based on data from 3 studies (e.g., sensitivity, specificity, and predictive values), which might impact the accuracy of the results obtained from these analyses. Second, a small number of articles (n = 2) reported test-retest correlation coefficients, making it impossible to conduct a comprehensive meta-analysis on the long-term consistency of test scores. Third, by including as an inclusion criterion that the studies report their reliability estimates, it was not possible to analyse the characteristics of those studies that induced or omitted their reliability estimates. Fourth, the high heterogeneity observed could not be explained by the potential moderating variables analyzed. Future research should explore additional moderators or methodological factors that may account for this variability. Fifth, the imbalance in group sizes across several moderator analyses may have compromised the reliability of between-group comparisons by reducing statistical power and increasing the potential for biased estimates. Therefore, the results of the moderator analyses should be interpreted with caution. Future research with a more balanced representation across moderator categories is necessary to improve the robustness of subgroup comparisons. Sixth, the meta-analysis of the validity coefficients was based on a limited number of primary studies (three), which restricted the possibility of assessing heterogeneity and reduced the precision of the pooled estimate. While this preliminary analysis provides an initial indication of the average validity of the measure, the findings should be interpreted with caution. Future research including a greater number of validity studies is needed to obtain more robust and generalizable conclusions. Seventh, including only articles using the ZBI-7 limits the study's potential. For future research, a meta-analysis could be performed to generalize the reliability of all versions of the ZBI. Finally, future studies should include test-retest reliability, which measures the consistency of measurements over time. Without this information, it is unclear whether changes in scores reflect actual changes in the measured variable or only variations due to test-retest reliability.

4.2. Implications for professional practice

Given the high values for sensitivity and negative predictive value on the one hand, and specificity and positive predictive value on the other hand, the ZBI-7 is very useful as a screening method in primary healthcare settings to identify severely overburdened caregivers. It also helps evaluate the care and support needs of these caregivers. The ZBI-7 is a valuable tool that can be used during annual check-ups for individuals with chronic diseases. Additionally, the ZBI-7 can be utilized in informal caregivers during medical check-ups of elderly individuals and medical visits for those receiving palliative care. This tool is practical and easy to implement, and its results can aid healthcare providers and researchers in developing effective preventive measures and intervention strategies for caregivers.

4.3. Relevance for clinical practice

This study presents statistical evidence of the validity and reliability of the scores from the ZBI-7 caregiver burden instrument. As a result, this study supports high-quality clinical practice by demonstrating that the ZBI-7 provides accurate and reliable scores for its intended use, thus enabling healthcare professionals to make accurate clinical decisions. The high average reliability shows that the results remain reliable in different contexts, with formal and informal caregivers and people with diverse needs. The high sensitivity and specificity of the ZBI-7 demonstrate a low risk of misidentification and adequately discriminate between caregivers with and without strain. This is beneficial at a clinical level, providing evidence that intervention resources are targeted to carers who need them.

5. Conclusion

In conclusion, the short version ZBI-7 scores demonstrated good properties in terms of reliability and validity in detecting caregiver overload in the health care setting.

Grants

The author JG-P is the beneficiary of a grant (code: FPU22/03053) from the Ministry of Science, Education and Universities (Spain).

CRediT authorship contribution statement

Júlia Gisbert-Pérez: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation. **Elena Cejalvo:** Writing – review & editing, Writing – original draft, Methodology, Data curation, Conceptualization. **Manuel Martí-Vilar:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Data curation, Conceptualization. **Laura Badenes-Ribera:** Writing – review & editing, Writing – original draft, Visualization, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Ethical considerations

Not required.

Funding

This research has not received any specific funding from any public, commercial or non-profit organisation.

Declaration of competing interest

The authors declare no conflicts of interest.

Acknowledgements

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jad.2025.119744.

References[1]

Atella, V., Piano Mortari, A., Kopinska, J., Belotti, F., Lapi, F., Cricelli, C., Fontana, L., 2019. Trends in age-related disease burden and healthcare utilization. Aging Cell 18, e12861.

Bachner, Y.G., O'Rourke, N., 2007. Reliability generalization of responses by care providers to the Zarit Burden Interview. Aging Ment. Health 11, 678–685.

Badenes-Ribera, L., Rubio-Aparicio, M., Sánchez-Meca, J., 2020. Metaanálisis de generalización de la fiabilidad. Informació Psicològica 17–32.

Badenes-Ribera, L., Duro-García, C., López-Ibáñez, C., Martí-Vilar, M., Sánchez-Meca, J., 2023. The Adult Prosocialness Behavior Scale: a reliability generalization meta-analysis. Int. J. Behav. Dev. 47, 59–71.

Bonett, D.G., 2002. Sample size requirements for testing and estimating coefficient alpha. J. Educ. Behav. Stat. 27, 335–340.

Borenstein, M.J., Hedges, L.V., Higgins, J.P., Rothstein, H.R., 2021. Introduction to Meta, second ed. Wiley, UK.

Breinbauer, K.H., Vásquez, V.H., Mayanz, S.S., Guerra, C., Millán, K.T., 2009. Original and Abbreviated Zarit Caregiver Burden Scales. Validation in Chile, Revista Médica de Chile, p. 137.

Cejalvo, E., Martí-Vilar, M., Gisbert-Pérez, J., Badenes-Ribera, L., 2025. The CarerQol instrument: a systematic review, validity analysis, and generalization reliability study. J. Clin. Med. 14.

Chang, H., Chiou, C., Chen, N., 2009. Impact of mental health and caregiver burden on family caregivers' physical health. Arch. Gerontol. Geriatr. 50, 267–271.

Chong, E., Crowe, L., Mentor, K., Pandanaboyana, S., Sharp, L., 2022. Systematic review of caregiver burden, unmet needs and quality-of-life among informal caregivers of patients with pancreatic cancer. Support Care Cancer 31.

Cicchetti, D.V., 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol. Assess. 6, 284–290.

Duval, S., Tweedie, R., 2000. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 56, 455–464.

Egger, M., Davey Smith, G., Schneider, M., Minder, C., 1997. Bias in meta-analysis detected by a simple, graphical test. BMJ 315, 629–634.

Fekete, C., Tough, H., Siegrist, J., Brinkhof, M.W., 2017. Health impact of objective burden, subjective burden and positive aspects of caregiving: an observational study among caregivers in Switzerland. BMJ Open 7, e017369.

Gort, A.M., March, J., Gómez, X., de Miguel, M., Mazarico, S., Ballesté, J., 2005. Short Zarit scale in palliative care. Med. Clin. 124, 651–653.

Gräßel, E., Chiu, T., Oliver, R., 2003. Development and Validation of the Burden Scale for Family Caregivers (BSFC). Comprehensive Rehabilitation and Mental Health Services, Toronto.

Gronlund, N.E., Linn, R.L., 1990. Measurement and Evaluation in Teaching, 6th ed. Macmillan, New York.

Hartung, J., Knapp, G., 2001. On tests of the overall treatment effect in meta-analysis with normally distributed responses. Stat. Med. 20, 1771–1782.

Hernández-Padilla, J.M., Correa-Casado, M., Granero-Molina, J., Cortés-Rodríguez, A.E., Matarín-Jiménez, T.M., Fernández-Sola, C., 2019. Psychometric evaluation and cultural adaptation of the Spanish version of the "Scale for End-of Life Caregiving Appraisal. Palliat Support Care 17, 314–321.

Higgins, J.P., Thompson, S.G., 2002. Quantifying heterogeneity in a meta-analysis. Stat. Med. 21, 1539–1558.

Higgins, J.P., Thompson, S.G., Deeks, J.J., Altman, D.G., 2003. Measuring inconsistency in meta-analyses. BMJ 327, 557–560.

Huedo-Medina, T.B., Sánchez-Meca, J., Marín-Martínez, F., Botella, J., 2006. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? Psychol. Methods 11, 193–206.

Knapp, G., Hartung, J., 2003. Improved tests for a random effects meta-regression with a single covariate. Stat. Med. 22, 2693–2710.

Kokorelias, K.M., Gignac, M.A.M., Naglie, G., Cameron, J.I., 2019. Towards a universal model of family centered care: a scoping review. BMC Health Serv. Res. 19, 564.

Langan, D., 2021. Assessing Heterogeneity in Random-effects Meta-analysis (Methods In Molecular Biology).

Liu, Z., Heffernan, C., Tan, J., 2020. Caregiver burden: a concept analysis. Int. J. Nurs. 7, 438–445.

López-López, J.A., Botella, J., Sánchez-Meca, J., Marín-Martínez, F., 2013. Alternatives for mixed-effects meta-regression models in the reliability generalization approach: a simulation study. J. Educ. Behav. Stat. 38, 443–469.

López-López, J.A., Marín-Martínez, F., Sánchez-Meca, J., Van den Noortgate, W., Viechtbauer, W., 2014. Estimation of the predictive power of the model in mixed-effects meta-regression: a simulation study. Br. J. Math. Stat. Psychol. 67, 30–48.

Marvardi, M., Mattioli, P., Spazzafumo, L., Mastriforti, R., Rinaldi, P., Polidori, M.C., Cherubini, A., Quartesan, R., Bartorelli, L., Bonaiuto, S., Cucinotta, D., Di Iorio, A., Gallucci, M., Giordano, M., Martorelli, M., Masaraki, G., Nieddu, A., Pettenati, C., Putzu, P., Mecocci, P., 2005. The Caregiver Burden Inventory in evaluating the burden of caregivers of elderly demented patients: results from a multicenter study. Aging Clin. Exp. Res. 17, 46–53.

Navarro Jiménez, J.M., Cruzado Rodríguez, J.A., Boya Cristiá, M.J., 2022. Concordance of patient-caregiver information in advanced cancer. Psicooncologia 19, 283–298.

Nunnally, J.C., Bernstein, I.H., 1994. Psychometric Theory. McGraw Hill, New York.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C., et al., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372, n71.

Raudenbush, S.W., 2009. Analyzing effect sizes: random-effects models. In: Cooper, H., Hedges, L.V., Valentine, J.C. (Eds.), The Handbook of Research Synthesis and Meta-analysis, second ed. Russell Sage Foundation, New York, pp. 295–331.

Regueiro Martínez, A.A., Pérez-Vázquez, A., Gómara Villabona, S.M., Ferreiro Cruz, M. C., 2007. Short Zarit interview on burden of care for caregivers in primary care. Aten. Primaria 39, 185–188.

Rubio-Aparicio, M., Badenes-Ribera, L., Sánchez-Meca, J., Fabris, M.A., Longobardi, C., 2020. A reliability generalization meta-analysis of self-report measures of muscle dysmorphia. Clin. Psychol. 27.

Sánchez-Meca, J., López-López, J.A., López-Pina, J.A., 2012. Some recommended statistical analytic practices when reliability generalization studies are conducted. Br. J. Math. Stat. Psychol. 66.

Sánchez-Meca, J., Marín-Martínez, F., López-López, J.A., Núñez-Núñez, R.M., Rubio-Aparicio, M., López-García, J.J., et al., 2021. Improving the reporting quality of reliability generalization meta-analyses: the REGEMA checklist. Res. Synth. Methods 12, 516–536.

Schmidt, F.L., Oh, I., Hayes, T.L., 2008. Fixed- versus random-effects models in meta-analysis: model properties and an empirical comparison of differences in results. Br. J. Math. Stat. Psychol. 62, 97–128.

Starr, L.T., Bullock, K., Washington, K., Aryal, S., Parker Oliver, D., Demiris, G., 2022. Anxiety, depression, quality of life, caregiver burden, and perceptions of caregiver-centered communication among black and white hospice family caregivers. J. Palliat. Med. 25, 596–605.

Tu, J.Y., Jin, G., Chen, J., Chen, Y., 2022. Caregiver burden and dementia: a systematic review of self-report instruments. J. Alzheimer Dis. 86, 1527–1543.

Tyo, M.B., McCurry, M.K., 2020. An integrative review of measuring caregiver burden in substance use disorder. Nurs. Res. 69, 391–398.

Vélez Lopera, J.M., Berbesí Fernández, D., Cardona Arango, D., Segura Cardona, A., Ordóñez Molina, J., 2012. Validation of the abbreviated Zarit scales for measuring burden syndrome in the primary caregiver of an elderly patient. Aten. Primaria 44, 411–416.

Viechtbauer, W., 2010. Conducting meta-analyses in R with the metafor package. J. Stat. Softw. 36, 1–48.

Viechtbauer, W., 2021. Model checking in meta-analysis. In: Schmid, C.H., Stijnen, T., White, I.R. (Eds.), Handbook of Meta-analysis. CRC Press, Boca Raton, pp. 219–254.

Viechtbauer, W., Cheung, M.W.L., 2010. Outlier and influence diagnostics for meta-analysis. Res. Synth. Methods 1, 112–125.

Viechtbauer, W., López-López, J.A., Sánchez-Meca, J., Marín-Martínez, F., 2015. A comparison of procedures to test for moderators in mixed-effects meta-regression models. Psychol. Methods 20, 360–374.

Vitaliano, P.P., Russo, J., Young, H.M., Becker, J., Maiuro, R.D., 1991. The screen for caregiver burden. Gerontologist 31, 76–83.

Zarit, S.H., Reever, K.E., Bach-Peterson, J., 1980. Relatives of the impaired elderly: correlates of feelings of burden. Gerontologist 20, 649–655.

Zarit, S.H., Orr, N., Zarit, J.M., 1985. The Hidden Victims of Alzheimer's Disease: Families Under Stress. New York University Press, New York.

[1] Articles included in the systematic review and meta-analysis are referenced in Appendix B.