Reference	Sample	Measurement Instrument			Psychometric Characteristics	
		Name of the Scale	Domains and Constructs	Length and Format of Instrument	Validity	Reliability
Hebert *et al.* (2000)[30] Canada	ADRD	Zarit Burden Interview (ZBI)	Burden Two factors: (1) Personal strain; (2) Role strain	12 items; 5-point Likert scale (0=Never to 4=Nearly always)	The structural validity of ZBI was assessed through a stepwise process that began with an exploratory analysis of the original 22-item ZBI scale followed by CFA. After comparing several CFA competing models, authors went back and ran an EFA model producing a 2-factor solution ("personal strain" and "role strain") with a reduced 12-item ZBI scale that was further tested for goodness of fit with a CFA model. Compared to previous competing CFA models, the 2-factor solution produced the best goodness-of-fit indexes (e.g., AGFI=0.98, RMR=0.10). Concurrent validity was established by significant Spearman's correlations (*p-value*s < 0.001) between scores on the 12-item ZBI and a) CG depression as measured by the CES-D (rho=0.57), b) behavior problems, measured by the Dementia Behavior Disturbance scale (rho=0.58).	Cronbach's α, full scale =0.91 Guttman's split-half reliability estimate for the full 12-item scale=0.91
Guberman *et al.* (2001)[31] Canada	Mixed	The Caregiver Risk Screen (CRS)	Risk to CG mental and physical wellbeing Two domains: (1) Level of risk to CG mental wellbeing; (2) Level of risk to CG physical wellbeing	12 items, 4-point Likert scale (0=Totally disagree, 1=Somewhat disagree, 2=Somewhat agree, 3=Totally agree)	Content validity was established through several scale development stages. First, a literature review of validated tools on caregiving psycho-social scales measuring burden, depression, social support, etc. was conducted. Second, non-validated CG assessment tools were also collected from key informers representing public, private, and non-profit agencies as well as research on non-validated tools which described what key CG risk elements should contain. Third, nine focus groups were conducted with family CGs, administrators, and community care practitioners to identify the key elements to be included in a measure of risks to caregiving mental and physical wellbeing. Informal pretests were also conducted to assess the relevance of preliminary items. No formal tests were conducted to study the dimensionality of the scale. Concurrent validity was assessed by calculating a Pearson's correlation coefficient between the total scores on the 12-item Caregiver Risk Screen (CRS) and the Caregiver Burden Screen (Rankin et al, 1994), as the external criterion. The correlation was statistically significant (r=0.83, p<0.005). Note: The Caregiver Burden Screen was chosen as the external criterion to establish the CRS validity because it was short, validated in English and French, and contained two relevant dimensions: CG depression and patient level of care/demands.	Cronbach's α, full scale =0.88
Gitlin *et al.* (2002)[32] United States	ADRD	Task Management Strategy Index (TMSI)	CG strategies to simplify everyday self-care tasks of patients One Factor: CG actions to cope with deficits in functioning, orientation, and awareness of patients	19 items, 5-point Likert scale (ranging from 1=Never to 5=Always)	Content validity was demonstrated by gathering an initial 20-item pool from observational research and clinical techniques used by occupational therapists that address a particular action that is designed to change the external environment by simplifying task requirements and interactions for the person with dementia. As such, a score on the TMSI scale reflects behavioral actions that are designed to compensate for the patient's functional loss. After eliminating an item due to interpretability difficulties, a final 19-item TMSI included items that reflected constructive strategies that would benefit both CG and care recipient. The structural validity of the 19-item TMSI was examined in an independent sample of 202 CGs (Sample 1) using an EFA with principal-axis factoring extraction method. EFA identified one factor accounting for 60.2% of the variance in items. Factor loadings ranged from 0.35 to 0.87. Using an independent sample of 255 CGs with similar characteristics as the sample used in the EFA, the concurrent validity was demonstrated by computing Pearson's correlations between TMSI scores and a) functional dependency of Only ADRD patients as measured by "ADL dependence" (0.237, p<0.001), b) CG self-efficacy, as measured by "ADL self-efficacy" (0.173, 0<.05), and c) use of positive coping strategies, as measured by a subscale of the Dementia Management Scale, DMS (0.507, p<0.001). Discriminant validity. As expected, TMSI scores were not associated with a) level of CG upset with disruptive behaviors, as measured by the Disruptive Behaviors subscale of the RMBPC (-0.002, p>.05) and b) CG use of criticism-based strategies, as measured by a subscale of the DMS (-0.055, p>.05).	Cronbach's α, full scale, in Sample 1 (N=202) was 0.81. Note: The Cronbach's α estimate for the full scale in Sample 2 was slightly lower (0.74), but above recommended thresholds.
Fortinsky *et al.* (2002)[33] United States	ADRD	Family caregiver self-efficacy for managing dementia	Perceived Self-Efficacy Two factors: (1) Symptom management; (2) Use of community support services	9 items, 10-point Likert scale (ranging from 1=Not at all certain to 10=Very certain)	Content validity was shown by reviewing the literature on perceived self-efficacy, applying its principles to family CGs of persons with Alzheimer's disease, defining relevant conceptual domains, and developing a preliminary bank of 10 items. The item count was kept to a minimum to limit the scale's burden on respondents. The structural validity was demonstrated by an EFA with PAF as factor extraction method and Varimax rotation to simplify factor interpretation. EFA identified two factors accounting for 54% of the variance. A scree plot confirmed the two-factor solution. One item was eliminated due to low factor loading resulting in a final 9-item scale. Concurrent validity was established by calculating Pearson's correlations between a "Global CG competence measure" and the two perceived self-efficacy factors: a) symptom management (r=0.49, p < 0.01) and b) use of community support services (r=0.27, p < 0.01).	Cronbach's α by subscales: Symptom management (α=0.77) Community support service use (α=0.78)

Marwit *et al.* (2002)[34] United States	ADRD	Marwit–Meuser Caregiver Grief Inventory (MM-CGI)	CG grief Three factors: (1) Personal Sacrifice Burden; (2) Heartfelt Sadness and Longing; (3) Worry and Felt Isolation	50 items, 5-point Likert scale (ranging from 1=Strongly disagree to 5=Strongly agree)	Content validity was demonstrated by conducting 16 focus groups with N=90 dementia CGs. Focus groups resulted in the generation of a pool of 184 grief statements/items. A preliminary analysis of the skewness of the items distribution led to the reduction of the pool to 164 items. The structural validity of the MM-CGI was established through a stepwise procedure. First, iterative PCAs (both unrotated and rotated) led to the elimination of items with high unique variances resulting in a final pool of 69 items with three distinct components that were confirmed with a scree plot. Next, using the pool of 69 items, an EFA, with PAF extraction method and oblique rotation yielded a three-factor solution explaining 34% of the item variance. Items with double loadings were dropped resulting in a final 50-item MM-CGI scale. Concurrent validity was established by significant positive Spearman's rank correlations between scores of the MM-CGI scale and: a) depression measured by Beck Depression Index (rho=0.758, p < 0.01) and the Geriatric Depression Scale (rho =0.714, p < 0.01); b) grief measured by scores on the Anticipatory Grief Scale (rho =0.798, p < 0.01); and c) strain measured by the Caregiver Strain Index (rho =0.656, P < 0.01). Statistically significant (p < 0.01) negative correlations between scores of the MM-CGI and a) the Caregiver Wellbeing Scale-Basic Needs subscale (rho=-0.66) and b) the Perceived Social Support Questionnaire-Family subscale (rho=-0.36) also supported the convergent validity of this MM-CGI scale.	Cronbach's α, full scale =0.96. Cronbach's α by subscales: Personal Sacrifice Burden (α=0.93) Heartfelt Sadness and Longing (α=0.90) Worry and Felt Isolation (α=0.91) Guttman's split-half estimate, full scale=0.91 Guttman's split-half by subscales: Personal Sacrifice Burden (Guttman's split-half=0.91) Heartfelt Sadness and Longing (Guttman's split-half=0.86) Worry and Felt Isolation (Guttman's split-half=0.91)
Steffen *et al.* (2002)[35] United States	ADRD	Revised Scale for Caregiving Self-Efficacy (R-SCSE)	Caregiving self-efficacy Three factors: (1) Obtaining respite; (2) Responding to disruptive behavior; (3) Controlling upsetting thoughts	15 items, Confidence in doing activity (ranging from 0=Cannot do at all to 100=Certain can do)	To expand the content validity of the original self-efficacy measure developed by Zeiss et al. (1999), the authors conducted a thorough literature review and added 37 items mostly representing a new domain (management of distressing thoughts). These 37 items were added to the original 14-item self-efficacy measure which contained two domains: self-care and problem-solving. Two independent samples were used to assess the structural validity of the revised scale. After examining the item distributions using responses from the first independent sample (N=169), the initial pool of 51 items was reduced to 33 items. After iterative EFAs using PAF for factor extractions and Promax rotations, items were further eliminated due to low factor loadings. The final EFA yielded a three-factor structure underlying a final 15-item scale accounting for 62% of the variance. A CFA conducted on the second independent sample (N=145) produced an adequate fit for the three-factor, 15-item solution (e.g., CFI=0.93 and the $\chi 2/df = 1.59$). (Of note, a value less than 3 is a commonly used indication of adequate fit.) The concurrent validity was established by significant Pearson's correlations between scores on different R-SCSE subscales (factors) and (a) depression, as measured by Short Form Beck Depression Inventory (Obtaining respite: r=-0.38; *p*<0.001, Responding to Disruptive Behavior: r=-0.34, *p*<0.001, Controlling upsetting thoughts: r=-0.38; *p*<0.001), (b) anger, measured by the Spielberger's Trait Anger (Responding to disruptive behavior: r=-0.41, *p*<0.001), (c) anxiety, measured by Spielberger's Trait Anxiety (Controlling upsetting thoughts: r=-0.62, *p*<0.001), and (d) perceived social support network, measured by the Arizona Social Support Interview Schedule (Obtaining respite: r=0.16, *p*<0.05).	Cronbach's α by subscales: Obtaining respite (α=0.88); Responding to disruptive behavior (α=0.84); Controlling upsetting thoughts (α=0.86) Test-retest reliability was calculated with a subset participants (N=100) after a 2-week interval using Pearson's correlation coefficients. Test-retest reliability by subscales: Obtaining respite (r=0.76); Responding to disruptive behavior (r=0.70); Controlling upsetting thoughts (r=0.76) Note: Reliability estimates by subscales were obtained in both independent samples. The pattern of estimates was the same in the second sample.
Suwa (2003)[36] Japan	ADRD	Assessment Scale for Caregiver's Experience with Dementia (ASCED)	Stages in caregiving experience Three factors (subscales): (1) Empathetic caregiving experience; (2) Disciplinary caregiving experience; (3) Resigned caregiving experience	24 items, 5-point Likert scale (1=Never, 2=Very infrequently, 3=Sometimes, 4=Frequently, 5=Continually)	Content validity was established by using prior qualitative research results that included CG interviews to develop a pool of items. The draft of the ASCED scale was guided by a 7-stage CG experience model. Ten items were written for each stage resulting in an initial 70-item measure. A panel of experts judged the appropriateness of the items for each caregiving stage, and another panel of CGs judged the legibility of items. After administering the 70-item ASCED tool to the sample (N=90), the correlation coefficients were computed for each of the 10 items at all seven stages. Using item-total correlation coefficients greater than 0.40 as item selection criterion, a final pool of 35 items were retained (5 items per the seven stages). Structural validity. An EFA with Varimax rotation was conducted on the 35-item tool to identify underlying patterns or "factors". EFA yielded a 3-factor model, but 11 items were dropped due to low factor loadings. The final EFA using the 24-item ASCED tool also showed a 3-factor structure accounting for 51.4% of the variance. Concurrent validity was demonstrated by "moderate" Pearson's correlations between scores on the ZBI and (a) scores on the "Disciplinary caregiving experience" subscale (r=0.38, *p* < 0.01) and (b) the "Resigned caregiving experience" subscale (r=0.41, *p* < 0.01). The correlation between the "Empathetic caregiving experience" and ZBI scores, however, was insignificant (r=0.08, *p*=0.45).	Cronbach's α by subscales: Empathetic caregiving experience (α=0.89); Disciplinary caregiving experience (α=0.78); Resigned caregiving experience (α=0.81) Test-retest (temporal) reliability was evaluated with Pearson's correlations between scores in the ACED scale obtained in two administrations (1 to 4 weeks apart) using a subsample of respondents (N=30). Test-retest reliability for subscales: Empathetic caregiving experience (r=0.34*); Disciplinary caregiving experience (r=0.75);Resigned caregiving experience (r=0.71)
Mahoney *et al.* (2003)[37] United States	ADRD	Caregiver Vigilance Scale (CVS)	Caregiver vigilance or oversight of patient activities One factor: Oversight of patient activities	4 items, Items 1 and 2 are scored with two scales: 2-point, binary scale (0=No, 1=Yes) and time estimate in hours and	Content validity was established through a year-long qualitative study collecting data from discussions with 70 family CGs on vigilance and oversight of care recipients. The study led to the key finding that a vigilant CG is actively involved and perceives herself as responsible for the care recipient even when not "actively" providing care. As a consequence, four vigilant questions/items were developed that reflected "being there" and "doing things" for the care recipient. The items were pilot tested with 15 family CGs resulting in the refinement and re-wording of questions.	Cronbach's α, full scale =0.66.

				minutes per day Items 3 and 4 have one scale: Time estimate in hours per day. Note: CVS items are recoded as total number of hours per day.	A PCA was conducted to study the structural validity of the 4-item scale. The analysis yielded a single component accounting for 50% of the variance. Concurrent validity was supported by a significant negative Pearson's correlation between CVS scores and scores on the MMSE (r=-0.34, *p*<0.001). The greater the cognitive impairment (lower MMSE score), the greater the score in the CVS scale. The correlation between CVS scores and total scores on the Revised Memory and Behaviors Problems Checklist, RMBPC was, as expected, positive and significant (r=0.15, *p* < 0.001).	
Goolieb & Rooney (2003)[38] Canada	ADRD	RIS Eldercare Self-efficacy Scale	Caregiver Self-Efficacy Beliefs Three factors: (1) Relational self-efficacy; (2) Instrumental self-efficacy; (3) Self-soothing efficacy	10 items, 5-point Likert scale (1=I'm certain I can't do this, 2=I probably can't do this, 3=Maybe I can and maybe I can't do this, 4=I probably can do this, 5=I'm certain I can do this)	Content validity was shown by developing 13 items from a review of the perceived self-efficacy literature. Among the many dimensions of perceived self-efficacy in the literature, authors focused on three dimensions they believed were universally experienced by CGs: CG beliefs about their ability to manage caregiving, to maintain a cooperative relationship with a care recipient, and to sustain personal wellbeing in demanding situations. Prior to the inspection of the underlying structure of the scale, inter-item correlations were calculated and separate internal consistency analyses were performed for each of the hypothesized subscales to identify items that may reduce reliability estimates. This analysis reduced the item pool to 12 items. Iterative PCAs with oblique rotations were subsequently conducted to determine the factorial structure of the scale. A scree plot and eigenvalues inspection suggested a three-component/factor solution. Factor loadings led to removing two additional items. The final PCA applied to the 10-item scale also yielded a 3-component solution (subscales) that accounted for 66% of the total variance in items. Concurrent validity was demonstrated by expected significant positive Pearson's correlations between Perceived social support and the three RIS subscales a) Relational self-efficacy (r=0.20, *p* <0.05), b) Instrumental self-efficacy (r=0.23, *p* <0.01), and (c) Self-soothing efficacy (r=0.30, *p* <0.001). All three subscales (*Relational*, *Instrumental*, and *Self-soothing*) were (as expected) significantly (*p*-values < 0.05) associated with CG personality traits such as: a) Optimism (r=0.28; r=0.41; r=0.36, respectively), b) Agreeableness (r=0.31; r=0.22; r=0.25, respectively), and Conscientiousness (r=0.33; r=0.40; r=0.29, respectively). The RIS Relational and Instrumental subscales correlated significantly (*p*-values <0.01) with a "Coping "measure (r=0.32 and r=0.31, respectively). Finally, the RIS Relational subscale correlated (as expected) negatively with "anger expression" (r=-0.26).	Cronbach's α by subscales: Relational self-efficacy (α=0.72) Instrumental self-efficacy (α=0.74) Self-soothing efficacy (α=0.79) Test-retest reliability was calculated with Pearson's correlations between RIS scores obtained at baseline and 4-6 months later for a subsample of respondents (N=105). Rest-retest reliability for subscales: Relational self-efficacy (r=0.48, p<0.001) Instrumental self-efficacy (r=0.69, p<0.001) Self-soothing efficacy (r=0.60, p<0.001)
Gräßel *et al.* (2003)[39] Canada	ADRD	Burden Scale for Family Caregivers (BSFC)	Subjective burden One factor: Subjective burden	28 items, 4-point Likert scale (0=No, definitely not, 1=No, not really, 2=Yes, generally, 3=Yes, definitely)	Content validity was demonstrated through a multi-step item-development process. Statements recorded from CG discussion groups and interviews were developed into items and a prototype or preliminary scale. The preliminary scale was compared to published CG burden scales and reviewed by an expert panel prior to pilot testing. The scale was pilot tested and items were revised for comprehensibility and acceptability. Finally, the scales were translated from German into English before further psychometric testing. To study the structural validity of the 28-item BSFC scale, the authors conducted PCAs without rotation on two independent samples: one of dementia CGs (N=1143) and a second of non-dementia CGs (N=548). (Forty-five percent of the non-dementia CGs were caring for elderly people with relatively unimpaired cognition and the remaining 55% were carers of individuals with neurological disorders.) The PCA of the dementia CG responses yielded a one-component/factor solution explaining 29.1% of the variance. (The PCA of the non-dementia sample yielded a similar one-factor structure explaining 31.5% of the variance.) Using the dementia CG sample, the concurrent validity was established by a significant (*p* < 0.001) positive Pearson's correlation between the BSFC scores and patient behavioral disturbances (r=0.39) measured by the Sandoz Clinical Assessment-Geriatric. (The non-dementia CG BFSC scores produced a similarly significant positive Pearson's correlation with SCAG scores (r=0.44, *p* <0.001).)	Sample 1 (Dementia CGs) Cronbach's α, full scale =0.90 Split-half reliability was calculated with the Spearman-Brown correlation coefficient=0.88 Note: Reliability calculations for the second independent sample of non-dementia CGs yielded similar results: Cronbach's α=0.91 and the split-half reliability coefficient=0.90.
Stevens *et al.* (2004)[40] United States	ADRD	The Leisure Time Satisfaction (LTS)	Satisfaction with leisure time One factor: Satisfaction with Leisure Time (impact of caregiving on leisure activities).	6 items, 3-point Likert scale (0=Not at all, 1=A little, 2=A lot)	Content validity was established through an extensive literature review that revealed only one existing measure of leisure that assessed the concept of satisfaction with leisure: the 51-item Leisure Satisfaction Scale (LSS).[41] However, this measure had not been evaluated with CGs of older adults and had an estimated administration time judged inappropriate as a brief measure to assess changes in leisure after caregiving interventions. Taken into account the review of literature and limitations of the existing LSS tool, the authors developed a short 6-item scale to assess the distinct psychological dimension of satisfaction with the amount of time spent in leisure activities relevant to family CGs of those with Alzheimer's disease or a related dementia. To establish the structural validity of the 6-item scale, the baseline sample (N=1225) with non-missing item data was randomly split into two subsamples to perform a PCA (N=900-roughly 75% of the sample) and CFA (N=291-roughly 25% of the sample). A PCA, oblique rotation, and weighted least squares	Cronbach's α, full scale =0.80.

					estimation yielded a one-factor solution explaining 57.8% of the variance. The CFA indicated a good fit for the one-factor solution with a RMSEA statistic of 0.069. Concurrent validity was shown by "small to moderate" and significant (p<0.001) Spearman's rank correlations between scores on the LTS and a) a 3-item measure of *CG satisfaction with social support* (rho =0.32), b) *social network*, measured by the Lubben Social Network (rho =0.25), and c) wellbeing, measured by the CES-D-wellbeing subscale (rho =0.28). Expected negative correlations with LTS scores included *time spent on ADL activities* (rho =-0.21) and *depression* measured by the CES-D (rho =-0.37).	
Gaugler *et al.* (2004)[42] United States	ADRD	Perceived Unmet Need (PUN)	Perceived unmet need at different "stages" of the caregiving career Seven domains: (1) ADL care tasks; (2) IADL care tasks; (3) Dementia symptoms; (4) Timing of care (5) Formal support (6) Information; (7) Confidante/family support	34 items, 2-point/binary scale (0=No, 1=Yes) (Respondents are asked: "Do you need more help with/help providing…?" The "yes" responses for each domain are summed to create "unmet need" scores.)	Content validity was demonstrated by researchers identifying seven domains of unmet CG need from literature review and consultation with experts in dementia caregiving. The instrument under development to measure unmet need was administered to three groups of dementia CGs based on the "stage" of the care recipient: still living in the community, institutionalized, or deceased. Although authors do not study the underlying dimensional structure of the PUN measure and do not provide analyses establishing the concurrent validity of the seven domains or the full scale using simple correlations, they conducted three independent multivariate regression *path analyses* by the "stage" of care recipient to study the associations between unmet needs domains and measures of subjective stress of CGs while controlling for demographic variables. (Three outcome measures of subjective stress were simultaneously examined in each path model: (a) a three-item *role overload scale*, (b) a three-item *role captivity scale*, and (c) three-item *scale assessing CGs' loss of intimate exchange* (*feelings of emotional/physical separation*). All models produced acceptable fit indexes (e.g., RMSEA ranged from 0.02-0.03 and the GFI ranged from 0.92 to 0.97). Among CGs of individuals in the community, scores on the *Confidant/family support domai*n were significantly associated with scores on all three outcomes (*role overload*, *role captivity*, and *loss of intimate exchange*). For CGs with institutionalized care recipients, scores on the *ADL care tasks* domain were significantly associated with all three outcomes. For those in the deceased care receiver group ("bereave CGs"), scores on the *Confident/family support domain* were associated with both *role overload* and *loss of intimate exchange*.	Cronbach's α by subscales: ADL care tasks (α=0.85) IADL care tasks (α=0.86) Dementia symptoms (Pearson's correlation, r=0.54, p<0.01) (Only two items) Timing of care (α=0.79) Formal support (α=0.77) Information (α=0.68) Confidante (α=0.79)
Tarlow *et al.* (2004)[43] United States	ADRD	The Positive Aspects of Caregiving (PAC)	Positive Aspects of Caregiving Two factors: (1) Self-Affirmation; (2) Outlook on Life (Positive aspects of caregiving refer to the CGs' sense that their caregiving experience is generally satisfying and rewarding.)	9 items, 5-point Likert scale (1=Disagree a lot, 2=Disagree a little, 3=Neither agree or disagree, 4=Agree a little, 5=Agree a lot)	Content validity was established by a literature review of studies of CGs for persons with dementia that included a measure for positive aspects of caregiving. The studies provided operational definitions of positive aspects of caregiving that authors used to develop the PAC tool that differed from prior measures in three ways (1) response options were changed from the yes/no format to a 5-point Likert scale to increase variability of responses and improve reliability, (2) questions were rephrased to accommodate different response options, and (3) instructions were modified to facilitate ease of administration. The initial PAC tool contained 11 items. To establish the structural validity of the 11-item scale, the sample (N=1229) was randomly split into two subsamples to perform a PCA (N=922) and a CFA (N=307). The PCA with oblique rotation and weighted least squares estimation yielded a two-component solution. After eliminating two items with low loading the final 9-item scale accounted for 45% of the total variance in items. The CFA indicated a good fit for the two-factor solution with a RMSEA statistic of 0.0689. Concurrent validity was examined by Spearman's rank correlations between scores in the PAC scale and scores in (a) the Somatic and Well-Being subscales of the CES-D, (b) the RMBPC (burden), and (c) the Satisfaction with Received Support and Negative Interactions subscales of the Inventory of Socially Supportive Behaviors (ISSB). The resulting correlations were significant (p-values < 0.001) and lower than expected (all < 0.30, "small to moderate") but in the anticipated directions. The PAC was positively associated with *wellbeing* (rho=0.24) and *satisfaction with support* (rho=0.15), but negatively associated with the *RMBPC-burden* (rho= -0.23), and *somatic aspects of depression* (rho= -0.17). PAC was not associated with *negative social interactions* (rho= -0.05, ns).	Cronbach's α, full scale =0.89. Cronbach's α by subscales: Self-Affirmation (α=0.86) Outlook on Life (α=0.80)
Mitrani *et al.* (2005)[44] United States	ADRD	Structural Family Systems Ratings--Dementia Caregiver (SFSR-DC)	Family interaction patterns Two second- or higher-order factors: (1) Intimacy-conflict resolution (2) Freedom from negativity Six first-order factors: (Intimacy-Conflict Resolution) (1) Enmeshment (2) Care recipient	40 items, 5-point Likert scale (ranging from 1=least adaptive family functioning to 5=most adaptive family functioning)	Note: The respondent for the SFSR-DC scale is not a family CG. Instead, an experienced rater analyzes videos obtained from interactions between the family CG and the care recipient. Content validity was demonstrated by experienced raters reviewing the coding manual, rating five tapes together, and rating five tapes independently followed by meetings to reconcile discrepancies. The instrument developed during this stage had 67 items organized into eight "subscales". PCA followed by CFAs were conducted to study the structural validity of the initial 67-item SFSR-DC scale. The PCA with Varimax rotation yielded nine components/factors. Seven items with low loadings were eliminated resulting in a 60-item scale. A scree plot confirmed a nine-component structure. Iterative CFAs testing alternative models further eliminated items resulting in a 46-item first-order CFA with eight factors. Subsequent analyses and item deletions yielded a final hierarchical confirmatory factor model with two "second order" factors (labeled as "Intimacy-conflict resolution" and "Freedom	Interrater reliability (degree of agreement between different raters assessing the same data) was calculated with the ICC using the results from the 46-item first-order CFA model with eight factors. ICCs ranged from 0.617 to 0.937.

			disengagement (3) Conflict resolution (4) Expressed positive affect (Freedom from Negativity) (5) Identified Patienthood (6) Expressed anger		from negativity") and six first-order factors underlying a 40-item SFSR-DC scale. This hierarchical factor model yielded the best fit among competing models (e.g., CFI=0.981, RMSEA=0.048). Concurrent validity was demonstrated by significant (*p-value*s < 0.001) negative Spearman's correlations between the SFSR-DC "Intimacy-conflict resolution" second-order factor and (a) depression (rho= -0.30) and (b) anxiety (rho= -0.41). "Freedom from negativity" second-order factor was negatively and significantly associated with subjective burden (rho= -0.30). Depression was measured by the CES-D, anxiety was measured by the State Anxiety Inventory, and subjective burden was measured by RMBPC.	
Gitlin *et al.* (2005)[45] United States	ADRD	Caregiver Assessment of Function and Upset (CAFU)	CG reaction to physical dependence Two factors: (1) Activities of Daily Living (ADL) dependence and upset; (2) Instrumental Activities of Daily Living (IADL) dependence and upset	15 items, (Items were scored using two ordinal scales: Dependence Scale and Upset Scale) Dependence scoring scale 7-point scale of physical dependence (ranging from 7=Complete independence to 1=Complete help or more than 75% assistance) Upset scoring scale If physical dependence ≤ 5, then the CG was asked to rate the reaction/upset to providing assistance using a 5-point scale.	The CAFU scale was developed by combining items from two existing scales: eight items from Lawton and Brody's (1969)[46] instrumental ADL scale and seven items from Hamilton and Fuhrer's (1987)[47] Functional Independence Measure scale. (The CAFU scale was developed to measure both the dementia patient's level of physical dependence (functional needs) and the CG's reaction (emotional upset) to providing assistance with daily activities.) To assess the structural validity of the 15-item scale, the sample (N=640) was randomly split into two subsamples (each N=320) to perform a PCA and a CFA. The PCA with Varimax rotation yielded a two component solution explaining 54.7% of the variance. A scree plot confirmed the two components. CFA with the second subsample further established that the two-factor model was the best fitting model for the data (e.g., the goodness-of-fit index, GFI=0.98, the normed fit index, NFI=0.98, and the root mean square error of approximation, RMSEA=0.04). Concurrent validity was established by significant (*p-values* < 0.001) Spearman's rank correlations between CAFU scores and selected criterion measures. CAFU scores (using the Dependence scoring scale) were associated with both vigilance items: more hours on duty (rho=0.24) and more hours doing things for patients (rho=0.24). Greater CG "upset" (using the Upset scoring scale) was significantly correlated with a) more depression (rho=0.32) as measured by the CES-D scale and b) more problem behavior (rho=0.47), as measured by the RMBPC. Greater CG "upset" was also significantly associated with more hours of vigilance for the ADL activities subscale/factor (rho=0.43), but not for the IADL activities factor.	Cronbach's α by subscales: ADL dependence scoring (α=0.91) ADL upset scoring (α=0.83) ADL mean upset scoring per dependence (α=0.90) IADL dependence scoring (α=0.81) IADL upset scoring (α=0.80) IADL mean upset scoring per dependence (α=0.84)
Andrén & Elmståhl (2005)[48] Sweden	ADRD	Carers' Assessment of Satisfactions Index (CASI)	Subjective experience of satisfaction Four factors: (1) Purpose; (2) Pleasure (3) Appreciation; (4) Reverse	20 items, 4-point Likert scale (1=Does not apply, 2=Applies, but does not provide a source of satisfaction, 3=Applies and provides quite a lot of satisfaction, 4=Applies and provides a great deal of satisfaction)	This study explores an existing 30-item CASI scale developed by Nolan et al. (1996)[49] for CGs of relatives with common geriatric diseases and not specifically dementia. The current study validates CASI in a sample of dementia carers. The structural validity of the initial 30-item scale was established by factor analysis with Varimax rotation reducing the measure to 20 items, yielding four factors, and explaining 64% of the variance. According to the authors, this reduction of items resulted in a scale that was more specific to conditions of dementia. Concurrent validity was examined by Spearman's rank correlations between the CASI subscales and several criterion measures for assessing (a) patient dementia syndromes such as intellectual, emotional and motor performance, measured by the Gottfries-Brane-Steen (GBS) scale), (b) social dependency, measured by the Berger Scale), (c) CG stress management (measured by the Sense of Coherence Scale), d) burden, as measured by the Caregiver Burden Scale, and perceived health, assessed by the Nottingham Health Profile scale. Only the CASI Purpose subscale was associated with the patients' social dependency scores (rho=0.17, p<0.05) and intellectual syndrome (cognitive symptoms) scores (rho=0.168, *p*<0.05). Group discriminant validity. "Satisfaction", as measured by the CASI-Purpose subscale, was influenced by the patient's severity of disease. For the care recipient group with high independence (defined as low Berger score) CGs had higher mean scores in the Purpose subscale compared to the group of CGs caring for individuals with high dependence (23.4 vs. 20.4, *p* = 0.023).	Cronbach's α, full scale =0.78. Cronbach's α by subscales: Purpose (α=0.77) Pleasure (α=0.80) Appreciation (α=0.70) Reverse (α=0.83)
Kuhn *et al.* (2005)[50] United States	Mixed	Knowledge about Memory Loss and Care test (KAML-C)	Knowledge of memory loss, Alzheimer's, and related care Three subscales: (1) Medical; (2) Caregiving; (3) Legal and financial planning	15 items/questions, Each item has 5-response options with a single-correct answer. Example: *Which of the following is the most common cause of memory loss in people over age 65?* 1. Alzheimer's disease* (Correct answer) 2. Senility 3. Normal aging	Content validity. A preliminary survey of CGs of individuals in the primary stages of Alzheimer's disease helped identify three key knowledge domains about memory loss and related care: medical information, caregiving, and legal/financial planning. These domains guided the writing of 31 multiple-choice items by a panel of seven health professionals. The 31-item pool was administered to three different samples (family CGs, N=45); experts, N=37, and medical students, N=39). (The sample of medical students was included as a comparison to the experts and the CGs.) Item discrimination and difficulty indexes were calculated using the sample of experts and CGs (N=92). The initial 31-item pool was reduced to 15 items after inspecting a) item difficulty and discrimination and b) the difference in an item's difficulty prior to and following a five-week education program (pre- and post-test difference index, PPDI) aimed at improving knowledge about memory loss and related care issues among carers. Group discriminant validity was established by demonstrating the KAML-C's test ability to distinguish	Cronbach's α, full scale =0.76. Cronbach's α by subscales: Medical (α=0.46) Caregiving (α=0.61) Legal and financial planning (α=0.53) Note: The full scale, but not the subscales, showed a level of internal consistency considered acceptable, with a Cronbach's α value above 0.70.

				4. Hardening of the arteries 5. Benign senescent forgetfulness	between three groups: CGs, experts, and medical students. A Kruskal-Wallis test revealed significant differences between the scores of the three groups (p<0.0005) and in post-hoc tests groups scored in the expected order. Experts scored significantly higher than the other two groups (p<0.05), and medical students scored significantly higher than CGs (p<0.05).	
Gitlin *et al.* (2006)[51] United States	ADRD	Perceived Change Index (PCI)	State of wellbeing (CG appraisals of self-improvement or decline in distinct areas of wellbeing) Three factors: (1) Emotional wellbeing; (2) Physical wellbeing; (3) Ability to manage caregiving	13 items, 5-point Likert scale (1=Became much worse, 2=Became somewhat worse, 3=Stayed the same, 4=Improved somewhat, 5=Improved a lot over the past month)	Content validity was shown by conducting a literature review and drawing content for item development that reflected areas amenable to change, evidence of being a wellbeing concern and potentially, decline, as a consequence of caregiving, which could affect health. A 13-item pool was then administered to a sample of N=255 consisting primarily of women and non-spouses CGs. Using a split sample (N=127), structural validity was established by EFA with a PAF extraction method and a Varimax rotation that yielded a three-factor solution explaining 63% of the variance. Using the second half of the sample (N=128), concurrent validity was established by significant (*p-values* < 0.001) Pearson's correlations between PCI scores and a) the CES-D (r=-0.48), b) the Positive Aspects of Caregiving scale scores (r=0.41), and c) the Social Activities Index (r=0.43). Discriminant validity was shown by expected non-statistically significant Pearson's correlations of PCI scores with *characterizations of the patients' dementia using the MMSE scores* (r=0.01, ns) and *activities of daily living--functional independence* (r=0.07, ns).	Cronbach's α, full scale =0.90. (Using half of the sample, N=127) Cronbach's α by subscales: Emotional wellbeing (α=0.87) Physical wellbeing (α=0.79) Ability to manage caregiving (α=0.75)
Reilly *et al.* (2006)[52] United States	ADRD	Partner-Patient Questionnaire for Shared Activities (PPQSA)	Shared activities between CG and patient One factor: Relationship interference	17 items (activities), 5-point Likert scale (ranging from 0=Not at all to 4=Extremely) to measure the extent that patient mood or mental state interfered with the activity Note: Average PPQSA score is used as the "scoring method." CGs also rate the *importance* of the 17 activities and the *frequency* (# of activities that occurred in the past 24 hours or the past week.	Content validity was shown by item development through a literature review on CG burden, anticipatory grief, marital relations, and emotion constructs as well as consultation with an Alzheimer's disease clinician. This phase resulted in the development of a bank of 17 shared activities. Spouse and non-spouse CGs were asked to add activities and judge the frequency, importance, and interference in shared activities due to the patient's mood or mental state. Added activities did not differ conceptually from the originals, so the final PPQSA contained the same original 17 items, yet respondents' input did change item wording. The PPQSA structural validity was examined through a PCA with Varimax rotation. Authors split the sample into spouses (N=71) and non-spouses (N=29) and conducted separate PCA's in each group. Results were similar from both groups yielding one component/factor labeled as relationship interference. Some evidence in support of concurrent validity was provided by fitting a multiple regression model using PPQSA scale interference scores as the outcome measure and several criterion scores as explanatory variables while controlling for age, gender, and relationship to the patient. The following explanatory (criterion) variables were significant predictors (*p-values* < 0.001) of PPQSA scores: Caregiver Reaction Assessment, CRA, Work Productivity and Activity Impairment, and Time Spent Caregiving. All CRA domain scores were also significant predictors of the PPQSA score (*p-values* ≤ 0.02).	Cronbach's α estimates were high for the sample of spouses (0.95) and non-spouses (0.96).
Charlesworth *et al.* (2007)[53] United Kingdom	ADRD	Carers Assessment of Difficulties Index (CADI)	Objective burden Eight factors: (1) Carer's reaction to caring; (2) Degree of physical help; (3) CG-patient relationship; (4) Restrictions on social life; (5) Professional support; (6) Family support; (7) Interpersonal demands; (8) Financial consequences	30-item, 3-point Likert scale (1=Never applies, 2=Sometimes applies, 3=Always applies)	The CADI scale was originally developed by Nolan and Grant (1989) to assess multiple dimensions of carer burden. The original 30 items were identified from theoretical and empirical literature on caring representing aspects of social life, economic situation, relationship with the patient and wider family, professional and family support, dependency factors and the carer's reaction to the demands of caregiving. However, its psychometric properties had not been examined with dementia CGs. The current study validates the scale in a sample of N=232 dementia CGs. The structural validity of the 30-item scale was established by PCA with oblique (direct Oblimin) rotation. It yielded an eight-component/factor structure accounting for 59% of the variance. Evidence of group discriminant validity was shown by the sensitivity of the CADI scale to differentiate carers' age groups and gender. The overall 'objective burden' score (as measured by CADI total scores) was significantly higher for females than male's t (187) = -3.40, p <0.001. A significant negative Pearson correlation was found with age (r= -0.25, p < 0.01) and a positive correlation was found with duration of caring (r=0.273, p < 0.001).	Cronbach's α by subscales: Carer's reaction to caring (α=0.77) Degree of physical help (α=0.67)* CG-patient relationship (α=0.67)* Restrictions on social life (α=0.76) Professional support (α=0.68)* Family support (α=0.64)* Interpersonal demands (α=0.71) Financial consequences (α=0.69)*
Losada *et al.* (2008)[54] Spain	ADRD	Revised Familism Scale (R-FS)	Familism Three factors: (1) Familial obligations; (2) Perceived support from the family; (3) Family as referents	9 items, 5-point Likert scale (ranging from 1=Strongly disagree to 5=Strongly agree)	This study validates the previously developed Familism Scale (FS) in a sample of dementia CGs and confirms its original 3-factor structure. (The factor/component structure of the scale was originally assessed in a non-CG sample of 679 adults (452 Hispanics and 227 non-Hispanics) using a PCA. The current study used CFA techniques to examine the underlying dimensionality (structural validity) of the previous 14-item FS scale. After deleting five items due to low loadings, the CFA analysis confirmed the original 3-factor structure. The model fit indexes for the final 9-item Revised FS scale (R-FS) were within recommended thresholds (e.g., chi-square=40.17, df=26, p=0.04; chi-square/df= 1.55; GFI=0.94; CFI=0.96; and RMSEA=0.06). No further validity estimates for the R-FS scale were provided.	Cronbach's α, full scale =0.75. Cronbach's α by subscales: Familial obligations (α=0.59) Support from the family (α=0.75) Family as referents (α=0.75)
Cooper *et al.* (2008)[55]	ADRD	The Brief-Coping Orientation to Problems	Coping strategies Fourteen domains/subscales organized by three	28 items, (2 items per subscale) 4-point Likert scale	The original 60-item COPE scale was developed by Carver et al. (1989)[56] and later simplified to a 28-item Brief COPE scale by Carver (1997)[57]. However, the scales were validated in non-CG samples (undergraduate students and other adults). The current study validates and further simplified the Brief	Cronbach's α by "composite" subscale: Emotion-focused (α=0.72) Problem-focused (α=0.84)

United Kingdom		Experienced (Brief-COPE)	"composite subscales": A. Problem-focused (1) Active coping; (2) Use of informational support; (3) Positive reframing (4) Planning B. Emotion-focused (5) Emotional support; (6) Venting; (7) Humor; (8) Acceptance; (9) Religion; (10) Self-blame C. Dysfunctional coping (11) Self-distraction; (12) Denial; (13) Substance abuse; (14) Behavioral disengagement	(1=Not doing it at all, 2=A little bit, 3=A medium amount, 4=Doing it a lot)	COPE scale with a sample of dementia CGs. No study of the underlying factorial structure of the scale is conducted to establish structural validity. Concurrent validity was established by calculating Pearson's correlations between the Brief COPE composite scores and existing measures of a) patient functional impairment (assessed by the AD Co-Operative Study Inventory-Activities of Daily Living-ADL), b) relationship quality (number of confidants), and c) subjective attachment style (secure, avoidant, and anxious/ambivalent) measured by the "Attachment questionnaire"). As predicted, scores on the Brief-COPE Dysfunctional composite subscale were significantly associated with avoidant attachment (r=0.40, p<0.001). The Brief-COPE Emotion-focused composite scores correlated with number of confidants (r=0.29, p<0.001). Finally, the COPE Problem-focused composite scores correlated with ADL scores (r=-0.22, p<0.05). Note: The Psychometric properties of the Brief-COPE scale are studied both using total scores on the three separate composite subscales and using total scores on the Brief-COPE scale.	Dysfunctional (α=0.75) Test-retest reliability was established by calculating Pearson's correlations between total Brief COPE scores at one-year after (r=0.67) and two-years after (r=0.54) the first administration. In CGs whose ZBI scores remained "stable" between baseline and two-years after (change within 1 SD), total baseline COPE scores were associated with total scores at one and two-years after (r =0.72, 0.57). Test-retest reliability over a year was also demonstrated for emotion-focused (r=0.51), problem-focused (r=0.71), and dysfunctional (r=0.64) subscales.)
Menne *et al.* (2008)[58] United States	ADRD	Decision-Making Involvement Scale (DMIS)	Involvement in daily decision making One factor: Involvement in decision making (The tool measures the CG's perception of the day-to-day patient's decision making involvement.)	15 item, 4-point Likert scale (0=Not at all involved, 1=A little involved, 2=Fairly involved, 3=Very involved)	Content validity. Although content validity is not addressed in the current study, prior work is cited[59] on the underlying theories used for DMIS scale development and item adaptation to individuals with dementia and their family CGs. The structural validity of the 15-item DMIS scale was established by EFA with a PAF extraction method and Promax rotation. EFA yielded a unidimensional (one-factor) structure explaining 46.72% of variance. Concurrent validity was demonstrated by expected associations, calculated with Pearson's correlation coefficients, between total DMIS scores and a) depression, as measured by the CES-D (r= -0.16, p<0.05), b) quality of life, as measured by the Quality of Life-Alzheimer Disease scale (r=0.187, p<0.01), and c) relationship strain, measured by the Dyadic Relationship Scale (r=-0.221, p<0.01).	Cronbach's α, full scale =0.92.
Wilks (2008)[60] United States	ADRD	Shortened Resilience Scale (RS-15)	Resilience One factor: Global resilience	15 items, 7-point Likert scale (ranging from 1=Disagree to 7=Agree)	The 25-item RS was originally developed by Wagnild & Young (1993)[61] and evaluated in a national sample of community-dwelling older adults. The current study examines the psychometric properties of a shortened 15-item version in a dementia CGs sample. Structural and concurrent/convergent validity studies were conducted in two separate samples. Structural validity was established through EFA with PAF extraction that yielded a single resilience factor with an eigenvalue of 9.61 and explained 64% of the variance in items. Concurrent validity was demonstrated by significant (p-*values* < 0.01) Pearson's correlations between scores in the RS-15 scale and scores in the Perceived Stress Scale-10 (r= -0.60) as well as significant correlations with scores in the Perceived Social Support Family Scale (r=0.30) and Perceived Social Support Friends Scale (r=0.34).	Cronbach's α, full scale =0.89.
Wilks (2009)[62] United States	ADRD	Shortened Perceived Social Support Scale (S-PSSS): S-PSSS Family (SSfa) (Scale appraising family support)	Perceived social support as provided by family Three factors: (1) Relationship, Togetherness; (2) Moral, emotional support; (3) Openness, reliance	10 items, 5 point Likert scale (ranging from 0=Strongly disagree to 4=Strongly agree)	The PSSS Family and Friends independent "subscales", originally developed by Procidano & Heller (1983)[63] and later shortened by Maton et al, (1996)[64] were previously tested using data from undergraduates. Content validity examination was previously described.[64] The current study validates the two independent scales in a sample of ADRD CGs. To examine the structural validity of the Family and Friends scales independently, the sample of N=229 Alzheimer's CGs was randomly split into two samples. The first half (N=115) was administered the Family scale and the second half (N=114) was administered the Friends scale. Separate EFAs with PAF extraction and Varimax rotation were then conducted. Analysis of the independent samples produced the same underlying three-factor structure and similar patterns of factor loadings across factors. The proportion of variance explained was 74% for the Family scale. Concurrent validity was demonstrated by significant negative Pearson's correlations between scores in the S-PSSS "Family" scale and scores in the Perceived Stress Scale (r= -0.18, p<0.05) as well as significant positive correlations with scores in the Resilience Scale (r=0.15, p<0.05).	Cronbach's α estimate for Family scale was 0.89. Cronbach's α by subscales (Family scale): Relationship, Togetherness (α=0.82) Moral, emotional support (α=0.79) Openness, reliance (α=0.79) Guttman's split-half reliability estimate for the Family scale was 0.92.
		Shortened Perceived Social Support Scale (S-PSSS): S-PSSS Friends (SSfr) (Scale appraising friends support)	Perceived social support as provided by friends Three factors: (1) Relationship, Togetherness; (2) Moral, emotional support; (3) Openness, reliance	10 items, 5 point Likert scale (ranging from 0=Strongly disagree to 4=Strongly agree)	Structural validity. The EFA with the PAF extraction method and Varimax rotation also yielded a three-factor structure explaining 76% of the variance in items for the "Friends" scale. Concurrent validity was demonstrated by significant negative correlations between scores in the S-PSSS Friends scale and the Perceived Stress Scale (r= -0.26, p<0.05) as well as significant positive correlations with scores in the Resilience Scale (0.23, p<0.05).	Cronbach's α estimate, Friends scale =0.90. Cronbach's α by subscales (Friends scale): Relationship, Togetherness (α=0.86) Moral, emotional support (α=0.79) Openness, reliance (α=0.81) Guttman's split-half reliability, Friends scale = 0.94.

Carpenter *et al.* (2009)[65] United States	Mixed	The Alzheimer's Disease Knowledge Scale (ADKS)	Knowledge of Alzheimer's disease Seven domains: (1) Risk factors; (2) Assessment and diagnosis; (3) Symptoms; (4) Course; (5) Life impact; (6) Caregiving; (7) Treatment and management	30 items, 2-point, binary scale (0=False, 1=True)	Content validity. The Alzheimer's disease Knowledge Scale (ADKS) is an update to the 30-year-old Alzheimer's disease Knowledge Test (ADKT) developed by Dieckmann et al. (1988).[66] The team conducted a review of existing scales, evaluated the items, and assigned them to content domains. Differences were reconciled in a series of consensus conferences resulting in a preliminary bank of 49-items organized in seven content domains. Before studying the psychometric properties of the full scale, authors first analyzed individual item properties via item discrimination indexes, item difficulty indexes, and item homogeneity using split samples from the targeted mixed sample. Results were used to further reduce the scale to 30 items. The structural validity was studied by repeated PCAs with both unrotated and rotated components that yielded no simple structure or meaningful interpretation. Authors concluded it was best to interpret the ADKS as a scale of overall AD knowledge rather than a set of separately scored subscales or domains. Concurrent validity was established by a positive and significant Pearson's correlation between the new ADKS and the older ADKT (r=0.60, p<0.001) . Predictive validity was demonstrated by a significant Pearson's correlation between self-reported knowledge of AD with ADKS scores using the overall sample (N=763) (r=0.50, p<0.001). Correlations within the examined subsamples were also significant but "moderate": dementia CGs (r=0.46), older adults (r=0.41), dementia professionals (r=0.39), and students (r=0.20).	Cronbach's α, full scale =0.71 Test-retest reliability for a subsample (N=40) at an interval of 2 to 50 hours between tests (r=0.81, p<0.001). Guttman's split-half reliability estimate for the full scale=0.55 (p<0.001).
Czaja *et al.* (2009)[67] United States	ADRD	REACH Risk Appraisal Measure (RAM)	CG risk Six domains: (1) Depression; (2) Burden (3) Self-care and healthy behaviors; (4) Social support (5) Safety; (6) Patient problem behaviors	16 items, (Mixed scale formats) 2-point/binary scale (0=No, 1=Yes), 3-point Likert scale (0=Never to 2=Often), 4-point Likert scale (from 0=Not at all to 3=Very), 5-point Likert scale (from 0=Poor to 4=Excellent), 6-point Likert scale (from 0=Never to 5=Nearly always)	Content validity was established by a multisite working group generating items from a literature review of instruments and prior research. The working group identified six domains of risk and an initial 59-item pool. Further selection of items based on the identification of clear and good indicators for the six domains, relevant to diverse groups, and amenable to intervention reduced the item pool to 16 items. The concurrent validity of RAM was demonstrated by significant Pearson's correlations between scores in the RAM domains and at least one of the proposed criterion measures predicted to have an association with the domain. For example, scores on the Burden and Depression domains were significantly (p-*values* < 0.001) correlated with the Burden Interview scale (r=0.79 and r=0.45 respectively) and the CES-D (r=0.51 and r=0.68, respectively). Scores on the Self-Care domain correlated with the Self-Care Scale (r=-0.27) and Social Support domain scores were correlated with the Social Support Scale (r=0.68). Safety domain scores were, as expected, negatively associated with ADL/IADL (functional impairment) measures (r= -0.21). Finally, and scores on the Patient problem behaviors domain were significantly correlated with the Burden Interview scale (r=0.27).	Cronbach's α, full scale =0.65.
Montorio *et al.* (2009)[68] Spain	ADRD	Dysfunctional Thoughts about Caregiving Questionnaire (DTCQ)	Dysfunctional thoughts about caregiving Two factors: (1) Perception of sole responsibility; (2) Perfectionism	16-item, 5 point Likert scale (ranging from 0=Totally disagree to 4=Totally agree)	The Dysfunctional Thoughts about Caregiving Questionnaire (DTCQ) was originally developed by Losada (2005)[69] to assess specific dysfunctional thoughts and provide a single summary score indicating a "maladaptive approach" to caregiving. The present study examined the psychometric properties of the scale in a sample of dementia CGs. The structural validity of the 16-item DCTQ was established by PCA with oblique rotation that produced a two component/factor solution accounting for 47.7% of the variance in items. (The two factors/components labeled: Perception of sole responsibility and Perfectionism, explained 39.3% and 8.6% of the variance, respectively.) Concurrent validity was demonstrated by a significant positive Pearson's correlation between total DCTQ scores and scores in the Dysfunctional Attitudes Scale (r=0.58, p<0.001). DCTQ scores also were, as expected, significantly and negatively correlated with a) social support, measured by the Psychosocial Support Questionnaire (r=-0.21, p<0.01), b) the "amount of help received" question from socio-demographic variables (r=-0.25, p<0.001), and c) seeking emotional support (r=-0.23, p<0.001) and seeking instrumental support (r=-0.26, p<0.001) both measured by items from the Coping Orientation to Problems Experienced (COPE) scale. The discriminant validity of the DTCQ was analyzed by computing a correlation between total scores on DTCQ and the Frequency of Behavioral Problems subscale from the RMBPC. As expected, the correlation was not significant (r= −0.08, p=0.23).	Cronbach's α, full scale =0.89. Test-retest reliability for a subsample (N=31) at an interval of four weeks between tests was calculated using a Pearson's correlation (r=0.60, p<0.01).
Vickrey *et al.* (2009)[70] United States	ADRD	Caregiver-targeted quality-of-life (CGQOL)	CG Quality-of-Life Three higher order factors Ten domains/subscales: Tangible Assistance (1) Assistance in ADLs (2) Assistance in IADLs (3) Personal time (4) Role limitations due to caregiving	80 items, Items have different scales and response categories. Note: The 80 items are distributed across 10 subscales. The final scoring for the CGQOL scale recodes the initial	Content validity was established through focus groups and cognitive interviews of CGs from diverse ethnic groups to generate a pool of 91 items in 10 domains assessing aspects of CG quality of life. The structural validity of the CGQOL was established by iterative EFAs with Promax rotations. Guttman's weakest lower bound, Cattell's scree plot, and parallel analysis were examined to determine the number of factors. A final higher order factor analysis identified a three-factor solution influencing the 10-subscales or factors. The three higher order factors were interpreted as: *Tangible assistance*, *Psychosocial*, and *Benefits/faith*. Associations between the three factors ranged from 0.04 to 0.52. Multitrait scaling was used to examine item and subscale internal consistency estimates, item-scale correlations, and correlations among scales. This process reduced the scale from 91 to 80 items.	Cronbach's α by subscales: Assistance in ADLs (α=0.88); IADLs (α=0.93); Personal Time (α=0.78); Role Limitations (α=0.83); Family Involvement (α=0.86); Caregiving Demands (α=0.86); Worry (α=0.82); CG Feelings (α=0.94); Spirituality/Faith (α=0.92); Benefits of caregiving (α=0.89).

			Psychosocial (5) Family involvement (6) Caregiving demands (7) Worry (8) CG feelings Benefits/Faith (9) Spirituality and faith (10) Benefits of caregiving	response categories into a 0-100 rating where higher is better quality-of-life.	Multitrait-multimethod analysis was used to assess the validity of the scale by examining the correlations between multiple traits measured using the 10 subscales. Concurrent validity was demonstrated by significant negative correlations between hours-per-week caregiving and all subscales (r=0.14 to 0.68; *p-values* < 0.01) except Caregiving Benefits and Spirituality/Faith (0.092, *p*>0.05). The association of duration of being a CG and IADLs was also significant (= -0.192, *p*=0.007; r= -0.163, *p*=0.02).	Test-retest reliability (within 21 days) was calculated with the ICC with N=38. Test-retest reliability by subscales: Assistance in ADLs (ICC=0.86); IADLs (ICC=0.86);Personal Time (ICC=0.63); Role Limitations (ICC=0.53); Family Involvement (ICC=0.74); Caregiving Demands (ICC=0.72); Worry (ICC=0.53);CG Feelings (ICC=0.65); Spirituality/Faith (ICC=0.83); Benefits of caregiving (ICC=0.89)
Epstein-Lubow *et al.* (2010)[71] United States	Mixed	Caregiver Self-Assessment Questionnaire (CSAQ)	Stress and depression Two domains: (1) Stress; (2) Depression	18 item, (Mixed item scales) 2-point/binary scale (0=No, 1=Yes); 10-point Likert scale (ranging from 1=Not Stressful to 10=Extremely Stressful) or (from 1=Very healthy to 10=Very ill)	Note: The CSAQ was originally developed and tested by the American Medical Association (AMA) targeting a general population of family CGs. AMA reported a Cronbach's α reliability of 0.78 during scale development. To our knowledge, no further details on content validation and underlying factorial structure have been reported. The field study by Epstein and Lubow (2010) reported here examined the concurrent and predictive validity of the CSAQ scale in a sample of 106 predominantly (91.5%) dementia CGs. Assuming unidimensionality, a "total" score for the CSAQ was used to report the results. The concurrent validity of the CSAQ was demonstrated by a significant positive Pearson's correlation with the CES-D (r=0.807, *p*<0.001). Similar significant positive associations (all *p*<0.001) were found between CSAQ and a) stress measured by the Rapid Screen for Caregiver Burden (r=0.707), b) grief, measured with the Inventory for Traumatic Grief, Pre-Loss Version (r=0.594), and c) stress assessed with the Perceived Stress Scale-4-Item Version (r=0.682). CSAQ's scores sensitivity to predict significant depressive symptoms was 0.98, with a specificity= 0.52.	Cronbach's α, full scale =0.82.
Gough *et al.* (2010)[72] United States	ADRD	Intrinsic Spirituality Scale (ISS)	Spirituality One factor: Intrinsic spirituality	6 items, 11-point scale (ranging from 0=Spirituality answers no questions about life to 10=Spirituality answers absolutely all questions about life)	Content validity of the scale was established through literature reviews on caregiving burden and spirituality as a coping resource. Authors use the ISS scale, originally developed by Hodge, (2003)[73], and re-evaluate its content appropriateness for AD CGs within a theoretical framework of risk and resilience. The total sample (N=304) was randomly split to conduct factor and reliability analyses (N=152) and validity analyses (N=152). Structural validity was established by EFA with PAF extraction resulting in a single-dimension solution explaining approximately 70% of the variance. Concurrent validity was demonstrated by significant positive Pearson's correlations between ISS scores and a) frequency of prayer (r=0.50), b) the Private Prayer as a Means of Coping (UPPMC, r=0.65)), Using Private Prayer as a Means of Coping scores (r=0.65), and c) the Connor-Davidson Resilience Scale scores (r=0.44). The correlation of ISS scores with the ZBI scores was not significant (r=0.06, *p*=0.31). Discriminant validity. ISS scores were not associated with relation to care recipient (r-0.07, *p*=0.43).	Cronbach's α, full scale =0.919. Guttman's split-half reliability estimate for the full scale=0.914.
Losada *et al.* (2010)[74] Spain	ADRD	Caregiver Guilt Questionnaire (CGQ)	Guilt Five factors: (1) Guilt about doing wrong by the care recipient; (2) Guilt about not rising to the occasion as CGs; (3) Guilt about self-care; (4) Guilt about neglecting other relatives; (5) Guilt about having negative feelings towards other people	22 item, 5-point Likert scale (0=Never, 1=Rarely, 2=Sometimes, 3=Several times, 4=Always or almost always)	Content validity was established by a literature review on guilt-related constructs and expert panel review of items resulting in an initial pool of 34 items. The structural validity was established by PCA using Varimax rotation that yielded a five-factor/component solution in a final 22-item tool that explained 59.25% of the total variability present in the total data set. Concurrent validity was demonstrated by significant positive correlations (*p-values* <0.01) between CGQ scores and a) guilt (r=0.46), measured by the ZBI Guilt factor, b) depression (r=0.46), measured by the CES-D, c) anxiety (r=0.46) measured by the Profile of Mood States Tension-Anxiety subscale, and d) both behavioral problem appraisal (r=0.51) and frequency (r=0.42) measured by the Revised Memory and Behavior Problems Checklist. In addition, there was a significant negative correlation (*p*<0.01) between CGQ scores and social support (r=-0.19, *p*<0.01), as measured by the Psychosocial Support Questionnaire.	Cronbach's α, full scale =0.88. Cronbach's α by subscales: Guilt about doing wrong by the care recipient (α=0.89) Guilt about not rising to the occasion as CGs (α=0.76) Guilt about self-care (α=0.69) Guilt about neglecting other relatives (α=0.86) Guilt about having negative feelings towards other people (α=0.61)
Wimo *et al.* (2010)[75] Sweden	ADRD	Resource Utilization in Dementia (RUD)	Informal caregiving time Three domains: (1) Basic Activities of Daily Living (ADL; e.g., eating, dressing, bathing) (2) Instrumental Activities of Daily Living (IADL; e.g., cooking, cleaning, budgeting) (3) Supervision/Surveillance	3 "items" or domains, Note: Scoring in RUD consists of CG recollections of time (e.g., minutes) spent on activities in each the 3 "items" or domains: ADL, IADL, and Supervision.	The content validity of RUD has been previously established in an institutional care setting.[76] The current study validates the accuracy of the caregiver time estimates provided with the RUD and tests its validity and reliability in a community care setting with dementia caregivers. Concurrent validity was shown by expected significant (*p-values* < 0.001) positive Pearson's correlations between CG estimates (recall) of time spent on caregiving activities (i.e., RUD scores) and *the time observed by a nurse*. Correlations between recalled and observed times were reported for the total scale (r=0.69) and each subscale: ADL (r=0.81), IADL (r=0.68), and Supervision (r=0.67). Note: Time spent caregiving was recorded in three ways: diary, observation, and recall. The CG recorded activities and their duration (in minutes) in a 24-hour diary period. CG recollections of activities and their duration were estimated after each diary period. *Nurse observations* were made in four-hour	Inter-rater reliability was calculated with the ICC for all recalled times (i.e., the full RUD) compared to diary was ICC=0.91 and compared to observation was ICC=0.80. Inter-rater reliability by subscales: ADL: Recalled versus diary (ICC=0.93) and versus observation (ICC=0.81) IADL: Recalled versus diary (ICC=0.85) and versus observation (ICC=0.74)

			(e.g., preventing dangerous episodes and managing behavioral problems)		sessions. CG recollections of activities and their duration (recall using RUD) were estimated after each observation session.	Supervision: Recalled versus diary (ICC=0.87) and versus observation (ICC=0.78)
Yap *et al.* (2010)[77] Singapore	ADRD	Gain in Alzheimer Care INstrument (GAIN)	Positive outcomes from caregiving; One factor: Gain, personal growth	10 items, 5-point Likert scale (ranging from 0=Disagree a lot to 4=Agree a lot)	Content validity established by deriving items and themes from a qualitative study of CGs and from focus groups of CGs confirming the preliminary pool of identified items. The structural validity was assessed by a PCA that yielded one component accounting for 52.8% of the total variability present within the original dataset. Concurrent validity was demonstrated by significant positive correlations between the GAIN scale scores and a) Positive Aspects of Caregiving (r=0.68, *p*<0.001) and b) both active/engaged management (r=0.42, *p*<0.001) and encouragement (r=0.35, *p*<0.001) subscales of the Dementia Management Strategies Scale (DMSS). GAIN scores were significantly and negatively correlated with scores on the criticism subscale (r=-0.14, *p*<0.05) of the DMSS and the ZBI scores (r=-0.15, *p*<0.05).	Cronbach's α, full scale =0.89. Test-retest reliability (2-week interval) was assessed with the ICC using a subsample (N=149) of participants. (ICC=0.70)
Savundranayagam *et al.* (2011)[78] United States	ADRD	Montgomery Borgatta Caregiver Burden Scale (MB-CBS)	CG burden Three factors: (1) Objective burden; (2) Subjective demand or relationship burden; (3) Subjective stress burden	14 items, 5-point scale (1=A lot less, 2=A little less, 3=The same, 4=A little more, 5=A lot more)	The study uses two Independent groups of family CGs of persons with dementia (spouses and children) to study underlying structure and psychometric properties of the MB-CBS scale across groups. Authors adopted the factorial structure proposed by Montgomery et al. (2000);[79] the original developers of the MB-CBS scale. The current study did not examine scale dimensionality in the AD CG sample. Measurement invariance tests using multiple group CFAs were conducted with the full sample (N=523). Results revealed that the MB-CBS factor structure had *configural* and *metric* invariance across the samples of caregiving spouses and adult children in the measurement of stress burden, relationship burden, and objective burden confirming the same factor structure and that the association between each item and the latent construct it measures per subscale is the same across spouses and adult children. That is, the interpretation of scale items can be considered consistent across these two groups of carers. Note: To provide some evidence of "criterion validity", authors test hypothesized relationships between the subscales and known caregiving burden measures fitting two separate structural equations models. The results showed that the MB-CBS-objective burden subscale and ADLs were significantly associated. Problem behavior scores were also significantly associated with all three MB-CBS burden factors. Both analyses with the spouses and children samples yielded the same pattern of results.	*Spouses*: Cronbach's α by subscales: Objective burden (α=0.85) Relational burden (α=0.87) Stress burden (α=0.86) *Children*: Cronbach's α by subscales: Objective burden (α=0.93) Relational burden (α=0.89) Stress burden (α=0.90)
Werner *et al.* (2011)[80] Israel	ADRD	Family Stigma in Alzheimer's Disease Scale (FS-ADS): Scale 1: Family Stigma	CG's stigma Eight factors/components: (1) Esthetics; (2) Shame; (3) Pity; (4) Fear; (5) Concealment from professionals; (6) Concealment from friends; (7) Helping with ADL/IADL; (8) Concealment from family	18 items, 5-point scale (ranging from 1=Lowest to 5=Highest)	Content validity. Authors report identifying an initial pool of 100 items from the literature and an earlier qualitative study including three dimensions representing/defining the scales (CGs' stigma, lay persons' stigma, and structural stigma). The structural validity of the FS-ADS was established with PCA and Varimax rotations to increase interpretability. Using the same sample of participants (N=185), the PCA analysis was conducted separately (and iteratively) in each of the three scales. For the Caregiver's Stigma scale, the final PCA yielded an 8-factor structure of an 18-item scale that explained 88% of the variance. Concurrent validity was demonstrated by significant positive Pearson's correlations (*p-values* < 0.05) between the ZBI and the following factors of the Caregiver's stigma scale: a) Esthetics (r=0.27), b) Shame (r=0.41), c) Fear (r=0.31), d) Pity (r=0.18), and e) ADL/IADL (r=0.38). Further evidence was shown by significant positive correlations between the Problematic Behavior Scale and the factors of Esthetics (r=0.30), Share (r=0.24), and ADL/IADL (r=0.27).	Cronbach's α by subscales: Esthetics (α=0.97) Shame (α=0.97) Pity (α=0.80) Fear (α=0.95) Concealment from professional (α=0.81) Concealment from friends (α=0.66) Helping with ADL/IADL (α=0.70) Concealment from family (α=0.41)
		Family Stigma in Alzheimer's Disease Scale (FS-ADS): Scale 2: Lay persons stigma	Lay persons stigma Nine factors/components: (1) Esthetics; (2) Cognitive functioning; (3) Distancing; (4) Willingness to help; (5) Pity/uneasiness; (6) Physical functioning; (7) Fear; (8) Shame; (9) Disgust	28 items, 5-point scale (ranging from 1=Lowest to 5=Highest)	Structural validity. A PCA approach yielded a 9-component/factor solution of the 28-item scale that explained 88% of the variance. Concurrent validity was demonstrated by significant positive Pearson's correlations coefficients between the ZBI and the following factors of the Lay person stigma scale a) Cognitive functioning (r=0.16, *p*<0.05), b) Physical functioning (r=0.19, *p*<0.05), c) Esthetics (r=0.25, *p*<0.01), d) Fear (r=0.25, *p*<0.01), e) Disgust (r=0.27, *p*<0.001), and f) Distancing (r=0.31, *p*<0.001). Further evidence was shown by the significant positive correlations between the Problematic Behavior Scale and a) Cognitive functioning (r=0.15, *p*<0.05), b) Physical functioning (r=0.35, *p*<0.001), c) Esthetics (r=0.30, *p*<0.001), d) Fear (r=0.15, *p*<0.05), e) Disgust (r=0.19, *p*<0.01), and f) Distancing (r=0.28, *p*<0.001).	Cronbach's α by subscales: Cognitive functioning (α=0.98); Disgust (α=0.95); Distancing (α=0.98); Esthetics (α=0.99); Fear (α=0.93); Physical functioning α=0.88); Pity/Uneasiness (α=0.81); Shame (α=0.97); Willingness to help (α=0.98)
		Family Stigma in Alzheimer's Disease Scale (FS-ADS): Scale 3: Structural stigma	Structural stigma Two factors/components: (1) Structural stigma; (2) Professionals' relationship	16 items 5-point scale (ranging from 1=Lowest to 5=Highest)	Structural validity. A PCA approach to factor extraction yielded an 2-factor/component solution of a 16-item scale that explained 72% of the variance. Concurrent validity was demonstrated by significant Pearson's correlation coefficients between the ZBI and the Structural stigma (r=-0.33, *p*<0.001) and Professionals' relationship (r=0.22, *p*<0.002) factors. Significant Pearson's correlations were also obtained between the Problematic Behavior Scale and a) Structural stigma factor (r=-0.25, *p*<0.001) and b) Professionals' relationship factor (r=0.24, *p*<0.001).	Cronbach's α by subscales: Structural stigma (α=0.96) Professionals' relationship (α=0.88)
Erder *et al.* (2012)[81]	ADRD	Caregiver-Perceived	Caregivers' Assessment of the Patient (CAP)	20 items, Likert scale (cut-points or	Content validity. The assessment goals and measurement domains of the CPBQ were initially informed via input from clinicians experienced in treating Alzheimer's disease (AD). The domains were as follows:	Cronbach's α, full scale =0.88. Test-retest reliability after a 4-week

United States		Burden Questionnaire (CPBQ): Scale 1: Caregivers' Assessment of the Patient (CAP)	(Caregiver-perceived patient functional engagement) Three "factors" from the EFA analysis (not labeled) (Rasch analysis suggested a unidimensional (one-factor) construct.)	thresholds not provided)	Functional Communication, Social Abilities, and Executive Functioning. The initially analyzed CPBQ item pool consisted of 42 items. An EFA with oblique rotation was first conducted with the entire 42-CPBQ item pool using a split-half sample from the total N=676. Based on further review of the results and the item content analysis, the CPBQ was divided into 2 scales: a 29-item Caregivers' Assessment of the Patient (CAP) scale and a 13-item Caregivers' Assessment of Themselves (CAT) scale. The structural validity for CAP was established by EFA on a split-half sample yielding a 3-factor structure. After deleting items with low loadings, 20-items were retained for CAP. A CFA was executed on the second split-half sample. The model failed tests of comparative fit index (CFI=0.863), root mean square error of approximation (RMSEA=0.073), and standardized root mean square residual (SRMR=0.065), but "items were judged by the experts as the most plausible and meaningful". Next, a Rasch analysis of the CAP scale was conducted showing good overall fit suggesting that it measured a single underlying construct, as the Rasch model assumes unidimensionality. Concurrent validity was shown by significant Spearman's rank correlations (*p-values*<0.001) between the CAP and the NPI (rho=0.38), the Severe Impairment Battery (rho= -0.45), the Alzheimer's Disease Cooperative Study-ADL Scale (rho= -0.57), the Clinician's Interview-Based Impression of Change-Plus Caregiver Input (rho=0.45), and the Functional Assessment Staging Tool (rho=0.36).	interval was estimated with an ICC=0.83. PSI (internal consistency under the Rasch model) estimate for the full scale=0.89.
		Caregiver-Perceived Burden Questionnaire (CPBQ): Scale 2: Caregivers' Assessment of Themselves (CAT)	Caregivers' Assessment of Themselves (CAT) (Caregiver-perceived burden in relation to the patient's engagement) Two "factors" from the EFA analysis (not labeled) (Rasch analysis suggested a unidimensional (one-factor) construct.)	10 items, Likert scale (cut-points or thresholds not provided)	The structural validity for CAT was established by EFA with on a split-half sample yielding a 2-factor structure. After deleting items with low loadings, 10-items were retained for CAT. A CFA was conducted on the second split-half sample. The model produced a satisfactory fit (e.g., CFI=0.918, RMSEA=0.084, and SRMR=0.056) yet again "items were judged by the experts as the most plausible and meaningful". Also, the Rasch analysis of the 10-item CAT scale showed good overall fit suggesting a single (unidimensional) construct. Concurrent validity was demonstrated by significant Spearman's rank correlations (*p*<0.001) between the CAT and the NPI (rho=0.35), the Severe Impairment Battery (rho=-0.19), the Alzheimer's Disease Cooperative Study-ADL Scale (rho=-0.24), the Clinician's Interview-Based Impression of Change-Plus Caregiver Input (rho=0.23), and the Functional Assessment Staging Tool (rho=0.14).	Cronbach's α, full scale =0.83. Test-retest reliability after a 4-week interval was calculated with the ICC=0.58. PSI (internal consistency under the Rasch model) estimate for the full scale=0.83.)
Quirk *et al.* (2012)[82] United Kingdom	Mixed	The Carer Well-being and Support (CWS) questionnaire	Quality of life/sense of meaning Two Factors: (1) Wellbeing; (2) Social support	49 items, 4-point scale (0=Very dissatisfied, 1=Somewhat dissatisfied, 2=Somewhat satisfied, 3=Very satisfied) or 5-point scale (0=Poor, 1=Fair, 2=Good, 3=Very good, 4=Excellent	Content validity was demonstrated by conducting workshops with carers for people with psychosis (N=5), common mental health problems (N=10) and dementia (N=8) and collecting feedback on how to improve a preliminary bank of items. As a result, a 74-item CWS measure was developed. After a "preliminary" pilot field test with a sample of 210 participants and an EFA, the CWS measure was further reduced to 49 items. The CWS was field tested using an independent sample of 361 carers. The structural validity was established through an EFA of the 49-item scale that produced a 2-factor structure (a 32-item well-being and a 17-item social support) accounting for 50.8% of the variance. Concurrent validity. The wellbeing subscale showed a "large" Pearson's correlation with the General health questionnaire, GHQ-12 (r=-0.66, p < 0.001) and the Involvement evaluation questionnaire, IEQ-EU (r=-0.70, p < 0.001). Discriminant validity. Wellbeing and support subscales were, as expected, uncorrelated with the age of the carer (r=0.14, ns).	Cronbach's α by subscales: Wellbeing (α=0.96) Social support (α=0.97) Test-retest reliability (2-week interval) was calculated with the ICC using a subsample (N=92). ICC by subscales: Wellbeing (ICC=0.92) Social support (ICC=0.88)
Riley *et al.* (2013)[83] United Kingdom	ADRD	Birmingham Relationship Continuity Measure (BRCM)	Relationship continuity One factor: (Items cover the following domains: Relationship redefinition, Same/different person, Same/different feelings, Couplehood, loss of relationship)	23 items, 5-point Likert scale (1=Disagree a lot, 2=Disagree a little, 3=Neither, 4=Agree a little, 5=Agree a lot)	Content validity was established by qualitative research on relationship continuity leading to a 42-item measure pilot tested on a sample of 51 spousal CGs. The structural validity of the BRCM was established through an EFA with PAF for factor extraction and Oblimin rotation producing a single-factor structure accounting for 46% of the variance in scores. A scree plot confirmed a one-factor structure. Concurrent validity was demonstrated by a significant positive Pearson's correlation coefficient between BRCM scores and the Closeness and Conflict Scale (r=0.411, *p*=0.002) and a significant negative correlation with the Heartfelt Sadness and Longing subscale of the Marwit-Meuser Caregiver Grief Inventory (r=-0.641, *p*<0.001).	Cronbach's α, full scale =0.947. Test-retest reliability was calculated (at one to three-week interval) using the ICC in a subsample (N=34) of participants (ICC=0.932).
Lopez & Guarino (2013)[84] United States	ADRD	Surrogate Decision Making Self-Efficacy Scale (SDM-SES)	Self-efficacy for decision making One factor: Self-efficacy	5 items, 4-point Likert scale (ranging from 1=Strongly disagree to 4=Strongly agree)	Face/content validity was established by three expert Gerontological nurses who reported on the instrument's credibility, accuracy, and relevance as a measure of self-efficacy for surrogate decision making. The reliability of agreement between the three experts was assessed with Fleiss' kappa coefficient (Fleiss' kappa=0.90). The structural validity of the scale was established through CFA of a hypothesized single underlying latent factor model for self-efficacy for decision making explaining the set of observed items. As expected, CFA produced a single-factor (unidimensional) model with factor loadings ranging from 0.63 to 0.86. The model goodness-of-fit measures were acceptable (CFI=0.99; TLI=0.98).	Cronbach's α, full scale =0.87
Tebb et al (2013)[85]	Mixed	The Caregiver Well-Being Scale:	Basic Needs Three factors:	8 items, 5-point Likert scale (from	Content validity was established by a review of the initial 43-item pool of the Caregiver Well-Being Scale (CWBS) by an expert panel (5 psychometricians and 1 social worker) and a lay panel (10 family CGs of	Cronbach's α estimate for the Basic Needs scale=0.73.

Canada		Short Form Rapid Assessment: Basic Needs Scale	(1) Emotional Needs; (2) Physical Needs (3) Self-Security	1=Rarely to 5=Usually)	people with Alzheimer's disease). As a result of the review, the original 43-item scale was further reduced to a 16-item scale. The current study reports on the validation of two subscales identified in the original CWBS measure using a mixed sample that included dementia CGs. The structural validity by subscale was estimated with a CFA. Using the same sample of CGs (N=486), the two subscales ("Basic Needs" and "Activities of Living") from the full 16-item Caregiver Well-Being Scale (CWBS) were analyzed using two separate CFAs to test whether each subscale was conceptually distinct and psychometrically valid as a stand-alone scale, and whether it reliably measured the specific construct it was intended to capture within the larger CWBS scale. For the Basic Needs scale , the model fit the data well (e.g., RMSEA=0.05; CFI=0.97, and TLI=0.95).	Note: Cronbach's α estimate for the full CWBS scale=0.83.
		Activities of Daily Living Scale	ADLs Three factors: (1) Self-Care; (2) Connectedness; (3) Time for Self	8 items, 5-point Likert scale (from 1=Rarely to 5=Usually)	Structural validity. For the Activities of Daily Living scale, the CFA analysis revealed that the hypothesized model fit the data. Fit indexes were acceptable (e.g., RMSEA=0.07, CFI=0.95, and TLI=0.92).	Cronbach's α estimate for the Activities of Daily Living scale=0.74. Note: Cronbach's α estimate for the full CWBS scale=0.83.
Bekhet & Zauszniewski (2013)[86] United States	ADRD	Depressive Cognition Scale (DPS)	Depressive cognitions One factor: Depressive cognitions	8 items, 6-point Likert scale (ranging 0=Strongly disagree to 5=Strongly agree)	The content validity of the scale was previously established by Zauszniewski et al., 2002[87]. The current study examined the structural validity of the scale with a PCA in a sample of ADRD CGs that resulted in two factors/components. Authors follow-up with a CFA that produced a single factor explaining 55.99% of the variance. This solution confirmed previous findings using the scale. The concurrent validity was assessed through an expected positive Pearson correlation between DPS scores and Caregiver burden (r=0.40, p<.001) measured by the ZBI and a significant negative correlation with resourcefulness (r= -0.54, p<.001) as measured by the Resourcefulness Scale.	Cronbach's α, full scale =0.88.
Orgeta *et al.* (2013)[88] United Kingdom	ADRD	Warwick–Edinburgh Mental Well-Being Scale (WEMWBS)	Mental wellbeing One factor: (Items cover the following domains: affective-emotional aspects, cognitive-evaluative dimensions, and psychological functioning.)	14 items, 5-point Likert-type scale (1=None of the time to 5=All of the time)	The structural validity was shown by a PCA that yielded a single-factor structure explaining 57% of the variance. Concurrent validity was established by significant negative correlations between WEMWBS scores and (a) anxiety (r=-053, p<0.001) and depression (r=-0.50, p<0.001) measured by the HADS (b) dysfunctional coping strategies (r=-0.51, p<0.001) measured by the Coping Orientations to Problems Experienced Scale, and (c) stress (r=-0.63, p<0.001) measured by the Relative's Stress Scale. Further proof of concurrent validity was provided by significant positive correlations with physical health (r=0.63, p<0.001), measured by the EuroQoL-Visual Analogue Scale, and social support (r=0.39, p<0.01), measured by the Multidimensional Scale of Perceived Social Support.	Cronbach's α, full scale =0.83.
Wilks *et al.*(2013)[89], United States	ADRD	Spiritual Support Scale (SSS)	Perceived spiritual support One Factor: (Items measure the use of spiritual support as a form of coping.)	12 items, 4-point Likert scale (1=Strongly disagree to 4=Strongly agree)	The structural validity was demonstrated by an EFA with Varimax rotation that yielded a single-factor structure explaining 79% of the variance by a rotated Varimax solution. The concurrent validity of the SSC scale was established by significant positive correlations with (a) the Task-Focused subscale of the Coping in Task Situations (CITS) measure (r=0.12, p<0.01) and (b) the Resilience Scale (r=0.25, p<0.01). Validity was also supported by a significant negative correlation between SSS scores and the Emotion-Focused subscale of the CITS measure (r=-0.12, p<0.01). However, SSS scores were not significantly correlated with the CITS's Avoidance-Focused subscale.	Cronbach's α, full scale =0.974 Split-half reliability was estimated by Guttman's coefficient showing a strong correlation between two random halves of the measure (Guttman's split-half reliability=0.940).
Crellin *et al.* (2014)[90] United Kingdom	ADRD	Caregiver Efficacy Scale (CES)	CG efficacy for managing behavioral and psychological symptoms in dementia Three factors/components: (1) Mood and hyperactivity; (2) Psychosis and nighttime disturbance; (3) Euphoria	12-items, 4-point Likert scale (ranging from 4=Not at all confident to 1=Very confident)	Content validity. Based on a literature review on the link between self-efficacy and experiences of CGs of individuals with dementia and their ability to cope with behavioral and psychological symptoms of dementia (BPSD), the CES was developed by the addition of a single item to each of the 12 domains of BPSD in the Neuropsychiatric Inventory (NPI).[91] CGs reporting the presence of a behavioral disturbance also reported their self-efficacy in dealing with the problem. The structural validity was established through PCA with Oblimin rotation to improve components interpretability and a scree plot examination to determine the number of components/factors. The PCA yielded a 3-factor/component solution accounting for 49.85% of the variance. Concurrent validity was evaluated using Spearman's rank correlations between the CES scores and the subscales of the Revised Scale for Caregiving Self-Efficacy: "obtaining respite" (rho=−0.268, p< 0.001), "responding to disruptive behavior" (rho=−0.386, p < 0.001), and "controlling upsetting thoughts" (rho= −0.384, p < 0.001). Highly significant correlations were also obtained between CES scores and the NPI subscales.	Cronbach's α, full scale =0.79.
Cole *et al.* (2014)[92] United States	ADRD	Impact of Alzheimer's Disease on Caregiver Questionnaire (IADCQ)	Caregiver burden One factor (Items cover the following domains: Caregiver burden across emotional, physical, social, financial, sleep, and time impact)	12-items, 5-point Likert scale (ranging from 0=Not at all to 4=Extremely)	Content validity. No formal statements on content validity are made. However, authors reported item generation being informed by reviewing the literature and identifying previous measures on AD caregiving burden and quality of life. Three focus groups were held to better understand the experience of caring for a patient with AD and to conduct a cognitive debriefing of an initial 9-item draft of the IADCQ. CGs provided input on the questions, response options, and instructions resulting in a revised 12-item IADCQ instrument. The structural validity of the 12-item IADCQ was assessed through a CFA that resulted in a final one-	Cronbach's α, full scale =0.927. Test-retest reliability (4-week interval) was assessed with the ICC using a subgroup of AD CGs (N=50). The ICC was moderate (0.68).

					factor (unidimensional) solution that provided acceptable goodness-of-fit indexes (e.g., GFI=0.934; RMSEA=0.076; CFI=0.944; and SRMR = 0.040). Concurrent validity was assessed through "moderate to large" Pearson's correlations between IADCQ scores and the Short Form-12 Health Survey (SF-12: V2) composite scores scales: Physical health (r= -0.26, p < 0.001) and Mental health (r= -0.58, P < 0.001). Pearson's correlations between IADCQ scores and other subscales from the SF-12: V2 were also "moderate to large" ranging from -0.20 to -0.57.	
Gillanders *et al.* (2014)[93] United Kingdom	ADRD	Cognitive Fusion Questionnaire (CFQ)	Cognitive fusion One factor (Items cover the following domains: Dominance of cognitive events in a person's experience, emotional reactions to thoughts and beliefs, and ability to view cognitive events from a different perspective	7 items, 7-point Likert scale (1=Never true, 2=Very seldom true, 3=Seldom true, 4=Sometimes true, 5=Frequently true, 6=Almost always true, 7=Always true)	Content validity. Experts from the British Association for Behavioral & Cognitive Psychotherapy acceptance and commitment therapy Special Interest Group were asked to comment on item clarity and rate how well the initial pool of 44 items (developed by the authors) represented cognitive fusion and defusion. The final revised scale had 42 items. Structural validity was first examined through iterative EFA with oblique rotations and Horn's parallel analyses to determine the number of underlying factors using a sample (N=592) of younger adults (not dementia CGs). After removing items with low loadings, only 7 items were retained in a final one-factor scale. Independent CFA models were subsequently estimated using five different samples of CGs. The results for the sample of dementia CGs presented here yielded acceptable goodness-of-fit indexes for the one-factor structure (e.g., RMSEA=0.101; CFI=0.962; and IFI=0.963). A measurement invariance test across the five samples supported metric invariance making it possible to meaningfully compare mean CFQ scores between the five groups of CGs on the underlying construct. Concurrent validity in the sample of dementia CGs: CFQ scores were significantly associated with scores on the CES-D (r=0.66, *p* < 0.001)	Cronbach's α, full scale =0.88.
Liu et al. (2014)[94] Taiwan	ADRD	Finding a Balance Scale (FBS)	Balance between the demands of caregiving and other competing needs One factor (A single factor is assumed; no analyses are conducted to determine the underlying structure of the scale.)	17 items, Items 1-17 (competing needs) 4-point Likert scale (0=Unable to handle either, 1=Able to handle only one, 2=Able to handle both, but not well, 3=Usually able to handle both well).	Content validity. Evidence of content validity is reported in a previous study while developing the scale for CGs of frail elders. The original scale was reviewed by a clinician, two sociologists, and three nurses who reported acceptable content validity. Structural validity. No formal analysis to assess the underlying structure of the 17 items in the FBS scale is presented with the current sample of dementia CGs. A unidimensional structure seems to be assumed. Concurrent validity was assessed by calculating Pearson's correlation coefficients between FBS total scores and (a) the Role Strain Scale (r=-0.48, p < 0.01), (b) SF-36-Physical health, SF-36-Physical Component (r=0.20, p < 0.01), and (c) the SF-36-Mental health (r = 0.44, p < 0.01). Discriminant validity was supported by the expected absence of a significant correlation between FBS total scores and total scores on the Mutuality Scale (r=0.04, p=.61). (The Mutuality scale measures the quality of the CG–care receiver relationship.) Group discriminant validity was shown by comparing a "well-balanced group" (FBS scores >2) with a "poor balance group" (FBS≤2) on role strain and mental health scores. As expected, an independent samples t-test showed that the well-balanced group had significantly lower Role Strain (*t*=-5.72, p < 0.01) and better SF-36-Mental health (*t*=7.07, p < 0.01) than those in the poorly balanced group.	Cronbach's α, full scale =0.92.
Losada et al. (2014)[95] Spain	ADRD	Experiential Avoidance in Caregiving Questionnaire (EACQ)	Experiential avoidance Three factors: (1) Active avoidant behaviors; (2) Intolerance of negative thoughts/emotions toward care recipient; (3) Apprehension concerning negative internal experiences related to caregiving	15 items, 5-point Likert scale (1=Not at all, 2= A little, 3=Somewhat, 4=Often 5=A lot)	Content validity. Based on a literature review and a previously developed scale measuring experiential avoidance, a pool of 15 items was developed and tested in a sample of 44 dementia CGs. As a result, changes were made to both item content and response options. Structural validity was established via PCA with Oblimin rotation and a scree plot to determine the optimal number of components. The PCA yielded a 3-factor solution explaining 44.5% of the total variance. Concurrent validity was assessed through Pearson's correlations between the total EACQ scores and (a) the Acceptance and the Action Questionnaire (AAQ) (r=0.14, p <0.05) (b) the dysfunctional thoughts about caregiving questionnaire (DTCQ) (r=0.22, p <0.01) and (c) the POMS-Tension-Anxiety subscale (r=0.14, p <0.01) Discriminant validity of the EACQ subscales is shown by fitting a series of a hierarchical regression models entering the factors one at a time and determining whether there was a significant incremental change in percentage of explained variance indicating a unique/distinct factor-specific contribution to the scale. A significant incremental change in percentage of explained variance was found for each of the EACQ factors, indicating an estimate of the unique, construct-specific contribution of each factor.	Cronbach's α, full scale =0.70. Cronbach's α by subscales: Active avoidant behaviors (α=0.63)* Intolerance of negative thoughts/emotions toward care recipient (α=0.71) Apprehension concerning negative experiences (α=0.60).
Solberg et al. (2014)[96] United States	ADRD	Caregiver Stress Impact Scale (CGQ-13)	Impact of stress on primary caregivers (adult children) One factor: Impact of stress	13 items, 3-point Likert type scale with varying labels.	Content validity. Authors developed a 32-item pool based on a literature review of the stress experienced by caregivers for older adults in general. Items were adapted to reflect the impact of the stress on adult children who were primary caregivers for their demented parents. (Adult children caregivers were the primary focus of this study.) The structural validity of the CGQ-13 scale was established via EFA with Oblique rotation and a scree plot to determine the optimal number of factors. After item deletions due to low factor loading, the	Cronbach's α estimate for the 13-item scale=0.74.

					scale was reduced to 13 items with high loadings on a single factor explaining 50% of the total variance.	
Toye et al. (2014)[97] Australia	Mixed	Dementia Knowledge Assessment Tool (DKAT2)	Dementia knowledge Two domains: (1) Knowledge of dementia and its progress; (2) Knowledge of dementia support and care (No factors are derived; the Items are organized by the two domains above)	21 items, Binary response options: Yes/No (with a "Don't Know" option)	Content validity was established by four experts with experience in supporting families of people with dementia and prior research in dementia and tool development. The panel examined items for clarity and consistency. After the review, the original pool of 25 items was reduced to 21 final items. The 21-item scale was pilot tested and further refined with 30 family carers (daughters, spouses, and other) and trained staff members (nurses and care workers). No further studies on the structural validity were conducted. Note: Although authors acknowledge the need to conduct validity studies with larger samples. They state that the results provide initial support for the tool's "validity" in that the care workers (who had formal education in dementia) obtained marginally higher scores than family CGs. No further studies on validity are provided.	Family CGs: Cronbach's α, full scale = 0.79 Care workers: Cronbach's α, full scale = 0.79
Kraijo et al. (2014)[98] The Netherlands	ADRD	The Perseverance Time (PT)	Perceived burden with capacity of CG to cope (The tool consists of a single question/item) Note: The single question states: "If the informal care situation stays as it is now, how long will you be able to cope with the care?"	One question, 6 ordered categories: < than one week; > than one week, but < than one month; > than one month, but < than six months; > six months, but < one year; > one year, but < two years; > two years	Content validity was evaluated by performing binary logistic regression analyses between Perseverance Time (PT) scores (dichotomized at three levels: >6 months: Yes/No; >1 year: Yes/No; and >2 years: Yes/No) and characteristics of dementia patients, informal carers, and care situations. Results showed that different categories of PT were associated with different sets of characteristics. Concurrent validity was assessed by estimating Spearman's rank correlations between PT scores and (a) measures of subjective burden (Caregiver Strain Index [CSI], Self-Rated Burden [SRB], and Care-related Quality of Life [CarerQol-7 D]) and (b) happiness (CarerQol-Visual Analogue). The convergent validity of PT was "moderate" with CSI (rho=-0.46, p < 0.001) and care-related quality of life (rho=0.33, p<0.001), good with SRB (rho=-0.63, p < 0.001), but poor with happiness (rho=0.22, p<0.01).	Not reported Note: Richters et al. (2016)[99] reports a study on the test-retest reliability of the Perseverance time instrument.
Sadak et al. (2015)[100] United States	ADRD	Partnering for Better Health-Living with Chronic Illness: Dementia (PBH-LCI: D)	CG activation (CGs' knowledge and skills in health care management of persons with dementia and the ability to meet their own needs.) Six factors: (1) Understanding dementia (2) Recognizing and anticipating symptoms and challenges (3) Managing care patient's medications (4) Managing day-to-day symptoms and challenges (5) Recognizing sudden changes in patients' health (6) Utilizing health services and managing sudden changes in person's self-care	32-items, 5-point Likert scale (ranging from 1=Disagree completely to 4=Agree completely; with an additional response option: 0=Not my responsibility)	Content validity was established through cognitive interviewing with 16 dementia clinical experts using an initial item pool of 86 questions. Experts were asked to reflect on the items they considered important for engaging CGs in patients' health care management and to identify skills that CGs must develop to support optimal health care. Cognitive interviewing was also conducted with 35 primary CGs. As a result of this step, a 35-item scale (23 "knowledge" and 12 "skills" items) emerged. Using the initial 35-item scale, the structural validity was established through a PCA and Varimax rotation explaining 93.8% of the total variance. A scree plot confirmed a 7-component/factor underlying structure. Instead of fitting a multidimensional model, authors conduct a unidimensional Rasch analysis with the initial 35-item pool. Despite the small sample size (N=130) and the underlying multi-dimensional structure found in the previous step, most items showed acceptable fit statistics under the unidimensional Rasch model. After eliminating 3 items due to poor performance in the Rasch analysis, the authors present the final scale as a "six-factor" 32-item scale. Concurrent validity was established through significant Pearson's correlations (p-values < 0.05) between total scores on the PBH-LCI: D and scores on (a) Preparedness for Caregiving(r=0.69), (b) Global Caregiving Self-Competence(r=0.41), (c) Global Caregiving Self-Confidence (r=0.43), and (c) the "mental health component summary" obtained from the SF-12 (r=0.35). Scores on the PBH-LCI: D were negatively correlated with anxiety measured by the General Anxiety Disorder Assessment (r=-0.33). (Sample sizes used in the reported correlations ranged from N=52 to N=130).	Cronbach's α, full scale = 0.95 Pearson's correlation coefficient was used to calculate the test-retest reliability (two-week interval) of the scale scores in a sample of 79 participants (r = 0.76).
Chang et al. (2016)[101] Taiwan	ADRD	Affiliate Stigma Scale	Self-stigma Three factors (components): (1) Cognitive; (2) Affective; (3) Behavioral (Each factor is tested independently to demonstrate unidimensionality of the separate scales.) (Authors also estimate scores for the full Affiliate Stigma Scale.)	22 items, 4-point Likert scale (ranging from 1=Strongly disagree to 4=Strongly agree)	Content validity of the 22-item scale is reported in Mak et al. (2008).[102] The scale was previously tested in a sample of CGs of individuals with mental illness or intellectual disability. The current study validates the scale in dementia CGs. Structural validity. PCA revealed a 3-factor structure of the 22-item Affiliate Stigma Scale. The PCA showed that the first component's eigenvalue for the entire Affiliate Stigma Scale was > 2. Next, the PCA was conducted *separately* for each subset of items defining the 3 domains (cognitive, affective and behavioral) measured by the full scale. Since each separate domain produced eigenvalues <2, the three scales were each considered "unidimensional." Therefore, instead of conducting a CFA using the full 22-item scale, the authors conducted three separate CFAs followed by Rasch models to establish the psychometric properties for each scale: Cognitive, Affective, and Behavioral. All fit indices produced by the CFA indicated satisfactory fit: CFI and TLI were > 0.95, and RMSEA <0.06. Finally, Rasch models confirmed the unidimensionality of the three scales, suggesting their use as separate scales. Most Infit and Outfit statistics obtained through the Rasch model were within the acceptable ranges. Concurrent validity was demonstrated through significant (p-values < 0.05) positive Pearson's correlations between both the total Affiliate Stigma Scale scores (including each domain score and the entire scale score) with criterion measures such as the Caregiver Burden Inventory (r=0.290 to r=0.628), the Taiwanese Depression Questionnaire (r=0.391 to r=0.612), and the Beck Anxiety Inventory (r=0.367	Cronbach's α, full scale =0.93. Note: Using the same sample, authors conduct three separate CFAs for the cluster of items defining the following domains: (1) Cognitive (Cronbach's α=0.855) (2) Affective (Cronbach's α=0.849) (3) Behavioral (Cronbach's α=0.822)

					to r=0.467). Concurrent validity was also shown via expected negative correlations with the World Health Organization Quality of Life questionnaire (r=-0.59 to -0.365).	
Powers & Whitlach (2016)[103] United States	ADRD	Cultural Justifications for Caregiving Scale (CJCS)	Cultural expectations and reasons for providing care (as a function of beliefs and norms about the caregiving role) Two factors: (1) Reciprocity (making a family contribution as a motivation for caregiving); (2) Duty (caregiving as a sense of duty or obligation)	10 items, 4-point Likert scale (1=Strongly disagree, 2=Somewhat disagree, 3=Somewhat agree, 4=Strongly agree)	The theoretical basis for the development of the scale is presented in Dilworth-Anderson[104,105]. Although, the scale has been tested in CGs of individuals with cognitive impairment before, the current study reports the detailed psychometric properties of the scale in a diverse sample of dementia CGs. Structural validity was assessed by PCA to extract the components/factors and Varimax rotation to facilitate the interpretation of item loadings. This analysis was conducted for the full sample and separately for the White and African American subsamples. The PCA analysis in the full sample produced a two-component/factor solution (labeled "Reciprocity" and "Duty") explaining 60% of the total variance. The pattern of loadings, however, differed across the White and African American subsamples suggesting lack of measurement invariance and the need to conduct formal invariance tests to meaningfully compare results between groups. Concurrent validity. Authors correlated CJCS scores on the subscales with CG characteristics and measures of wellbeing and found significant correlations between relationship strain (CG wellbeing) and scores in a) the "Duty subscale" in both African American and White CGs (r=0.28, p<0.01, r=0.32, p<0.01, respectively) and b) the "Reciprocity subscale" in White CGs (r=0.35, p < 0.01). However, more research is needed regarding the measurement invariance of the scale across subgroups.	Cronbach's α for the full 10-item scale in the total sample=0.79. Cronbach's α for the scale (White sample) =0.87. Cronbach's α for the scale (the African American sample) =0.86. Note: No estimates per subscale (2-factors) were provided for the total sample
Piersol et al. (2016)[106] United States	ADRD	Functional Capacity Card Sort (FCCS)	CG appraisal of patient functional capacity (CG estimation/appraisal of patient's "function" regardless of the level of cognitive impairment)	6 "cards", Six Allen Cognitive Levels from lowest Level 1 (automatic actions) to highest Level 6 (planned actions). The six cards describe an individual's ability to perform the daily activity of "washing self." Each card maps to a range of high/low modes within each Allen Cognitive Level, representing a hierarchy of functional capacity.	Content validity was assessed by seven experts (occupational therapists) who reviewed the original set of 12 cards and identified the intended Allen cognitive level of each card. Based on the level of accuracy achieved by raters the cards were collapsed into a final set of six cards and another group of five experts reviewed the cards achieving 100% accuracy. Three independent groups of CGs (N=72) also reviewed the final set of six cards for level of accuracy in terms of cognitive level and mode, level of difficulty, and clarity. Concurrent validity was examined estimating the Spearman's rank correlation between the score on the activities of daily living (ADL) index of the Caregiver Appraisal of Function and Upset (CAFU) scale and the CG ranking of function on the FCCS scale. A moderately positive association between the two variables (rho=0.43, p < 0.001, N=86), provided support for the convergent validity of the FCCS. As hypothesized, the CG FCCS ranking was not significantly associated the NPI scores (rho= −0.14, p =0.16, N=86), providing evidence for discriminant validity of the FCCS.	Interrater reliability: The level of interrater agreement was highest (90.3%) with the lowest level of function, next highest (86.1%) with the highest level of function, and less with the middle levels (74% and 76.4%). Overall agreement measured by Kendall's coefficient of concordance was high (0.83, p=0.0001).
Kiriake & Moriyama (2016)[107] Japan	ADRD	The Partnership Scale (PS)	Ability of family CGs to build partnerships inside and outside of the family while providing care for a family member with dementia. Three factors: (1) Ability for Receptive Coping; (2) Proactive Consultation and Information-Seeking; (3) Trust Formation and Role Coordination	13 items, 5-point Likert scale (ranging from 0=Not at all to 4=Extremely so)	Content validity was established through cognitive interviewing with five family CGs who provided information on the ability of the CG to build collaborative relationships with the patients and with others involved in providing care. Interview results and further literature review were used to create an initial 39-item pool. Next, a team of nine dementia care experts ranked the appropriateness of each item using a 4-point Likert scale from 1 (not appropriate) to 4 (concise and appropriate). The item-content validity index ranged from 78 to 100%. All items were deemed appropriate. Structural validity. To analyze the underlying structure and dimensions of the scale, the sample was randomly split into two groups. The first group (N=130) was used to conduct an EFA using PAF for factor extraction and Varimax rotation, followed by a Horn parallel test to examine the number of factors to retain. After eliminating items with low factor loadings, the scale was reduced from 39 to 14 items. Alternative CFAs with MLE were conducted in the second group (n = 131) for cross-validation purposes. The best fitting model retained 13 items confirming a 3-factor structure. Goodness-of-fit indices for the final CFA model were acceptable (e.g., RMSEA=0.033; CFI=0.977; and TLI =0.971). Concurrent validity. The total score of the PS was confirmed to have a positive Spearman's rank correlation with the Scale of Social Support score (r=0.488, p < 0.01), a negative correlation with the ZBI score (rho=-0.334, p <0.01), and a positive correlation with the Caregiver Positive Appraisal score (rho=0.370, p < 0.01).	Cronbach's α, full scale =0.78 Cronbach's α by subscales: Ability for Receptive Coping (α=0.84) Proactive Consultation and Information-Seeking (α=0.71) Trust Formation and Role Coordination (α=0.67) Test-retest reliability (stability) (one week interval) was assessed with N=50 participants calculating the ICC. ICC for the full scale=0.80. ICC by subscales: Ability for Receptive Coping (ICC=0.83); Proactive Consultation / Information-Seeking (ICC=0.61);Trust Formation and Role Coordination (ICC=0.68).
Maneewat et al. (2016)[108] Thailand	ADRD	Caregiver Resilience Scale (CRS)	Resilience Six factors: (1) Physical competence; (2) Relationship competence; (3) Emotional competence; (4) Moral competence; (5) Cognitive competence; (6) Spiritual competence	30 items, 4-point Likert scale (ranging from 0=Not true to 3=Mostly true)	Content identification began with a literature review of the concept of resilience and interviews with ten CGs of older persons with dementia. Content validity was established by a three-person expert panel review of an initial 36-item pool on relevancy and clarity. Six item were considered redundant and were omitted resulting in a final 30-item scale. The CVI of the final scale was 0.84. The structural validity or underlying factorial structure of the CRS scale was established via PCA with a Varimax rotation to maximize the variance of squared factor loadings and increase factor structure interpretability. The PCA produced a 6-component/factor solution explaining 63.67% of the variance of the items in the scale.	Cronbach's α, full scale = 0.87. Cronbach's α by subscales: ranged from 0.52 to 0.87.

Sullivan et al. (2016)[109] Australia	ADRD	The Thoughts Questionnaire (TQ)	Dysfunctional thoughts Seven "themes" represented in the measure: (1) Perfectionism; (2) Overinvestment and embarrassment; (3) Personal vulnerability and fatality; (4) Interpretation of behavior; (5) Self-neglect; (6) Sole responsibility; (7) Perceived social support	25 items, 5-point Likert scale (0=Totally disagree, 1=Disagree, 2=Neither agree nor disagree, 3=Agree, 4= Totally agree)	Content validity was determined by an expert panel of project team members and both professional and nonprofessional family CGs who reviewed and evaluated an initial 55-item bank for face validity, usability, theoretical coverage, and overall perceived utility. A final 25-item scale was also assessed for item readability level using the Flesch Kincaid grade level score. Concurrent validity was established with Pearson's correlations between the TQ scale and: The Dysfunctional Thoughts about Caregiving Questionnaire (DTCQ); the geriatric depression (GDS); and Perling's Stress and Coping (PSC) scales. TQ scores were not significantly associated with GDS (r=0.319, p=0.183) or DTCQ scores (r= 0.29, p=0.10). However, as expected, TQ was significantly associated with all stress risk factors from Pearling's scales except for "conflict over attitudes toward the person with dementia." (Pearson's correlation estimates ranged from r=0.359 to r=0.620, $p < 0.05$). The expectation that the TQ would be negatively associated with a measure of coping was not supported.	Cronbach's α, full scale =0.85
Sadak et al. (2017)[110] United States	ADRD	Kingston Caregiver Stress Scale (KCSS)	CG stress Three factors: (1) Personal-/Caregiving-related stress; (2) Family-related stress; (3) Financial stress	10 items, 5-point Likert scale (ranging from 1=no stress to 5=extreme stress)	Content/face validity was addressed briefly by the authors in the website description of the scale[111] indicating that content validity the KCSS was established by examining the scale questions and determining that they addressed the characteristics of caregiver stress. Structural validity was established using a PCA that yielded a three-component/factor solution explaining 71% of the total variance. The three components/factors mapped on to a priory identified "domains" labeled as: Personal/caregiving-related stress, Family-related stress, and Financial issues. Concurrent validity. Scores from subsamples completing the General Anxiety Disorder (N=51) scale and Patient Health Questionnaire (N=52) were significantly (*p-value*s < 0.001) and moderately correlated with KCSS scores (r=0.69, 0.57, respectively).	Cronbach's α, full scale =0.88. Cronbach's α by subscales: Caregiving (α=0.885); Family (α=0.871) Financial (1 item, n/a) Test-retest reliability (two-week interval) in a subsample (N=78): Pearson's r=0.88.
Piggott et al. (2017)[112] United States	ADRD	Caregiver Confidence in Sign/Symptom Management (CCSM) Scale	CG self-efficacy (confidence) in sign/symptom management; CG role strain Four factors: (1) Knowledge of signs/symptoms; (2) Management of cognitive signs/symptoms; (3) Management of medical signs/symptoms; (4) General medication management	25-items, 5-point Likert scale (ranging from 1=Not at all true/confident to 5=Extremely true/confident)	Content validity. Five CGs participated in cognitive testing assessing item difficulty and relevance of an initial 37-item bank. They were also asked to provide recommendations of additional questions concerning their relative's medical problems or about their own self-efficacy not measured in the current scale. Further revisions reduced the original scale to 26 items. The structural validity of the 26-item scale was established through EFA with a Promax (oblique) factor rotation followed by the examination of the scree plot to determine scale dimensionality. After eliminating an item, the final 25-item CCSM scale produced a four-factor solution. Concurrent validity of the CCSM scale was assessed by Pearson's correlations with 3 widely used CG measures: (1) the ZBI-role strain (r=-0.36, p <0.001) and the ZBI-personal strain (r=-0.14, p=0.06); (2) the Generalized Anxiety Disorder-7-item scale (r=-0.12, p=0.09); and (3) the 9-item Patient Health Questionnaire-Depression (r=-0.06, p=0.43). CGs with less role strain reported more confidence in all subscales (correlations ranged from 0.37 ($p < 0.001$) for general medical management to 0.15 (p=0.042) for knowledge about signs/symptoms. The association between the caregiver's self-report of medical training and CCSM scores was also significant (r=0.26, $p < 0.001$).	Cronbach's α, full scale =0.92. Cronbach's α by subscales: Knowledge of signs/symptoms (α=0.83); Management of cognitive signs/ symptoms (α= 0.85); Management of medical signs/symptoms (α=0.87); General medication management/ responsiveness (α= 0.85) Test–retest reliability (2-day interval) was assessed with N=17 CGs using Pearson's and ICC coefficients. Test-retest reliability for the total scale (r=0.92, ICC=0.91). Test-retest reliability by subscale: Knowledge of signs/symptoms (r=0.57, ICC=0.56); Management of cognitive signs/ symptoms (r=0.87, ICC=0.82); Management of medical signs/ symptoms (r=0.78, ICC=0.78); General medication management (r=0.95, ICC=0.94)
Romero-Moreno et al. (2017)[113] Spain	ADRD	Valued Living Questionnaire Adapted to Caregiving (VLQAC)	Personal values in the CG stress process Two factors: (1) Commitment to own values; (2) Commitment to family values	12 items, 10-point Likert scale (ranging from 1=Not at all important to 10=Extremely important)	Content validity was established in the original version of the scale developed by Wilson et al., 2010.[114] Authors added two caregiving-related items and validated the expanded scale in a sample of ADRD CGs. Structural validity was evaluated through EFA applying Oblimin rotation and followed by a Horn's parallel analysis to determine the optimal number of underlying factors. EFA identified two factors explaining 43.42% of variance between scale items. Concurrent validity. Pearson's correlation coefficients were used to study associations between scale factors (subscales) and criterion measures. Higher scores in "Commitment to Own Values" and "Commitment to Family Values" factors were significantly associated with lower scores in depression (measured by CES-D) (r=-0.31, $p < 0.01$; r=-0.18, $p < 0.01$, respectively) and anxiety, measured by POMS (r=-0.27, $p < 0.01$; r=-0.31, $p < 0.01$, respectively), as well as with a higher score in the Satisfaction with life scale (r=0.35, $p < 0.01$; r=0.40, $p < 0.01$, respectively). In addition, higher scores in the "Commitment to Own Values" factor were associated with higher scores in emotional acceptance, measured by the "Difficulties in Emotion Regulation Scale" (r=0.14, $p < 0.05$).	Cronbach's α, full scale =0.75. Cronbach's α by subscales: Commitment to Own Values (α=0.71) Commitment to Family Values (α=0.61)
Stott et al. (2017)[115]	ADRD	Hospital Anxiety and Depression	Anxiety and depression Three factors:	13 items, 4-point Likert scale with	Content validity. Previously established by Zigmond & Snaith (1983).[116] The current study validates HADS in a sample of AD CGs.	Cronbach's α estimates by subscales (factors):

United Kingdom		Scale (HADS)	(1) Anxiety; (2) Depression; (3) Negative affectivity	several labels per scale: (1) 0=not at all to 3=most of the time (2) 0=hardly at all to 3=as much as I ever did (3) 0=very seldom to 3=often	Structural validity. CFA with robust MLE was used to test the fit of three previously proposed factor structures (one-, two-, and three-factors) using HADS data from the sample of dementia CGs. After eliminating one item and re-fitting the model, a 3-factor structure produced acceptable goodness-of-fit indexes (e.g., RMSEA=0.06; GFI=0.96; and TLI =0.95). Cross-validation in an independent sample confirmed initial results. Concurrent validity was examined using bivariate correlations between the Positive and Negative Affect Schedule (PANAS) and HADS subscales. Correlations were large, significant (*p-value*s < 0.001), and in the expected direction ranging from -0.65 to -0.37 between scores on all HADS scales and those on PANAS-PA and from 0.57 to 0.69 for those in PANAS-NA. Measurement invariance tests across subgroups revealed possible systematic response bias between older (≥65) and younger (<65) adults that may render latent variable mean group comparisons uninterpretable due to measurement bias rather than true group differences.	Anxiety (α=0.87) Depression (α=0.85) Negative affectivity (α=0.77)
Losada et al. (2017)[117] Spain	ADRD	The Caregiving Ambivalence Scale (CAS)	Ambivalence attitudes or feelings (The scale measures the degree in which CGs' attitudes and feelings toward their relatives afflicted with dementia are mixed or conflicted.) One factor: (1) Ambivalence	6 items, 4-point Likert scale (0=Never, 1=Sometimes, 2=Frequently, 3=Always)	Although content validity is not formally addressed in the study, authors conduct a literature review and present research linking the caregiving experience to heightened ambivalence and conflicting emotions as a rationale for developing a caregiving ambivalence measure. Drawing upon a previous scale[118] and clinical experience, authors developed 6 items measuring ambivalent feelings in dementia CGs associated with caregiving. Structural validity. To analyze the underlying structure of the scale, the sample was randomly split into two groups. The first group (N=200) was used to conduct an EFA using MLE for factor extraction, followed by a Horn's parallel analysis to determine dimensionality. A CFA was conducted in the second group (N = 201) confirming a unidimensional scale structure. Goodness-of-fit indices for the CFA model were acceptable (e.g., RMSEA=0.058; GFI=0.91; and TLI =0.987). Concurrent validity was demonstrated by high Pearson's correlations between CAS scores and measures of disruptive behavior using the RMBPC (r=0.42, p<0.01); depression using the CES-D (r=0.32, p <0.01), and anxiety using POMS- tension subscale (r=0.46, p<0.01).	Cronbach's α, full scale =0.86.
Abdollahpour et al. (2017)[119] Iran	ADRD	Positive Aspects of Caregiving (PAC) Questionnaire	Gains in positive aspects of caregiving Two factors: (1) Patient and CG relationship; (2) CG's psychological wellbeing	10 items, 5-point Likert scale (ranging from 0=Strongly disagree to 4=Strongly agree)	Content validity was assessed using a panel of five content experts (four neurologists and one psychologist), five CGs as lay experts, as well as one methodologist for the content validation process. Items were evaluated for relevancy and clarity using "item and scale content validity indexes" (I-CVI and S-CVI, respectively) resulting in acceptable ranges. I-CVI for relevancy and clarity were 0.90 to 1 and 0.80 to 1, respectively. S-CVI for relevancy and clarity indices were 0.97 and 0.93, respectively. The structural validity was evaluated via an EFA with Varimax rotation identifying a two-factor structure that explained 47% of total variance in PAC. Concurrent validity-The Pearson's correlation of *Self-reported health* (SRH) and *PAC* scores was examined for establishing "concurrent" validity (r=0.343, p=0.01). Divergent validity was assessed by correlating *PAC scores* with a measure of CG burden (*The Iranian caregiver questionnaire*) (r= -0.291, p=0.001). Rather than showing lack of association between the two measures, authors contrasted the two measures.	Cronbach's α, full scale =0.79. Cronbach's α by subscales: Patient and CG relationship (α=0.711); CG's psychological wellbeing (α=0.707) Test–retest reliability (3-week interval) was evaluated with 20 randomly selected CGs calculating the ICC. The ICC for the full scale=0.95. ICC by subscales: Patient and CG relationship (ICC=0.80) and Caregiver's psychological wellbeing (ICC=0.87)
Fabà & Villar (2017)[120] Spain	ADRD	Gains Associated with Caregiving (GAC) scale	Gains associated with caregiving for a person with dementia One factor: Gains	22 items, 4-point Likert scale (0=Not at all; 1=Yes, slightly; 2=Yes, quite a lot; 3=Yes, very much so)	Content validity was established by three external expert judges in the field of psychogerontology and developmental psychology. The judges evaluated the semantic definition of the five key domains (Industry, Identity, Intimacy, Generativity, and Ego Integrity) identified by the authors from the literature and included in an initial 62-item GAC scale. Two of the three judges were also asked to indicate the domain to which they considered each item belonged. Judges' agreement was high (Cohen's kappa coefficients ranged from 0.77 to 0.90, p < 0.001) Structural validity was established by iterative EFA starting with a reduced 32-item scale using an independent sample of 152 participants. After eliminating items with low loadings and item-rest score correlations, the final EFA model produced a unidimensional (one-factor) 22-item scale. A scree plot confirmed the solution. Using the same initial protocol, an independent sample of 260 participants was selected to conduct a CFA on the resulting 22-items confirming a unidimensional GAC scale. With the exception of the SRMR=0.07, goodness-of-fit statistics, however, were below recommended thresholds (e.g., CFI=0.71). Concurrent validity was assessed by calculating Pearson's correlations between GAC scores and the ZBI (r=-0.229, p < 0.01), the Geriatric Depression Scale–Short Form (r=-0.237, p < 0.01), and the Satisfaction With Life Scale, SWLS (r = 0.257, p < 0.001).	Cronbach's α, full scale =0.95
Weisman de Mamani et al. (2018)[121]	ADRD	Stigma Impact Scale (SIS)	Stigma Four domains: (1) Social Rejection; (2) Financial Insecurity; (3)	24-items, 4-point Likert scale (ranging from 1=Strongly disagree to 4=Strongly	The structural validity of the scale is not established as part of the current study with dementia CGs. Authors relied on the 4-domains of SIS defined by Burgener & Berger (2008)[122] using an adapted version of the original scale in a different population of CGs. Content validity was also examined in the adapted version. Although the objective of the current study was not to establish the validity of the SIS scale in a	Cronbach's α, full scale =0.93.

United States			Internalized Shame; (4) Social Isolation	agree)	sample of dementia CGs, the study provides evidence of the concurrent validity and reliability of SIS among dementia CGs. Authors hypothesize an association between SIS measures and constructs measured by Expressed emotion assessed using the 20-item Family Questionnaire (FQ). FQ also has two subscales: Emotional Over involvement (EOI) and Criticism. As hypothesized, greater CG stigma was positively associated with Criticism (r=0.372, p < 0.001) and EOI (r= 0.398, p < 0.001). EE total scores (i.e., the sum of the Criticism and EOI subscales) were also significantly correlated with stigma (SIS) scores (r= 0.434, p < 0.01).	
Moholt et al. (2018)[123] Norway	ADRD	Carers of Older People in Europe (COPE) Index (Scale validation with family carers of people with dementia-Norway)	Support needs Three factors: (1) Negative impact of caregiving; (2) Quality of support; (3) Positive values of caregiving	15 items, 4-point Likert scale (ranging from 1=Never to 4=Always)	Content validity was established in the original version of the scale developed by Mckee et al., 2003.[124] The original version targeted informal caregivers of older adults in general. The current study validates COPE in a sample of dementia caregivers. Structural validity. To analyze the underlying structure and dimensions of the scale items, the sample was randomly split into two groups. The first group (N=215) was used to conduct an EFA using PAF method to extract factors followed by an examination of a scree plot of eigenvalues to examine the number of factors to retain. A CFA with robust MLE was conducted in the second group (N=215) for cross-validation purposes confirming a 3-factor structure. Goodness-of-fit indices for the CFA model were acceptable (e.g., RMSEA=0.050; CFI=0.951; and TLI =0.939). (A second order model also provided a good fit supporting the use of a global COPE Index score.) Concurrent validity. The Pearson's correlation between COPE-Index and the World Health Organization-5 Well-being Index (WHO-5) was=0.62, p < 0.001; the correlation of COPE-I and demands of caregiving item was=0.49, p < 0.001. As expected, negative and statistically significant correlations were obtained between Cope-Index scores and a) a general status item (r=-0.37, p < 0.001) and b) scores on a social restriction scale (r=-0.33, p< 0.001).	Cronbach's α estimates per subscale: Negative impact of caregiving (α=0.86) Quality of support (α=0.76) Positive values of caregiving (α=0.64) Test-retest reliability (4-week interval) was examined using Spearman's rank order correlation with a small subsample (N=32). Negative impact of caregiving (r=0.91) Quality of support (r=0.76) Positive values of caregiving (r=0.92)
Oliveira & Aubeeluck (2018)[125] United Kingdom	ADRD	Dementia Quality of Life Scale for Older Family Carers (DQoL-OC)	Quality-of-life of older family carers One factor: Quality of life	22 items, 5-point Likert scale (1=Always, 2=Frequently, 3=Occasionally, 4=Rarely, 5=Never)	Content validity and "practicality" were determined by a panel of six experts (four researchers and two older family carers) who assessed the relevance, length, clarity of language, and levels of difficulty of an initial item bank of 89 items that was further reduced to 81 items. The structural validity of the 81-item scale was determined by iterative EFAs using PAF for factor extraction and Promax rotations to account for factor correlations. Each iteration was followed up by a re-evaluation of parallel analyses and scree plots. The final EFA iteration produced a satisfactory 22-item unidimensional scale explaining 43.83% of the total variance. Concurrent validity. The total scores of the DQoL-OC showed significant Pearson correlations (*p-values* < 0.001) with (1) the World Health Organization Quality of Life Scale (r=0.74), (2) the Satisfaction with Life Scale (r=0.65), (3) the Perceived Health Status Visual Analogue Scale (r=0.39), and (4) the Overall Perceived Health-Related Quality of Life Visual Analogue Scale (r=0.44).	Cronbach's α, full scale = 0.936. Test–retest reliability (two-week interval) was established through the calculation of the ICC using a small subsample of 18 participants. (ICC=0.835; p<0.001).
Peipert et al. (2018)[126] United States	ADRD	Dementia Burden Scale – Caregiver (DBS-CG)	CG experience, CG burden Three factors: (1) Strain of caregiving; (2) Distress caused to the CG by the patient's behavioral symptoms; (3) Depressive symptoms	34 items, Likert type scales varying from: "On a regular basis," "Sometimes," "No"; or "Not distressing at all" to "Extreme or very severe"; or "Not at all" to "Nearly every day"	The DBS-CG scales was developed by combining 34 items from existing scales. The structural validity for the 34-item scale was established through two alternative CFA models: a 3-factor model and a bifactor model (one general factor and 3-specific factors) using items from three existing scales: The Modified Caregiver Strain Index (MCSI), the NPI Questionnaire-Distress scale, and the Patient Health Questionnaire (PHQ-9). The resulting models fit the data well but the bifactor model produced a slightly better fit: (RMSEA=0.05, CFI 0.95). The score in the general factor represented "caregiver burden." Responsiveness-Minimal important differences estimates of the amount of clinically relevant change on the scale ranged from 4 to 5 points (effect sizes associated with each of these differences were "small": 0.20–0.49).	McDonald's ω for the full scale=0.93.
Stansfeld et al. (2019)[127] United Kingdom	ADRD	Sense of Coherence Scale-13 (SOC-13)	Sense of coherence Three factors: (1) Meaningfulness; (2) Comprehensibility; (3) Manageability	13 items, 7-point Likert-type scales with labels that vary per cluster of items.	The content validity of the scale was established by Antonovki (1993).[128] The scale has been used but its psychometric properties have not been established. This study, evaluates the measurement properties of the scale in a sample of dementia CGs. The structural validity of the scale was assessed with a CFA. However, the solution did not confirm the originally proposed 3-factor structure. The proposed model did not produce an adequate fit; with indices falling below or above acceptable thresholds. Factor loadings, however, were significant and ranged from 0.419–2.124. Concurrent validity. SOC-13 scores were a) strongly and positively correlated with scores on the Resilience Scale-14 (r=0.56, p < 0.001), b) moderately and positively correlated with scores on the 7-item Sense of Competence Scale (r=0.42, p < 0.001), and d) scores of the Self-efficacy for managing dementia scale (r=0.46, p < 0.001). SOC-13 was also moderately and negatively correlated with *health-related quality of life,* measured by the EuroQol 5-Dimension 5-level questionnaire (r= −0.38, p < 0.001).	Cronbach's α, full scale =0.88. Cronbach's α by subscales: Meaningfulness (α=0.72) Comprehensibility (α=0.76) Manageability (α=0.705)
Davis et al. (2019)[129]	ADRD	Guilt After Placement	Guilt and ambivalence following nursing home	10-items, 5-point Likert scale	Although no formal statements on content validity are made, authors developed scale items through information obtained from focus groups and a literature review of the emotional aspects of placement.	Cronbach's α, full scale =0.92.

United States		Questionnaire (GAP-Q)	placement One factor: Decisional guilt-reflecting guilt associated with making the decision to place	(ranging from 0=Never to 4=Always)	The focus groups consisted of a study clinician attending caregiver support groups run by the Alzheimer's disease association to explore CGs' feelings regarding nursing home placement. An initial sample of 46 items was generated using this method. Using an initial 46-item pool, the scale's structural validity was assessed via EFA with PAF to extract factors and Varimax rotation to explore factor loadings. After several EFA iterations and refinements, a 10-item GAP-Q scale produced a single underlying factor (1-factor solution) as the best fitting model. Concurrent validity was evaluated in a subset of the sample (N=53) using Pearson correlations between the GAP-Q scores and concurrent measures of (a) depression using the CES-D (r=0.53, p <0.001), (b) CG burden using the ZBI (r=0.48, p <0.001), (c) conflict with staff using the Interpersonal conflict scale (ICS) (r=0.47, p <0.001), and (d) "wellbeing" using the short form health survey (SF-36) (r= -0.30, p <0.05).	
Ying et al. (2019)[130] Singapore	ADRD	Center for Epidemiological Studies Depression Scale (CES-D)	Depression in CGs of persons with dementia Four factors: (1) Depressed affect; (2) Somatic symptoms; (3) Positive affect; (4) Interpersonal problems	20 items, 4-point Likert scale (0=Rarely or none of the time, 1=Some or little of the time, 2=Moderately or much of the time, 3=Most or almost all the time)	The scale's structural validity was assessed with alternative CFA models varying in dimensionality (from a one-factor to the original 4-factor model). The 4-factor model produce the best fit (e.g., RMSEA=0.077; CFI=0.909; and TLI=0.895). (TLI was marginal; values above 0.90 are recommended.) Concurrent validity was evaluated by examining the correlations among the CES-D, the Gain in Alzheimer care instrument (GAIN), the ZBI, and their respective subscales, using the Pearson's correlation coefficient (all p-values < 0.01). CES-D correlated strongly with ZBI scores (r = 0.71) and most of the subscales of ZBI (r=0.60 to 0.70). Correlations were *weaker* between total CES-D and the Finances subscale of ZBI (r = 0.46) or the Caregiving gains scale (GAIN) (r= –0.16). The Positive affect subscale of CES-D was negatively associated with ZBI subscales (r= -0.18 to -0.34).	Cronbach's α, full scale =0.92. Cronbach's α by subscales: Depressed affect (α=0.91) Somatic symptoms (α=0.85) Positive affect (α=0.74) Interpersonal problems (α=0.69)
Barello et al. (2019)[131] Italy	Mixed	Caregiving Health Engagement Scale (CHE-s)	CG engagement in healthcare One-factor: Engagement in healthcare	7 items, 4 types of "ordered" narrative/storylines in the process of family CG engagement: 1=denial, 2=hyper-activation, 3=drowning and 4=balance	Content validity. An initial item pool was generated based on literature reviews and interviews with a sample of 22 CGs about feelings and experiences in their caring roles and feeling of engagement. The item pool was reviewed for content and face validity by the project steering committee, and by CGs who participated in the interview, to check the relevance and comprehensiveness of items, response options, and instructions. This resulted in a refined 7-item scale. The structural validity of the CHE scale was evaluated using different approaches. Given the ordinal nature of the 7-item scale, the authors first conducted a CATPCA that yielded a one-dimensional (one-component) structure explaining 67.0% of the total variability. A CFA also yielded a one-factor structure producing adequate GFI's (e.g., CFI=0.96, SRMR=0.03, and RMSEA=0.05). Finally, a Rasch analysis, confirmed the unidimensionality of the scale. All Infit and Outfit statistics were within the acceptable range (0.66 to 1.27). Concurrent validity was established through Pearson's correlations between CHE's factor scores and scores from (a) the Caregiver Burden Inventory; (Pearson's r coefficients ranged from -0.62- -0.40, all p-values < 0.001) and (b) the two subscales of Caregiving Self-Efficacy (SE); SE-Obtaining respite (r=0.25, p <0.001) and SE-Responding to Disruptive Patient Behaviors (r=0.48, p < 0.001).	Ordinal Cronbach's α =0.88 (Using a polychoric correlations matrix) PSI (reliability) produced by the Rasch analysis=0.907
Brown et al. (2019)[132] United Kingdom	ADRD	Carer Dementia Quality-of-Life (C-DEMQOL)	Quality-of-Life applicable across the range of caring situations and severity in dementia. Five domains: (1) Carer-patient relationship; (2) Carer wellbeing; (3) Meeting personal needs; (4) Confidence in the future; (5) Feeling supported	30 items, 5-point Likert scale (ranging from 5=Best to 1=Worst)	Content validity. Qualitative interviews with 32 family carers and 9 support staff, and two focus groups with 6 carers and 5 staff were conducted to generate measurable domains and indicators (items) of dementia carer quality of life. Pilot testing further refined the questionnaire items. The scale's structural validity was assessed by EFA with ordinal variables using a polychoric correlation and oblique rotation. A Horn's parallel analysis confirmed a 5-factor structure underlying the original 40-item pool. Given the high correlation of factors, an exploratory bifactor model was also tested. An independent sample was used to fit a graded response model with an underlying bifactor model to establish the final underlying structure of the scale and its psychometric properties. This resulted in a final 30-item scale with a bifactor structure (one general and five orthogonal specific factors). The fit of the model was within acceptable ranges (e.g., RMSEA=0.066; CFI=0.968; and SRMR = 0.072). Concurrent validity was established via positive, significant (*p-values* < 0.001) correlations between C-DEMQOL total (overall) scores and similar constructs: e.g., short form health survey, SF 12-mental (r=0.70); Personal Wellbeing Scale (r=0.63); World Health Organization (WHO) QOL: physical health (r=0.61) and psychological (r=0.63). Divergent validity was determined via "lower" (although not necessarily insignificant) correlations with conceptually unrelated constructs (e.g., correlations between C-DEMQOL scores and SF-12 physical (r=0.34, p < 0.001)). The average convergent correlation between C-DEMQOL and carer-focused external scales was 0.58, and the average divergent correlation with unrelated constructs was 0.40.	McDonald's ω, full scale=0.97. McDonald's ω estimates by subscales: Meeting personal needs (ω = 0.95) Carer wellbeing (ω = 0.91) Carer-patient relationship (ω = 0.82) Confidence in the future (ω = 0.90) Feeling supported (ω = 0.85)
Cheng et al. (2019)[133] China	ADRD	Caregiver Grief Questionnaire (CGQ)	Pre-death grief Two factors: (1) Relational deprivation (RD); (2) Emotional pain (EP)	11 items, 5-point scale (ranging from 1=Strongly disagree to 5=Strongly agree)	The current scale was assembled from existing measures of CG grief: 15 items from the Meuser-Marwit CG Grief Inventory[34] and 3 items from Pearlin's et al.[134] measure of "relational deprivation." After a content inspection by the team, 7-items were eliminated and the 11-item scale was validated in a sample of AD CGs. Structural validity. A hypothesized two-factor model was evaluated against the one-factor model using a	Cronbach's α, full scale = 0.90 Test-retest reliability (two-week interval) was evaluated with Pearson correlation in a sample N=46, r = 0.95.

					CFA. A two-factor model (RD & EP) provided a modest fit to the data (e.g., RMSEA=0.14; CFI=0.94; and non-normed fit index, NNFI=0.92). Concurrent validity was shown by significant (*p*-values < 0.001) positive Pearson's correlations of CGQ scores with ZBI (r=0.47), HAM-D Scale (0.31), and the Neuropsychiatric Symptoms scale (0.26). Discriminant validity. As expected, neither total CGQ scores nor RD or EP subscales were associated with "social network size."	
McCaffrey et al. (2020)[135] Australia	Mixed	Carer Experience Scale (CES)	Caregiving experience Six domains: (1) Activities outside caring; (2) Social support (family and friends); (3) Institutional support (public and private organizations); (4) Fulfillment from caring; (5) Control over the caring; (6) Relationship with patient	6 items, 3-point Likert scale by "amount" (1=A little/few, 2=Some, 3=A lot/most) or by "frequency" (1=Rarely, 2=Sometimes, 3=Mostly)	Content validity. The resulting CES scale was developed in a previous study (Al-Janabi et al, 2008)[136] using a meta-ethnography of existing qualitative data to determine key conceptual attributes of caring. Sixteen semi-structured interviews with carers of older people were conducted to refine attributes and develop them into the CES measure. In this study, concurrent validity was established through Spearman rank correlations between CES scores and (a) the Adult Social Care Outcomes Toolkit for Carers (rho=0.71, p<0.001) and (b) the Care-Related Quality of Life (rho=0.45, p<0.001). Group discriminant validity was established by a Kruskal-Wallis one-way analysis of variance. Higher carer-related scores were associated with lower hours of care provided per week for CES (Kruskal–Wallis 53.41, p < 0.001). There was a significant difference in mean CES scores between informal carers who provided <20 hours and ≥40 hours (p < 0.001), 20-29 hours and ≥40 hours (p <0 .001) and 30-39 hours and ≥40 hours (p< 0.05).	Cronbach's α, full scale =0.59. Test-retest reliability was estimated via the ICC=0.81. The follow-up survey was administered 2 weeks after the baseline survey to a sample N=104.
Wynn & Carpenter (2020)[137] United States	Mixed	The Frontotemporal Dementia Knowledge Scale (FTDKS)	Frontotemporal dementia knowledge One factor (Knowledge of FTD) and 4 content areas: (1) Risk factors (2) Symptoms (3) Disease course (4) Caregiving	18 items, 4-point Likert-type scale format (False, Probably false, Probably true, True) with an auxiliary "Don't Know" option	No formal statements on content validity are presented. However, authors reported reviewing the literature to ensure item content relevance and coverage. The research team also reviewed an initial 24-item pool, removed items with overlapping content, and rewrote items for clarity achieving a twelfth-grade reading level. No factor analysis to examine the underlying factor structure of the scale is reported, but authors state that the scale "measured a unidimensional construct of knowledge about FTD". In the CG sample, convergent validity was demonstrated by correlating FTDKS and level of care provided to people with FTD (Pearson's r=0.231, p < 0.05). In contrast to expectation, scores on the FTDKS were not correlated with the number of people with FTD known (r=0.179, ns).	*CG Sample:* Cronbach's α, full scale =0.846. Split-half reliability (Spearman-Brown) = 0.814. *Professional Care workers:* Cronbach's α, full scale = 0.704. Split-half reliability (Spearman-Brown) = 0.728.
Van Houtven et al. (2020)[138] United States	ADRD	Caregiver Perceptions About Communication with Clinical Team Members (CAPACITY) Instrument	Perception of support. CGs perceptions of support from the patient's health care team and their communication experiences with the team. Two factors: (1) Capacity/preferences; (2) Communication	12 items, 4-point Likert scale (1=Rarely, 2=Sometimes, 3=Most of the time, 4=Always)	Content validity. Authors reported item generation being informed by a previous measure (the Patient Perceptions of Integrated Care), literature on patient perceived satisfaction and quality of care with health care encounters (e.g., Consumer Assessment of Healthcare Providers and Systems [CAHPS]), and an organizing framework of CG skills. No further pilot testing steps are provided. The structural validity of the CAPACITY scale was established by CFA. A model with a two-factor structure (with factors labeled as "Capacity/preferences" and "Communication") was the best fitting model. Goodness-of-fit indices were acceptable (e.g., RMSEA=0.085; CFI=0.973; and TLI=0.967).	McDonald's ω by subscales: Communication (ω=0.90) Capacity (ω=0.94)
Doherty et al. (2020)[139] Australia	ADRD	Consumer Access, Appraisal and Application of Services and Information for Dementia (CAAASI-Dem)	Dementia-specific health literacy (ability to locate, navigate and use dementia services and information-either for oneself or in providing care for others) Five factors: (1) Evaluation and engagement; (2) Readiness (3) Social Supports; (4) Specific Dementia Services; (5) Practical Aspects	26 items, Mixed format items: 5-point Likert scale: (ranging from "Not at all confident" to "Extremely confident" or "Strongly agree" to "Strongly disagree" or a binary scale: Yes/No)	The content validity of an initial pool of 70 items was assessed by three experts in the field with extensive experience and expertise in both scale development & dementia subject matter. The reviewers also checked content relevance, comprehensiveness, comprehensibility, and technical quality. The item pool was reduced to 65 items as a result of the content validity assessment. Using an independent sample of 1412 participants, items were pilot tested and data was used to make further revisions based on item-total and inter-item correlations and Cronbach's α--if-item-deleted. This revision resulted in the removal of 34 items. The underlying factorial structure of the scale was studied with an initial 31-item pool. The structural validity of the reduced 31-item scale was established by EFA with a PAF extraction method using response data from an independent sample of 3146 participants. After eliminating items with low loadings and item-rest score correlations, and re-running the EFA with an Oblimin rotation, the final EFA model produced a five-dimensional 26-item scale that explained 69.7% of the total variance.	Cronbach's α estimates by subscales: Evaluation and engagement (α=0.953) Readiness (α=0.911) Social Supports (α=0.887) Specific Dementia Services (α=0.926) Practical Aspects (α=0.888)
Furukawa & Greiner (2020)[140] Japan	ADRD	Social Capital Scale for Caregivers of People with Dementia	Social capital: social networks, reciprocity, and trust Three factors: (1) Support for people with dementia and their CGs; (2) Trust in providing dementia care; (3) Support from neighbors	17 items, 5-point Likert scale (ranging from 1=Strongly disagree to 5=Strongly agree)	Content validity was established by five experts using a scale to rate the relevance of 41 items. The ratings returned a content validity index, CVI= 0.94. Based on the CVI, 35 of original 41 items were retained. After further refinement, a final pool of 27 items were submitted to factor analysis. The structural validity of the scale was established by EFA using a ML likelihood factor extraction and oblique rotation. EFA produced a 3-factor solution and a final set of 17 scale items explaining 46.5 % of the total variance. Concurrent validity was demonstrated by a positive and significant Pearson correlation between the total scale scores and the Positive Aspects of Caring (PAC) scale (r=0.62, p<0.01). Each factor on the scale was also significantly correlated with the PAC scale (*Factor 1*: r = 0.42; *Factor 2*: r = 0.58, and *Factor 3*: r = 0.40).	Cronbach's α, full scale =0.85. Cronbach's αs by subscales: Support for people with dementia and their CGs (α=0.86); Trust in providing dementia care (α=0.74); Support from neighbors (α=0.78) Test-retest reliability (4-week interval) was estimated with the ICC in a sample of 50 respondents. (ICC=0.71)

Sakanashi & Fujita (2020)[141] Japan	ADRD	The Empowerment Scale for Family Caregivers of Community-dwelling People with Dementia (EFCD)	Empowerment Four factors: (1) Excellent Practice in Dementia Care (2) Understanding the Essence of Dementia Care (3) Caring for Oneself as well as for the Person with Dementia (4) Having Peers with Shared Support Activities	16-items, 4-point Likert scale (0=Disagree, 1=Somewhat disagree, 2=Somewhat agree, 3=Strongly agree)	Face/content validity was examined by asking five administrators from the Alzheimer's Association of Japan to evaluate an initial pool of 44 items for appropriateness. This review and further item analyses resulted in the reduction of the scale to 31 items. Structural validity was assessed by EFA using PAF as factor extraction and Promax rotation to account for factor correlations. Sixteen items remained after deleting item factor loadings less than 0.40. A scree plot indicated a 4-factor solution. A CFA supported a 4-factor structure. Goodness-of-fit indices for the CFA model were, overall, satisfactory (e.g., RMSEA=0.08; CFI=0.91; and GFI=0.898). (The GFI was "marginally acceptable".) Concurrent validity was established estimating Spearman's rank correlations between EFCD and known measures of self-efficacy and general health. For example, statistically significant correlations were obtained between the EFCD and the Japanese versions of (a) the revised scale for caregiving self-efficacy, RSCSE (rho=0.52, p<0.01) and (b) the general health questionnaire-12, GHQ12 (rho=-0.27, p<0.01).	Cronbach's α, full scale =0.90. Cronbach's α by subscales: Excellent Practice in Dementia Care (α=0.86); Caring for Oneself as well as for the Person with Dementia (α=0.72); Having Peers with Shared Support Activities (α=0.70). Test-retest reliability (7-28 days interval) for the full scale was estimated with the ICC in a sample of 101 respondents. (The ICC=0.51; "moderate" test-retest reliability).
Galvin et al. (2020)[142] United States	ADRD	The Positive and Negative Appraisals of Caregiving (PANAC) Scale	Positive and negative experiences associated with caregiving Two factors: (1) Positive Appraisals (PAs) (2) Negative Appraisals (NAs)	16 items 5-point Likert type scale (ranging from 1=Strongly disagree to 5=Strongly agree	After a systematic review of the literature on constructs covering positive and negative aspects of caregivers, the authors developed PANAC and study its psychometric properties among AD CGs. The structural validity of the scale was determined by EFA with principal components as the factor extraction method and Varimax rotation that produced a 2-factor/component solution explaining 46.7% of the cumulative variance. Concurrent/discriminant validity was assessed using Pearson's correlation coefficients between PANAC PA subscale scores and a) the Applied Mindfulness Process Scale, AMPS (r=0.31, p=0.001), b) CG burden, ZBI (0.014, ns), c) CG depression scores, PHQ4 (0.026, ns), and (d) a 4-item CG self-efficacy measure developed by the authors (0.073, ns). Negative Appraisal (NAs) of caregiving were associated with AMPS-low positive emotional regulation (r=-0.25, p=0.013), lower self-efficacy (r=-0.55, p < 0.001), higher ZBI scores (r=0.52, p < 0.001), and greater CG depression (r-0.37, p < 0.001).	Cronbach's α estimates by subscales: Positive appraisals (Pas) (α=0.84) Negative appraisals (Nas) (α=0.82)
Losada et al. (2020)[143] Spain	ADRD	Revised Familism Scale (RFS)	Familism is dementia Three factors: (1) Familial interconnectedness ; (2) Familial obligations; (3) Extended family support	21 items, 5-point Likert scale (ranging from 0=Very much in disagreement to 4=Very much in agreement)	To enhance content validity, the authors combined 25 items from two existing scales: 14 items from the Familism Scale[144] and 11 items from the Attitudinal Familism Scale.[145] The structural validity of the initial 25-item scale was determined by EFA employing a polychoric correlation matrix and a weighted least square method for factor extraction and Oblique rotation to account for factor correlations. After eliminating four items and repeating EFA and a Horn's parallel analysis, a 3-factor model accounted for 53.22% of variance of the assessed construct. Goodness-of-fit indices for the EFA model were acceptable (e.g., RMSEA=0.06; CFI=0.97, SRMR=0.05; and TLI=-0.95). Divergent validity was established through a hierarchical regression model using the RFS total scores as outcomes through a series of hierarchical regression analyses. One "Familism" factor was entered in each of the regressions in a first step. In a second step, a "Familism" factor different from that entered in the first step was entered. A significant incremental change in percentage of explained variance (R^2) provided an estimate of the *unique*, *construct-specific* component for each factor.	Cronbach's α, full scale =0.85 Cronbach's α by subscales: Familial interconnectedness (α=0.82) Familial obligations (α=0.74) Extended family support (α=0.74)
Maltby et al. (2020)[146] United Kingdom	Mixed	Adult Carers for Older Adults Quality-of-Life Questionnaire	Quality-of-life (including both the traditional and nontraditional roles of caregiving). Six factors: (1) Feelings of exhaustion; (2) Adoption of a traditional role; (3) Ability to care; (4) Personal growth; (5) Caring support; (6) Financial matters	24 items, 4-point Likert scale (1=Never, 2=Some of the time, 3=A lot of the time, 4=Always)	Authors *combined items from two previous scales*: 40 items from the original version of Adult Carers Quality of Life[147] Questionnaire and 21 items developed by Lawrence et al. (2008).[148] Content validity was assessed through the examination of item wording by authors until they reached consensus on clarity and content relevance. The structural validity of the initial 61-item scale was established through EFA using a mixed sample of CGs from the United Kingdom (N=308). PAF extraction followed by a Promax rotation resulted in a 6-factor solution that was confirmed by a Horn parallel analysis. *Two replication studies* using competing model formulations (CFA and a bifactor model) were conducted using two independent samples from the United States (N=164) and China (N=131) using a reduced 24-item scale and the same 6-factor structure. The *bifactor model* was the best fitting model producing satisfactory goodness-of-fit indices per sample: United States (RMSEA=0.06; CFI=0.947; and non-normed fit index, NNFI=0.93). China (RMSEA=0.04; CFI=0.94; and non-normed fit index, NNFI=0.92).	Cronbach's α estimates by subscales and country (USA, China): Feelings of exhaustion (α=0.83; α=0.77) Adoption of a traditional role (α=0.90; α=0.51) Ability to care (α=0.88; α=0.58) Personal growth (α=0.84; α=0.59) Caring support (α=0.85; α=0.76) Financial matters (α=0.84, α=0.82)
Mckenna et al. (2020)[149] United Kingdom	ADRD	Alzheimer's Patient Partners Life Impact Questionnaire (APPLIQue) (Questionnaire specific to AD spousal carers	Needs-based quality-of-life One factor: Quality-of-life	25 items, 3-point scale (1=Lower, 2=Medium, 3=Higher)	Content validity was established through cognitive debriefing interviews with 76 CGs, across the five countries included in the study, to assess and comment on the applicability, relevance and comprehensiveness of the questionnaire, instructions, and omitted aspects of their experiences. Structural validity was demonstrated by a Rasch analysis producing a unidimensional scale supporting internal construct validity. Concurrent validity was assessed by correlating scores on the APPLIQue with other scores that tap into related constructs: the Nottingham Health Profile (NHP) and the General Well-Being Index (GWBI). Spearman's rank correlations between NHP subscales and APPLIQue were "moderate" ranging from	Internal consistency reliability was assessed by the polychoric-based ordinal version of coefficient α (α=0.93) Test-retest reliability (two-week interval) was assessed with Spearman's correlation with a sample of 95 respondents (r=0.88).

					0.27 to 0.74. Spearman's rank correlations between GWBI and APPLIQue was also moderate (rho= 0.67). All correlations were significant (p < 0.05).	PSI produced by the Rasch analysis =0.85.
Perry-Duxbury et al. (2020)[150] Germany; Ireland; Italy; The Netherlands; Norway; Portugal; Sweden; United Kingdom	ADRD	Investigating Choice Experiments for the Preferences of Older People Capability-based measure of general quality of life: the ICEPOP Capability (ICECAP-O) instrument.	General capability wellbeing Five domains: (1) Attachment (Love & friendship) (2) Security (Thinking about the future) (3) Role (Doing things that makes you feel valued) (4) Enjoyment (Enjoyment and pleasure) (5) Control (Independence)	5 items (only one item per domain), 4-point Likert scale (1=No capability, 2=A little capability, 3=A lot of capability, and 4=Full capability)	Content validity of ICECAP has been ascertained in earlier work (e.g., Grewal et al., 2006) through in-depth interviews with selected informants aged 65 and over to explore their views about what is important to them in terms of quality of life. Concurrent validity. There was a moderate positive Spearman's correlations between the ICECAP-O scores and the EQ-5D-5L utility tariff (rho=0.46, p <0.01) and EQ-VAS scores (rho=0.45, p < 0.01), a moderate negative correlation with the EQ-5D-5L health problems index (rho=-0.45, p < 0.01), and a strong positive correlation with the CarerQol tariff (rho=0.53, p <0.01) and CarerQol-VAS scores (rho=0.54, p < 0.01). Group discriminant validity. Student's t tests or ANOVA were performed to identify significant differences in ICECAP-O scores by subgroups. ICECAP-O scores significantly discriminated between informal carers who were (a) old or young, (b) employed or unemployed, (c) with low or high "positive affect index" (PAI) scores, (d) in danger or not in danger of social isolation scores (LSNS), and (e) who felt they could or could not continue caregiving for 2 years or more "perseverance time" (PT) scores.	This study did not assess reliability in the international population of informal carers of people with dementia. Note: Two prior studies, however, reported "good" test-retest reliabilities of the scale but in older 70 year-olds (non-patients) (Horder et al., 2016)[151] and frail older adults (Van Leeuwen et al., 2015).[152]
Teresi et al. (2020)[153] United States	ADRD	Perceived Stress Scale (PSS)	Perceived stress One factor: Stress	10-item, 5-point Likert scale (0=Never, 1=Almost never, 2=Sometimes, 3=Fairly often, 4=Very often)	The structural validity of the PSS scale was established through PCA using Varimax rotation and polychoric correlations followed by CFAs. Dimensionality was also examined with the bifactor CFA model with polychoric correlations. PCA was conducted separately for the total sample and selected subgroups: age, education, and language of the test (Spanish-English). The PCA suggested a one-factor/component model explaining 54% of the variance for the total sample and ranged from 50% to 57% for subgroups. The bifactor CFA confirmed the unidimensionality of the scale producing goodness-of fit indices within acceptable thresholds (RMSEA=0.044, CFI=0.996). To provide validity evidence based on the internal structure of the PSS scale, differential item functioning (DIF) was examined for *age*, *education*, and *language* using the graded response IRT model. In general, the magnitude and impact of DIF were minimal across the groups examined.	Cronbach's α ordinal estimate, full scale=0.902, McDonald's ω=0.904, and the bifactor model explained common variance, ECV=68.34. IRT-based reliability measures were examined at selected points along the underlying latent continuum (attribute levels). The average reliability estimate for the total sample was 0.89 and ranged from 0.88 to 0.90 for subgroups. Test-retest reliability (6-month interval) examined over three follow-up waves (with samples N=343, 301, and 219). McDonald's ω estimates were about 0.90 across waves.
Thompson et al. (2020)[154] United States	ADRD	Fear of Incompetence—Dementia Scale (FOI-D)	Fear of incompetence in the context of relationships with a close family member diagnosed with dementia. Three factors: (1) Interaction Concerns; (2) Caregiving Concerns; (3) Knowledge Concerns	58 items, 7-point Likert scale (1=Not at all concerned to 7=Extremely concerned)	Content validity was established through a literature review on instruments measuring related constructs and focus groups that resulted in an initial 80-item pool that was pilot tested with 15 dementia caregivers for clarity and suitability. Based on the feedback, seven items were added and a preliminary 87-item scale was field-tested. The structural validity of the scale was established by iterative EFAs, using ML as the factor extraction approach, followed by CFAs to cross-validate the identified factors structure. The iterative analyses resulted in a final 58-item scale that supported a 3-factor structure. Goodness-of-fit indices for the CFA model were acceptable (e.g., RMSEA=0.05; CFI=0.91; and TLI=0.91). Concurrent validity. Only the "Interaction Concerns" subscale was significantly and negatively correlated with a single item assessing "relationship quality/satisfaction" (Pearson's r = −0.11, *p* = 0.01). The "Knowledge Concerns subscale" was significantly and negatively correlated with scores on the Dementia Knowledge Scale (DKS) (r =-0.20, p < 0.001). All FOI-D subscales were significantly and negatively related to Dementia Attitudes Scale (DAS) (r=-0.30 to -0.09) and the Burden Scale for Family Caregivers (BSFC-S) (r=-0.18 to -0.16). Discriminant/divergent validity. Pearson's correlations between scores on all FOI-D subscales and the Caregiver Self-Efficacy Scale (CSES) scores were, as expected, relatively low ranging from -0.13 to -0.07.	Cronbach's α estimates by subscales: Caregiving Concerns (α=0.90) Knowledge Concerns (α=0.90) Interaction Concerns (α=0.96) Test-retest reliability (with N=58 and approximately 10-week interval) was estimated with the ICC per subscale (all ICC's ≥ 0.75).
Voormolen et al. (2021)[155] The Netherlands; Germany; Ireland; United Kingdom; Sweden;	ADRD	The Care-Related Quality of Life (CarerQol) questionnaire	CG burden and wellbeing (happiness) Seven dimensions: (1) Fulfillment; (2) Relationship problems; (3) Mental health problems; (4) Daily activity problems; (5) Financial problems; (6) Social support with care; (7) Physical health problems	CarerQol-7D: 7 items, 3-point Likert scale (1=No, 2=Some, 3=A lot) CarerQol-VAS: 1 item, Visual analog scale (ranging from 0=Completely unhappy to 10=Completely happy).	Content validity. A previously published study on the initial phase of the scale development by Brouwer et al. (2006),[156] reported reviewing the literature and existing burden measures to create a comprehensive set of dimensions of family CG burden that were likely to be most important describing their experience. The authors also conducted a small pilot to gather preliminary information of dimensions of CG burden that might have been ignored in the instrument. The pilot also showed that the instrument was clear and understandable for CGs and easy to use. The previous study tested the tool in a heterogeneous (non-disease specific) sample of informal CGs (N=175). The current study tested the tool in a sample of family CGs of individuals with dementia (N=433). Concurrent validity was established by a significant positive Spearman's rank correlation coefficient (rho=0.530, p<0.001) between total scores on the 7-item CarerQol and the "ICEpop Capability measure	No reliability of the scale in the population of dementia CGs is reported.

Norway; Italy;Portugal					for Older people" (a broad measure of wellbeing) as well as a significant negative correlation (rho= -0.44, p<0.001) with the "EuroQol-5D-L" (a measure of health-related quality-of-life).	
Rose et al. (2021)[157] United States	ADRD	Family Quality of Life in Dementia (FQOL-D) scale.	Impact of dementia caregiving on family quality-of-life Four factors: (1) Family interactions (2) Wellbeing (3) Disease-related support/medical care (4) CG support	41 items, 5-point Likert-type scale (ranging from 1=Very dissatisfied to 5=Very satisfied)	Face validity was established by 2 persons with early stage ADRD and six family CGs who provided input regarding the clarity, readability, and content of the items included on the proposed FQOL-D instrument. Content validity was established by a panel of experts in ADRD research and care from across the United States who reviewed items for clarity of expression. A Delphi method was employed to identify important factors of family quality of life in dementia given 5 previously identified domains and preliminary items. Items were retained by panel consensus. Experts were given the opportunity to write in additional items not originally included. The final item pool comprised 43 items. Structural validity was assessed by factor analysis with PCA as the extraction method and Varimax rotation to increase interpretability of the factors/components. The PCA provided support for a 4-factor solution that explained 52% of the variance in the scale items. Concurrent validity was established by correlating the FQOL-D scale with three scales: 1) the "Family Resource", 2) the Family "Adaptation, Partnership, Growth, Affection, Resolve" (APGAR), and 3) the "Surrogate Decision Making Self-Efficacy scales". Increased FQOL-D scores were associated with higher scores in each of these scales. Pearson's correlations ranged from 0.39 to 0.46 (*p*-values < 0.01).	The Cronbach's α, full scale = 0.951.
Clemmensen et al. (2021)[158] Denmark	ADRD	Dementia Carer Assessment of Support Needs Tool (DeCANT)	Support needs Four factors: (1) Environmental factors (2) Activity and participation components (3) Personal factors (4) Body structure/function components (wellbeing)	25-item, 4-point Likert scale (0=No; 1=Yes, A little more; 2=Yes, quite a bit more; 3= Yes, very much more)	Face and content validity were established iteratively. Face validity was conducted through cognitive interviews with carers. Content validity was conducted by a panel of experts representing dementia carers, in general, or professionals in the area of dementia from different professions and care settings. The expert panel independently evaluated the representativeness, relevance, and clarity of the items using a rating scale. The structural validity of the scale was established through CFA and the evaluation of competing models. The final 4-factor structure produced acceptable goodness of fit indices (e.g., RMSEA=0.073; CFI=0.946, and TLI=0.938).	Cronbach's α by subscales: Environmental factors (α=0.84) Activity and participation components (α=0.80) Personal factors (α=0.73) Body structure/function components (α=0.84)
Durepos et al. (2021)[159] Canada	ADRD	Caring Ahead: Preparing for End-of-Life with Dementia Questionnaire	Preparedness for end-of-life Four factors: (1) Actions; (2) Dementia Knowledge; (3) Communication; (4) Emotions and Support Needs	20 items, 7-point Likert scale (ranging from 1=Strongly disagree to 7=Strongly agree)	Content validity was established by first conducting semi-structured interviews with a sample of bereaved CGs of persons with dementia to identify preparedness core concepts and generate measurable indicators (items). Indicators of preparedness were defined as questionnaire items and further reduced and refined through a Delphi survey with CGs and professional experts. The structural validity of the scale was determined by PCA as the factor/component extraction method and Promax rotation producing a 4-factor model that explained 61.7% of the cumulative variance in the scale items. Concurrent validity was demonstrated by Pearson's correlations between a single-global "preparedness question" and the scores on the four subscales. Correlations ranged from (0.43-0.55, $p < 0.001$).	Cronbach's α by subscales: Actions (α=0.85) Dementia Knowledge (α=0.86) Communication (α=0.78) Emotions and Support Needs (α=0.80) Test-retest reliability was calculated with the ICC and an N=32 (average of 28.9 days interval). Estimates by subscales: Actions (ICC=0.89); Dementia Knowledge (ICC=0.95); Communication (ICC=0.87); Emotions and Support Needs (ICC=0.91)
Wuttke-Linnemann et al. (2021)[160] Germany	Mixed	Resilience and Strain Questionnaire (ResQ-Care)	Resilience (inner attitude towards caregiving and energy sources) Strain (caregiving difficulties and burden) Four factors: (1) Inner attitude (IA); (2) Sources of energy (SE); (3) Difficulties dealing with the person in need of care (DIFF); (4) General burdens of my living situation (GB)	20 items, 4-point Likert scale (0=No, 1=Rather no, 2=Rather yes, 3=Yes)	Content validity. Authors developed a-20-item pool based on a literature review on CG burden constructs underlying published scales. The structural validity of ResQ-Care was established through an EFA with ML likelihood factor extraction method and Oblimin rotation to interpret the factor structure. The EFA revealed a 4-factor structure explaining 43.3% of variance in scale items. (Authors acknowledge that sample size was not adequate for a cross-validation study.) Concurrent validity was examined by calculating Pearson's correlations between the ResQ-Care subscales and the Brief Resilience Scale (BRS), the Perceived Stress Scale (PSS-4), and the Geriatric Depression Scale (GDS-15). The results confirmed the convergent validity for the subscales. For example, correlations between the two strain subscales (DIFF and GB) and the resilience BRS scores were negative (-0.27 and -0.37, respectively). As expected, however, correlations between the resilience subscales (IA and SE) and resilience BRS scores were positive and low to moderate in magnitude (0.52 and 0.37, respectively).	Cronbach's α and McDonald's ω estimates by subscales: Inner attitude (α=0.67; ω=0.68) My sources of energy (α=0.71; ω=0.72) Difficulty dealing with the person in need of care (α=0.81; ω=0.81) General burdens of living situation (α=0.82; ω=0.83).
Gallego-Alberto et al. (2021)[161] Spain	ADRD	The Interpersonal Triggers of Guilt in Dementia Caregiving Questionnaire (ITGDCQ)--	Guilt : guilt-triggering behavior employed by the care recipient Two factors: (1) Care recipient's criticism of the CG's role; (2) Personal disparagement	10 items, Each item is scored on two scales: frequency and magnitude of guilt. Frequency: 5-point Likert scale (ranging from 0=Never to 4=Always	Although the authors do not address the content validity of the scale, a literature review is conducted establishing the rationale for the development of the ITGDCQ subscales to address the lack of measures capturing the occurrence and frequency of behaviors performed by the care recipient and other relatives that may act as guilt triggers. The structural validity of the scale was established by EFA followed by a Horn's parallel test to determine the scale dimensionality, and a CFA. The analyses supported a 2-factor structure. Goodness-of-fit indices for the CFA model were acceptable (e.g., RMSEA=0.04; CFI=0.97; and TLI=0.94).	The Cronbach's α, full scale = 0.81. Cronbach's α by subscales: Care recipient's criticism of the CG's role (α=0.73) Personal disparagement (α= 0.80)

		Scale I: Care Receiver (ITGDCQ-CR)		Magnitude: 5-point Likert scale (ranging from 0=Not at all to 4=Extremely)	Concurrent validity was established by calculating Pearson's correlations between the Caregiver Guilt Questionnaire (CGQ) developed by Losada et al. (2010)[74] and the two subscales: (1) Care recipient's criticism of the CG's role (r=0.33, p<.01) and (2) Personal disparagement-CG guilt (r=0.44, p<.01)	
		Scale II. Other Relatives (ITGDCQ-OR)	Guilt: guilt-triggering behavior employed by other relatives (e.g., siblings, husband) Two factors: (1) Accusations of harming the patient; (2) Shifting responsibility onto the CG	8 items, As above, each item is scored on two scales: frequency and magnitude of guilt.	The structural validity was established by EFA followed by a Horn's parallel test to determine the scale dimensionality, and a CFA. The analyses supported a two-factor structure. Goodness-of-fit indices for the CFA model were acceptable (e.g., RMSEA=0.01; CFI=0.99; and TLI=0.99). Concurrent validity was established by Pearson correlations between the Caregiver Guilt Questionnaire (CGQ) developed by Losada et al. (2010)[74] and the two subscales. Only the "shifting responsibility onto the CG guilt" subscale was associated with CGQ (r = 0.25, p < 0.01).	The Cronbach's α, full scale =0.78. Cronbach's α by subscales: Accusations of harming the care recipient. (α=0.81) Shifting responsibility onto the CG (α= 0.80)
Horton et al. (2021)[162] United Kingdom	ADRD	Impact of DEmentia on CARers (SIDECAR) Battery: SIDECAR-D: Direct Impact on Carers	Carers needs and quality-of-life (QoL) One factor: Direct Impact on Carers	(The Impact of DEmentia on CARers (SIDECAR) battery has a total of 39 items. The following are the items per scale) 18 items, binary response options: Agree/ Disagree	Content validity was established through interviews with 42 carers of a relative with dementia living in the community and generating an initial bank of items based on the interviews. Items were further subject to checks regarding ambiguity, content, and face validity. Twenty-two cognitive interviews with carers were conducted to pretest and assess response formats. The structural validity for the original 70-item bank was established by EFA followed by Geomin (Oblique) rotations to increase factor structure interpretability. EFA revealed a 4-factor solution. Within each identified factor, a Rasch analysis for scale refinement was conducted iteratively producing three final separate scales: SIDECAR-D, SIDECAR-I, and SIDECAR-S. The concurrent validity of the SIDECAR scales was assessed with Spearman's rank correlations (all *p-values* < 0.001) between total scores in each of the scales and (a) a measure of wellbeing (the Short Warwick–Edinburgh Mental Well-being Scale, SWEMWBS) and (b) a measure of health valuation (the EuroQoL Group Visual Analogue Scale, EQ-5D VAS). (Hypothesized to be negatively correlated with SIDECAR scales scores.) Spearman's rank correlation (SWEMWBS, SIDECAR-D) r= -0.57; Spearman's rank correlation (EQ-5D VAS, SIDECAR-D) rho= -0.35 Responsiveness: SIDECAR-D demonstrated a "moderate" responsiveness, ES=0.43.	The Cronbach's α, full scale =0.83. Test-retest reliability (within 6 weeks with N=100 carers) was estimated with the ICC=0.86. PSI obtained from a Rasch analysis of the scale=0.81.
		SIDECAR-I: Indirect Impact on Carers	Carers needs and QoL: One factor: Indirect Impact on Carers	10 items, binary response options: "agree"/"disagree"	Concurrent validity Spearman's rank correlation (SWEMWBS, SIDECAR-I) rho= -0.40 Spearman's rank correlation (EQ-5D VAS, SIDECAR-I) rho= -0.21 Responsiveness: SIDECAR-I demonstrated a "small" responsiveness effect size, ES=0.29	The Cronbach's α, full scale =0.70. PSI obtained from a Rasch analysis of the scale=0.58. Test-Retest reliability (within 6 weeks) estimated with ICC= 0.86.
		SIDECAR-S: Support and Information	Carers needs and QoL: One factor: Support and information	11 items, binary response options: Agree/ Disagree	Concurrent validity Spearman's rank correlation (SWEMWBS, SIDECAR-S) rho= -0.36 Spearman's rank correlation (EQ-5D VAS, SIDECAR-S) rho= -0.24 Responsiveness: SIDECAR-S demonstrated a "small" responsiveness effect size, ES=0.11	The Cronbach's α, full scale =0.81. PSI obtained from a Rasch analysis of the scale=0.69. Test-retest reliability (within 6 weeks) estimated with ICC=0.85.
Schlomann et al. (2021)[163] Germany	ADRD	Berlin Inventory of Caregiver Stress-Dementia (BICS-D) Note: A test battery with 6 dimensions)	CG Stress: Subjective & objective burden Six dimensions: (1) Objective practical caregiving tasks (5 subscales-25 items) (2) Subjective burden from behavior change (6 subscales-26 items) (3) Conflicts in perceived needs and positive aspects of care (6 subscales-28 item) (4) Role conflict (2 subscales-9 items) (5) Aggression toward the patient (one scale-6 items) (6) Coping (5 subscales-27 items)	121 items, (across 25 subscales) Mixed response options per domain: (1) 5-point Likert scale (from Hardly to Not at all) (2) 5-point Likert scales (varied per subscale) (3) & (4) & (5) 5-point Likert scale (from Never to Always)	Content validity. The development of the inventory is based on stress-theory models that conceptualize burden as a situation-specific, multidimensional construct. An initial literature review and extensive qualitative data on stress from a pilot testing of N=80 caregiving relatives served as the basis from the generation of the initial pool of items. The pilot testing resulted in the refinement of the item pool and item reduction. Face to face interviews with CGs were applied to discuss the item selection. Structural validity. A total of six separate PCAs with Varimax rotation and inter-item correlations were applied to examine the factorability of each domain. The proportion of variance explained per domain varied from 56.6% to 64.5%. Concurrent validity. The 25 subscales were significantly (p-values < 0.05) correlated with the following criterion measures: wellbeing (assessed with CES-D, Self-esteem, Quality of life management and positive relationships to others) and a measure assessing "the sum of physical illnesses. Most of the subscales measuring "Objective practical caregiving" had low, but statistically significant correlations with the wellbeing criterion scales. Most of the subscales included in the "Coping" domain had relatively low correlations with both the wellbeing and the "Sum of physical illness" criterion measures. The responsiveness (sensitivity to change) of some of the BICS-D subscales was demonstrated by significant burden-reducing effects over a period of 3 months on a) practical caregiving tasks, b) subjective burden, and c) subjectively perceived need conflicts. (These results were obtained by comparing responses from 36 CGs using day-care and a matched sample of 30 non-day care users.)	The Cronbach's α estimates across the 25 subscales ranged between 0.72 to 0.95. Guttman's split-half reliability estimate per subscale varied from 0.21 to 0.90.
Cheon et al. (2022)[164]	ADRD	The Competence Scale in	Competence in managing behavioral and psychological	28 items, 5-point Likert scale	Face validity was assessed by asking five family CGs to identify inappropriate questions checking the ease of understanding, item length, and readability. Content validity assessments were conducted by	The Cronbach's α, full scale =0.922. Each sub-factor estimate ranged from

Korea		Managing Behavioral and Psychological Symptoms of Dementia (CS-MBPSD)	symptoms of dementia. Six factors: (1) Person-centered attitude, (2) Introspection for improvement, (3) Symptom occurrence analysis, (4) Application of various strategies, (5) Awareness of symptoms, (6) Caring for one's own mind and body.	(ranging from 1=Strongly disagree to 5=Strongly agree. Note: The last item is a single general question for the self-evaluation of overall competence in managing behavioral and psychological symptoms of dementia.	eight experts with the initial pool of items. Items were deleted or revised according to the experts' opinions. After the content validation, 39 of the initial 48 items remained. The structural validity of the CS-MBPSD was established through EFA and CFAs. EFA used principal components to extract and identify the factors/components followed by Varimax rotations. After further revisions, the analysis with 28 items revealed six factors/components with loadings per factor ranging from 0.493 to 0.789. Next, CFA models using the six-factor structure underlying the 28 items were estimated with a cross-validation sample (N=230). Goodness-of-fit indices, however, were found to be below recommended thresholds (RMSEA = 0.08, CFI = 0.81, and TLI = 0.79) indicating poor model-data fit. Standardized regression weights, (SRW), CR and AVE were used to assess the *reliability and convergent validity* of the factors extracted through the CFA model. The resulting SRWs ranged from 0.529 to 0.769; CR values ranged from 0.726 to 0.889; and the AVE values from 0.385 to 0.538. (Note: recommended thresholds are SRW>0.50, CR>0.70, and AVE>0.50.) Concurrent validity was established estimating Pearson's correlation between the CS-MBPSD total scores against, respectively, the Behavior Management Skill-BMS, the Visual Analogue Scale-VAS, and one general question (the last item) of the CS-MBPSD. (CS-MBPSD total scores were moderately correlated with a general question (CS-MBPSD item 29) (r=0.534, p < .01), the BMS (r=0.396, p < .01), and the VAS (r=0.339, p < .01).	0.610 to 0.846. Test-retest reliability (two-week interval) was calculated with the ICC with nine participants. The ICC for the total score was 0.781 (p=0.004) The ICC of Factors 1 to 6 ranged from 0.151 to 0.701 (very poor to moderate).
Wawrzicznyet al. (2022)[165] France	ADRD	Control and Stimulation in Dementia Caregiving (CSDC-13) Scale	CG management behaviors and approaches. Two factors/components: (1) Negative control behaviors, (2) Positive stimulation behaviors.	13 items, 5-point Likert scale (ranging from 1=Strongly disagree to 5=Strongly agree)	Content validity was established by five expert reviewers and 10 CGs who assessed items in terms of expression of a single, unambiguous idea; ease of understanding; and relevance and usefulness in clinical practice. Structural validity was established through PCA and CFA. The PCA used a Varimax rotation to explore item loadings and scree plots to define a final 2-component solution of the 13-item scale explaining 46.20% of the cumulative variance. CFA analyses for the 13-item scale exhibited a satisfactory goodness of fit indexes (e.g., RMSEA=0.08, CFI=0.91; TLI=0.90). Concurrent and discriminant validity were established through Pearson's correlations between factors (subscales) and criterion measures. For example, "Negative control" scores were significantly (p-values < 0.001) correlated with anxiety (0.25), burden (0.25) and impact on finances (0.22). "Positive stimulation" scores were significantly correlated with self-esteem (r = 0.44). As expected, "Positive stimulation" scores were not associated with anxiety (r= -0.06) or depression (r= -0.10).	Cronbach's α estimates by subscales: Negative control (α=0.82) Positive stimulation (α=0.70) Test-retest reliability (15-day interval, N=63) was 0.62 for the "Negative control" factor and 0.71 (p < 0.001) for the "Positive stimulation"
Gallego-Alberto et al. (2022)[166] Spain	ADRD	Caregiving Compassion Scale (CCS)	Compassion and distress Two factors: (1) Distress from witnessing the care recipient suffering (2) Motivation/disposition for helping or alleviating distress of their relative with dementia	10 items, 5-point Likert scale (ranging from 1=strongly agree to 5=strongly disagree)	Support for the content validity of the CSS is provided by its original developer (Schulz et al., 2017).[167] The present study analyzes its psychometric properties in a sample of dementia CGs. The structural validity was established through EFA using a ML likelihood estimator and Geomin rotation followed by a Horn parallel analysis to determine the optimal number of factors to retain. The solution supported a two-factor structure. Concurrent validity. Scores of the total compassion scale (CCS) showed significant and positive Pearson correlations with guilt levels (r=0.23, p < -0.01), overall frequency of behavioral and psychological symptoms of dementia (BPSD) (r=0.20, p < 0.01), and frequency (r=0.31, p < 0.01) and reactions (r=0.26, p < 0.01) of the RMBPC depressive behaviors subscale.	The Cronbach's α, full scale = 0.81. McDonald's ω, full scale=0.83 Cronbach's α and McDonald's ω by subscales: Distress from witnessing the care-recipient suffering (α=0.79; ω=0.79) Motivation/disposition for helping (α=0.72; ω=0.79)
Park et al. (2022)[168] United States	ADRD	Pre-Loss Grief-10-Dementia (PG-10-D)	Pre-loss grief One factor: Grief symptoms	10 items, 5-point Likert scale (from 1=Almost never to 5=Always)	The PG-12 was originally developed for non-AD carers and contained 12 items.[169] The current study reduces, adapts, and validates the scale with a sample of dementia CGs. The structural validity of the scale was assessed through iterative CFA producing a final one-factor (unidimensional) model with 10 items and factor loadings ranging from 0.53 to 0.85. Goodness-of-fit indices were within acceptable ranges (e.g., RMSEA=0.06; CFI=0.97; and TLI=0.96).	The Cronbach's α, full scale = 0.89
Bernaards et al. (2022)[170] United States United Kingdom Australia Canada Czechia France Germany Italy Korea	ADRD	27-item Zarit Caregiver Interview for Alzheimer's Disease (ZCI-AD-27)	Burden impact of caregiving Twelve factors/domains: (1) Physical (2) Emotional (3) Social (4) Daily life (5) Exhaustion (6) Dependence (7) Worry (8) Role perception (9) Financial impact (10) Difficulty with medication,	27 items, 11-point numerical rating scale with items ranging from (0=None to 10=All of the time) or (0=Not at all to 10=Extremely)	The structural validity of the scale was evaluated through iterative CFAs. A final CFA model with a second order factor (comprised of Physical, Emotional, Social, and Daily life) named "Humanistic Impact", provided a satisfactory fit to the data. Loadings for the multi-item factors Exhaustion, Dependence, Worry, and Role perception, and the single-item "factors" Overall Difficulty of caregiving, Difficulty with medication, Financial impact, and Sadness did not meet the stringent fit criteria. Authors, however, considered the full 12-factor model "acceptable" based on less stringent criteria. Goodness-of-fit indexes were below usual thresholds (e.g., RMSEA=0.07; CFI=0.91; and GFI=0.87). Convergent validity. Correlations between the items *with their own dimensio*n were satisfactory (≥ 0.40) for the following 8 domains: Physical, Emotional, Social, Daily life, Exhaustion, Dependence, Worry, and Role perception. Discriminant validity was met by all items in the *Dependence* and *Worry* scores and by all the *Humanistic impact domains* and *Role perception*. No items from the *Exhaustion* score met the discriminant validity criterion.	The Cronbach's α estimates for the subscales ranged from 0.66 for the Exhaustion score to 0.93 for the Humanistic Impact-Total score. Test-retest reliability was assessed with a subset of 219 care partners at Week 24 calculating the ICC. The ICC for the 12 domains ranged from 0.71 to 0.53 (Emotional Impact to Difficulty with medication respectively).

Poland Spain Sweden			(11) Overall difficulty of caregiving, (12) Sadness		Concurrent validity. Stronger Spearman's correlations were observed between the ZCI-AD-27 domains and scales with related concepts (e.g., the Alzheimer's Disease Cooperative Study-Basic ADLs and the Humanistic Impact-Total domain; rho= -0.30, p < 0.001). Also the correlation between ADL Total score and the Dependence scores was rho=0.35, p < 0.001. Responsiveness. A subset of 312 caregivers was used to assess responsiveness of ZCI-AD-27 to "detect change" at Week 52. Effect sizes showed a small increase in ZCI-AD-27 scores for those reporting an "improved experience" on the *Caregiver Global Impression of Change-Alzheimer's Disease*.	
Bhatt et al. (2022)[171] United Kingdom	ADRD	Family Stigma Instrument (FAMSI)	Contribution of stigma to burden among carers of people with dementia. Three domains: (1) Stigma by Association; (2) Positive Aspects of Caring; (3) Affiliate Stigma (with 3 subdomains: affective, behavioral and 'perceived')	26-items, 5-point Likert scale (1=Strongly disagree, 2=Somewhat disagree, 3=Neither disagree/agree, 4=Somewhat agree, 5=Strongly Agree)	The theoretical basis for the development of FAMSI is presented in Mitter et al., 2018.[172] The current study validates the instrument in a sample of dementia CGs. Only the concurrent validity was examined. The Rosenberg Self-Esteem Scale (RSES) was used to measure self-esteem of CGs. Authors hypothesized that stigma by association and affiliate stigma would be negatively correlated with RSES, whereas Positive aspects of caring subscale would be positively correlated with RSES. However, *no significant correlations between the FAMSI scales and RSES were observed.* Correlations ranged from r= 0.04 (p=0.74) to r=0.12 (p=0.32). Note: Authors define "stigma" directed at family carers of a stigmatized individual as 'stigma by association' or 'courtesy stigma'. When stigma by association becomes internalized, termed 'affiliate stigma', it can have negative affective, behavioral and cognitive consequences, such as unhappiness, withdrawal and sense of inferiority.	Cronbach's α by subscales: Stigma by association (α =0.917) Positive aspects of caregiving (α =0.72) Affiliate stigma (total) (α = 0.858) Subdomains of Affiliate Stigma: Affective (α =0.857); Perceived (α=0.875); Behavioral (α =0.759) Test-retest reliability estimates (2-week interval, N=70) obtained with ICC's ranged, from 0.73 (Affiliate stigma total) to 0.82 (Stigma by association).
Cartwright et al. (2022)[173] United Kingdom	ADRD	Multidimensional scale of perceived social support (MSPSS)	Perceived adequacy of social support Three Factors: (Factors are the sources of social support) (1) Family (2) Friends (3) Significant	12-Items, 7-point Likert type scale (ranging from 1=Very strongly disagree to 7=Very strongly agree)	The content validity of MSPSS was established by its original developer (Zimet et al., 1988).[174] The current study evaluates the measure's full psychometric properties in a sample of dementia CGs. The structural validity of the MSPSS scale was established through CFA yielding a 3-factor solution: All 12 items significantly loaded onto their hypothesized factor. Standardized factor loadings ranged from 0.79 to 0.93. The CFA analysis replicated the 5-factor structure and indicated a good model fit (e.g., GFI =0.967, CFI=0.959, and RMSEA=0.048). Concurrent validity. HADS scores were significantly and negatively correlated with the total MSPSS scores (r = 0.48, p<0.001), as well as 'significant other' (r= 0.34, p<0.001), 'family' (r=0.33, p<0.001) and 'friends' (r=0.45, p<0.001). The total MSPSS score was significantly positively correlated with the SF-12 physical component score(PCS) (r=0.17, p=0.003) and mental component score (MCS) (r=0.32, p<.0001)	Cronbach's α by subscales: Significant other (α=0.93); Family (α=0.94); Friends (α=0.92) Test-retest reliability (28 to 42.5 days interval) of the full MSPSS scale was estimated in a subsample of 58 participants with the ICC=0.90. Test-retest reliability per subscales: Significant other (ICC=0.89); Family (ICC=0.86); Friends (ICC =0.84)
Kim et al. (2022)[175] Australia	Mixed	Dementia Public Stigma Scale (DePSS)	Stigma Five factors: (1) Fear and discomfort (2) Incapability and loss (3) Acknowledgement of personhood (4) Burden (5) Exclusion	16 items, (items were statements about dementia and people living with dementia) 7-point Likert-type scale (ranging from 1=Strongly disagree to 7=Strongly agree)	Content validity was established by an expert panel who reviewed items for relevance and clarity of expression. The structural validity of DePSS was evaluated through EFA and CFA. EFA used ML likelihood as factor extraction method and Oblique rotation to increase factor interpretability producing a 5-factor structure. The CFA analysis replicated the 5-Factor structure and indicated a good model fit (e.g., GFI=0.967, CFI=0.959, and RMSEA=0.048). Tests of measurement invariance were conducted to examine the generalizability of the DePSS between gender and exposure groups (knowing or not knowing someone with dementia). The fit of the model was consistent with that of the configural model for both gender and exposure groups. That is, the findings indicated that all items designed to measure the public stigma of dementia are operating equivalently across gender and exposure groups.	Cronbach's α, full scale =0.818. Cronbach's α by subscales showed moderate to high reliability. Cronbach's α ranged from 0.738 to 0.805.
Hosseini et al. (2022)[176] Iran	ADRD	Family Caregivers' Hardiness Scale (FCHS)	Hardiness Five factors: (1) Religious Coping; (2) Self-Management; (3) Empathic Communication; (4) Family Affective Commitment; (5) Purposeful Interaction	21 items, 5-point Likert scale (1=Never, 2=Rarely, 3=Sometimes, 4=Often, 5=Always)	Face validity was attained by asking 11 family caregivers to examine items in terms of the level of difficulty, relevancy, or ambiguity. Content validity. Twelve experts in nursing, psychology, and instrument development were asked to evaluate the items in terms of grammar, wording, item allocation, and scaling. The structural validity of the FCHS scale was examined using EFA and CFA on a split sample of participants (N=435 was split into two subsamples: EFA sample with N=210 and a cross-validation sample for the CFA analysis with N=225. EFA used ML for factor extraction and Promax rotation. Horn's parallel analysis and Exploratory Graph Analysis revealed a two-factor structure. A CFA supported the 2-factor structure determined by EFA. The results showed all of the model fit indices were in the acceptable range (e.g., CFI=0.93, TLI=0.92, and RMSEA=0.065). Four out of the five factors (religious coping, self-management, empathic communication, and family affective commitment) showed *convergent validity* (AVE ranged from 0.50 to 0.62 and CR range from 0.75 to 0.89). The full factor structure also showed *discriminant validity*. Note: AVE > 0.50 and CR > 0.70 (or CR > AVE) are considered minimum requirements of convergent validity. Discriminant validity is achieved if the heterotrait-monotrait ratio (HTMT) of the correlations matrix values are all < 0.85.[177]	With the exception of the factor "purposeful interaction", the internal consistency reliability estimates (Cronbach's α and McDonald's ω) for the subscales were > 0.70. Religious coping (α=0.889, ω= 0.900) Self- management (α=0.880, ω=0.882) Empathic communication (α= 0.764, ω=0.766) Family affective commitment (α=0.749, ω=0.773) Purposeful interaction (α=0.691, ω=0.692) The stability of the CCS was also assessed by the ICC with the test-retest reliability method (two-week

						interval) with N=15 cases. (ICC=0.903).
Sharif-Nia et al. (2022)[178] Iran	ADRD	The Care Challenge Scale (CCS)	Caregiving challenges Two factors: (1) Effective role-play challenges reflecting physical, emotional, and psychological aspects of CGs' health. (2) Lack of social-financial support reflecting effects of caregiving on social life.	10-items, 5-point Likert scale (1=Never, 2=Rarely, 3=Sometimes, 4=Often, 5=Always)	Face validity was performed by asking 10 family CGs to examine items in terms of the level of difficulty, relevancy, or ambiguity in answering. Content validity. Twelve experts in nursing, psychology, and instrument development were asked to evaluate the items in terms of grammar, wording, item allocation, and scaling. The structural validity of the scale was examined using EFA and CFA on a split sample of participants. That is, N=435 was split into two subsamples: EFA sample with N=210 and a cross-validation sample for the CFA analysis with N=225. Horn's parallel analysis and Exploratory Graph Analysis revealed a two-factor structure. CFA confirmed the factor structure determined by EFA. Commonly used goodness of fit indexes indicated a satisfactory solution (e.g., CFI=0.929, TLI=0.903, and RMSEA=0.042). Only the first factor (Effective Role Play Challenges) showed discriminant validity (heterotrait-monotrait ratio of correlations matrix (HTMT=0.765) and *convergent validity* (AVE=0.537 and CR=0.848).	Cronbach's α and McDonald's ω by subscales: Effective role play challenge (α=0.838; ω=0.837) Lack of social - financial support (α=0.765; ω=0.773) The stability of the CCS was assessed by evaluating the ICC with the test-retest reliability method (two-week interval) in 30 family CGs. The scale stability was acceptable (ICC = 0.902).
Sharif-Nia et al. (2023)[179] Iran	ADRD	Care Stress Management Scale (CSMS)	Stress management Two factors: (1) Emotional-focused coping; (2) Problem-focused coping	8 items, 5-point Likert scale (1=Never, 2=Rarely, 3=Sometimes, 4=Often, 5=Always)	Face and content validity were established as in the previous study by Sharif-Nia et al. (2022)[178] The structural validity of the scale was examined using EFA and CFA on a split sample of participants: EFA sample (N=210) and a cross-validation sample for the CFA (N=225). EFA yielded a 2-factor solution explaining 51% of the total variance. Horn's parallel analysis and Exploratory Graph Analysis also revealed a 2-factor solution. CFA confirmed the factor structure determined by EFA. Goodness of fit indexes indicated a satisfactory solution (e.g., CFI=0.980, TLI=0.971, and RMSEA=0.052). Only the first factor (Emotional-focused coping) showed discriminant validity (heterotrait-monotrait ratio of correlations matrix, HTMT=0.76) and *convergent validity* (AVE=0.537 and CR=0.848).	Cronbach's α and McDonald's ω by subscales: Emotional-focused coping (α=0.774; ω=0.778); Problem-focused coping (α=0.791; ω=0.802) The stability of the CSMS was assessed by evaluating the ICC with the test-retest method (two-week interval) in 25 family CGs. (ICC=0.844).
Kuzmik et al. (2023)[180] United States	ADRD	Modified Caregiver Strain Index (MCSI)	Caregiving strain Two factors: (1) Individual experiences of burden; (2) Repercussions on the CG's life Factors include the following "domains:" financial, physical, psychological, social, and personal.	13 items, 3-point scale (0=No, 1=Yes, sometimes, 3=Yes, on a regular basis)	Content validity was assessed by the original developer of CSI.[181] The scale was later modified by Thornton & Travis, (2003).[182] The current study validates the modified scale among dementia CGs. Structural validity. CFA was performed to test the one- and two-factor models of the MCSI identified in prior studies. The two-factor model provided a better fit. Factors were labeled: individual experiences of burden and repercussions on the CG's life. Reported "goodness-of-fit" measures were within acceptable thresholds (e.g., CFI=0.932; RMSEA=0.076, and SRMR=0.027. Predictive validity was evaluated using three separate linear regression models controlling for CG's gender, age, race, education and living status.. Higher MCSI scores were significantly associated with higher outcome scores on the HADS-Anxiety; Subscale Depression, HADS-Depression and the Short-Form of the ZBI. (All p-values < 0.001.) Measurement Invariance. Tests of measurement invariance by race (configural, metric, and scalar) were conducted to determine whether the factor structure of the MCSI scale was invariant by race. A multigroup CFA model produced results confirming measurement invariance by race.	This study did not report reliability measures for the sample dementia CGs). Note. A previous study by Thornton and Travis (2003)[182] using the MCSI reported a Cronbach's α of 0.90 and a test–retest (2-week interval) reliability coefficient of 0.88. However, these estimates were obtained from a *mixed* sample of CGs of older adults with an unspecified disease status.
Olthof-Nefkens et al. (2023)[183] The Netherlands	ADRD	Experienced Communication in Dementia Questionnaire-Caregiver (ECD-C)	Self-perceived communication (Three domains/themes: (1) Experience communication from the perspective of the CG; (2) Judgment/assessment of the conversation quality; (3) Experienced emotions with communication problems	29 items, 4-point Likert scale for agreement (0=Fully disagree to 3=Fully agree) or frequency (0=Never to 3=Every conversation)	Content validity was demonstrated in a previous study in collaboration with experts in the field of dementia and through interviews with dyads (people with dementia and their CG). Reflexive thematic analysis of the interviews was used to generate items. Further pilot testing with a small sample of dyads and discussions with dementia experts contributed to the final version of the questionnaire. Concurrent validity was assessed with Pearson correlation coefficients. All parts of ECD-C correlated substantially with both the Dementia Quality of life Instrument by the Caregiver (DQI-C) and the ZBI short form (ZBI-12). Correlations were significant ($p < 0.05$) and in the predicted direction ranging from 0.36 to 0.47. Discriminant validity coefficients were, as expected, not significant and less than 0.20 when comparing ECD-C to the MMSE, the ADLs, and the IADLs.	Cronbach's α estimates by subscales: Experience communication (α=0.78) Judgment/assessment of the conversation quality (α=0.82) Experienced emotions (α=0.75) Test-retest reliability (2-week interval, N=49) was measured by intra-class ICC's: Experience communication (ICC=0.76); Judgment/assessment of the conversation quality (ICC=0.75); Experienced emotions (ICC=0.78)
Potter et al. (2023)[184] United Kingdom	ADRD	The Long-Term Conditions Questionnaire for Carers (LTCQ-Carer)	Effectiveness of carer support (a quality-of-life measure for carers) One factor: Effectiveness of caregiving support	21 items, 5-point Likert scale (0=Never, 1=Rarely, 2=Sometimes, 3=Often, 4=Always)	Content validity was established through cognitive interviewing with carers of people living with MCI on the comprehensibility, clarity, appropriateness and content of a draft questionnaire. Structural validity. An EFA using PAF as the factor extraction method followed by a Horn's parallel analysis provided support for a one-factor solution. To evaluate concurrent validity, gold standard measures for health-related quality of life were correlated with LTCQ-Carer scores: 1) the EuroQoL five-dimensional descriptive system with visual analogue scale: EQ-5D-5L with EQ VAS; and 2) a measure for social-care-related quality of life (ASCOT-Carer). Associations with EQ-5D and ASCOT-Carer supported construct validity. Concurrent validity was supported by Pearson's correlation estimates between the LTCQ-Carer scores and the following criterion measures: a) EQ-5D-5L index value ($r=0.52$, $p < 0.001$), b) EQ VAS ($r = 0.61$, $p < 0.001$), and c) the ASCOT-Carer ($r = 0.85$, $p < 0.001$).	Cronbach's α, full scale =0.95.
Risch et al.	ADRD	The Caregiver	Dysfunctional thoughts	28 items,	Content validity. Six experts (five German, one Australian) with experience in cognitive behavior therapy	Authors provide estimates of

(2023)[185] Germany		Thoughts Scale (CTS)	Four "domains:" (1) Dysfunctional caregiving standards; (2) Self-care; (3) Dysfunctional assumptions about dementia; (4) Acceptance	5-point Likert scale (ranging from 0=Never to 4=Very often)	for dementia CGs rated each potential question for content representativeness with possible classifications of 1 (should be excluded), 2 (would need to be revised) or 3 (should be included). This process resulted in a 28-item scale. The authors conceptualized CGs' thoughts as being *formative constructs* and allocated the 28 items into four domains (subscales) based on theoretical considerations. Therefore, construct validity was evaluated through the relationship of these four subscales with theoretically meaningful correlates. Concurrent and discriminant (divergent) validity were assessed through significant ($p<0.05$) correlations between the CTS subscales and several scales: a) depression (General Depression Scale) (r=0.36), b) anxiety (HADS) (r=0.36), c) grief (Caregiver Grief Scale) (0.39), d) quality of life (WHO Quality of Life) (psychological, r=-0.31; physical, r=-0.27), e) dysfunctional thoughts (Dysfunctional Thoughts about Caregiving Questionnaire-DTQC) (r=0.29). As expected, no significant associations were obtained between the CTS subscales and the number of care recipients' behavior problems (divergent measure) (pairwise correlations ranged from 0.02 to 0.18).	interrater agreement (for the six expert raters) using the ICC for the complete initial item pool. The obtained ICC=0.77) was considered a "good" measure of the scale reliability. Note: CTS is a formative scale. *Formative constructs* don't need to be internally consistent.[186]
Pendergrass et al. (2023)[187] Germany	Mixed	Benefits of Being a Caregiver Scale (BBCS)	Benefits (or positive aspects of caregiving) One factor: Benefits conferred by caregiving and benefits leading to personal enrichment)	14 items, 5-point Likert scale (4=Strongly agree, 3=Agree, 2=Neutral, 1=Disagree, 0=Strongly disagree)	Content validity was established in a "participatory" manner by including assessment of items by experts from different disciplines and also by family CGs. Structural validity. An EFA yielded one-factor solution explaining 49.8% of the total variance of the 14-item scale. A scree plot supported the solution. Concurrent validity. The Pearson's correlation coefficient between BBCS and the Positive Aspects of Caregiving Scale (PACS) was significant (r=0.75, $p<0.001$). Expected associations were found between BBCS scores and better a) emotion-focused coping (r=0.18, $p<0.001$) and b) problem-focused coping (r=0.23, $p<0.001$). Discriminant validity. BBCS scores were not associated with a) subjective burden (r= -0.05, p=0.240) and b) dysfunctional coping (r= -0.07, p=0.142).	Cronbach's α, full scale =0.922
Pione et al. (2023)[188] United Kingdom	ADRD	Positive Psychology Outcome Measure-**Carer Version** (PPOM-C)	Hope and Resilience in family carers of persons with dementia Two factors: (1) Hope; (2) Resilience	14 items, 5-point Likert scale (0=Not true at all to 4=True nearly all of the time) Note: The reference to answer each item is the last month.	The content validity of PPOM was previously reported by Stoner et al., 2018.[189] The current study validates the scale in a sample of dementia CGs. Structural validity. A CFA supported the hypothesized two-factor structure (hope and resilience). Commonly used goodness of fit indices showed an acceptable model fit (CFI=0.904; RMSEA=0.114; SRMR=0.057. The two-factor structure showed convergent validity, AVE=0.61. Concurrent validity. HADS-D scores were significantly (*p*-values < 0.001) and negatively correlated with PPOM-C total scores (r=-0.66) and the hope and resilience subscales (r= -0.67; r= -0.58, respectively). The hope and resilience subscales were positively correlated with the SF-12 mental component score (r=0.62, r=0.57, respectively.) in addition to the PPOM-C (r=0.63). The PPOM-C, and its hope and resilience subscales were significantly correlated with the SF-12 physical component score (r=0.19, r=0.17, r=0.19, respectively). Lastly, total MSPSS scores were significantly correlated with the PPOM-C (r= 0.39), the hope (r=0.45) and resilience (r=0.29) subscales.	Cronbach's α, full scale =0.948. Cronbach's α by subscales: Hope (α=0.912) and Resilience (α=0.918) Test-retest reliability (4-week interval, N=48) was estimated using the ICC. Full PPOM-C scale (ICC=0.908) Test-retest reliability by subscales: Hope (ICC=0.891) and Resilience (ICC=0.874)
Suganuma et al. (2024)[190] Japan	ADRD	Caregiving Competence Scale for Dementia (CCSD)	Caregiving competence Five factors: (1) Positive Emotions; (2) Presence of Consultation Partners/Family Support; (3) Caregiving Burden/Coping Skills; (4) Dementia Literacy; (5) Involvement & Emotion Control	27 items, 5-point Likert scale (ranging from 5=Strongly agree (always or frequently) to 1=Strongly disagree (never)).	Face validity was assessed by asking 15 family CGs of persons with dementia to review a preliminary pool of 45 items compiled by authors from previous studies. Content validity assessments were conducted by five experts (faculty and medical professionals specializing in dementia care) with the 45-item pool. The structural validity of the scale was established through iterative EFAs and CFAs. The EFA analyses used ML and Promax rotation to extract the underlying factors and a scree plot to determine the optimal number of factors to retain. The repeated EFA models resulted in a final 27-item scale with 5 factors. CFA analyses for the 27-item scale exhibited satisfactory commonly used goodness of fit indexes (e.g., RMSEA=0.07, CFI = 0.905).	Cronbach's α, full scale =0.892 Cronbach's α by subscales: Positive Emotions (α=0.903); Presence of Consultation Partners/family support (α=0.802); Caregiving Burden/Coping Skills (α=0.743); Dementia Literacy (α=0.782); Involvement & Emotion Control (α=0.783)

Note: AD = Alzheimer's disease; ADRD = Alzheimer's disease and related dementias; ADL = Activities of Daily Living; AGFI = adjusted goodness-of-fit index; AVE = average variance extracted. A recommended threshold for convergent validity is an AVE > 0.50; CG = Caregiver; CATPCA = categorical principal component analysis; CES-D = Center for Epidemiological Studies Depression Scale; CFA = confirmatory factor analysis; CFI = comparative fit index; CR = composite reliability. A recommended threshold for convergent validity is a CR > 0.70; CVI = content validity index;[191] EFA = exploratory factor analysis; GFI = goodness of fit index; Hamilton Depression Rating Scale = HAM-D; Hospital and Anxiety Depression Scale = HADS; IADL = instrumental activities of daily living; ICC = Intra-class correlation coefficient; IFI = incremental fit index; IRT = item response theory; LSNS= Lubben Social Network Scale; ML = maximum likelihood; MLE = maximum likelihood estimation; MMSE = Mini-Mental State Examination; NPI = Neuropsychiatric Inventory; NFI = Normed Fit Index; NNFI = non-normed fit index; PAF = principal axis factoring; PCA = principal components analysis; POMS= Profile of Mood States; RMPBC = Revised Memory and Behavior Problems Checklist; RMSEA = root mean square error of approximation; SF-36 = Short form 36 Health Survey; SRMR = standardized root-mean-square residual; TLI = Tucker-Lewis Index; ZBI = Zarit Burden Interview; PSI = person separation index.[192] PSI values above 0.70 indicate good to excellent reliability in differentiating persons along the measured trait. Proposed rule of thumb thresholds for ICCs are: between 0.50 and 0.75 (moderate); ≥ 0.75 (good), and ≥ 0.90 (excellent).[193] Generally accepted threshold for "good" Cronbach's α test of reliability is considered to be ≥ 0.70. Responsiveness (longitudinal validity) refers to the ability of an instrument to detect clinically important changes over time.[194] Measures such as minimal important change (MIC), smallest detectable change (SDC), effect size (ES), and area under the receiver operating curve (ROC) can be used to describe responsiveness.