Reference	Sample	Measurement Instrument			Psychometric Characteristics	
		Name of the Scale	Domains and Constructs	Length and Format of Instrument	Validity	Reliability
Kinney & Stephens (1989)[2] United States	ADRD	Caregiver Hassles Scale (CHS)	Stress or hassles of daily living Five domains: (1) Assisting with ADLs; (2) Assisting with IADLs; (3) Cognitive status of patient; (4) Behavior of patient; (5) Social network of CG	42 items, 4-point Likert scale (1=It wasn't, 2=Somewhat, 3=Quite a bit, 4=A great deal)	Content validity was established by literature reviews to derive key domains to be measured and further discussions with CGs to refine the domains. No formal tests of structural validity were conducted. Authors reviewed correlations between an item and the total score on the assigned "domain" or subscale (minus the item). Items with weak correlations were dropped resulting in a reduction from an initial pool of 110 item to 42 items. Concurrent validity was assessed by significant Pearson correlations between (a) the CHS-ADL subscale and the London Psychogeriatric Rating Scale (LPRS) measures of physical limitations (r=0.44, *p*<.001), and (b) the CHS-behavior hassles subscale and the LPRS-irresponsible behavior (r=0.331, *p*<.02). The CHS-cognitive status of patient subscale did not correlate significantly with the LPRS measure of "cognitive confusion."	Cronbach's α estimate, full scale=0.91 Cronbach's α by subscales: ADL (Cronbach's α=0.79) Instrumental ADL (Cronbach's α=0.75) Cognitive (Cronbach's α=0.82) Behavior (Cronbach's α=0.89) Social network (Cronbach's α=0.74) Test-retest reliability (1-day interval, N=60) was estimated with Pearson's correlations. The reliability coefficient for the full scale=0.83 Test-retest reliability by subscales: ADL=0.86; IADL=0.71; Cognitive=0.80; Behavior=0.87; Social network=0.66
Lawton *et al.* (1989)[3] United States	ADRD	Caregiver Appraisal Scale (CAS)	Appraisal of caregiving stress Three factors: (1) Subjective burden; (2) Caregiving impact; (3) Caregiving satisfaction	19 items, 5-point Likert scale (ranging from 1=Never True to 5=Nearly Always True or 1=Strongly Disagree to 5=Strongly Agree)	The structural validity of CAS was evaluated first with PCA using two independent samples and secondly through a CFA. (The first independent sample reported here (N=632) consisted of AD caregivers. The second cross-validation sample comprised a mixed sample of CGs.) The results of the PCAs with the two independent samples were used to refine the original 47-item scale with 5 components/factors resulting in a reduced 19-item scale with 3 factors. The CFA was conducted with the same two independent samples confirming an underlying 3-factor structure. The first sample (N=632) yielded acceptable fit indexes (e.g., GFI=0.94, NFI=0.90). Results in the cross-validation sample were lower (e.g., GFI=0.86, NFI=0.78). Concurrent validity was established through correlations of the three CAS subscales with the following measures: Burden rating; Quality of relationship; Emotional burden; Relationship to impaired person. *Subjective burden* was highly related to Burden rating scores (r=0.65) and less strongly but significantly to all of the other scales (r's=0.28-0.33). *Caregiving satisfactio*n was less strongly related to the Burden rating (r=0.24) but strongly related to the quality of the relationship to the impaired person (r=0.50). *Caregiving impact* was highly correlated with Burden rating (r=0.57).	Cronbach's α by subscales: Subjective burden (α=0.85) Caregiving impact (α=0.70) Caregiving satisfaction (α=0.67)
Ellis *et al.* (1989)[5] United States	ADRD	Caregiver Reactions Scale (CRS)	Reactions to caregiving Seven factors/dimensions: (1) Financial impact; (2) Impact on schedule; (3) Restrictions in social activities; (4) Impact on health; (5) Caregiving role responsibility; (6) Negative reactions; (7) Family abandonment of CG	34 items, 5-point Likert scale (ranging from 1=Strongly disagree to 5=Strongly agree)	Authors did not explicitly talk about content validity of the items in the scales developed but report conducting a review of the literature to define the concepts included in each of the scales and key relationships between concepts that needed to be considered in the development of the scales. In particular, the development of the "Caregiver Reactions" scale included in-depth interviews with CGs of persons with various types of physical and cognitive impairments. A pool of 101 items were identified from both the literature review and the analysis of the interviews. The structural validity of the CRS was established through a CFA to test a theorized 7-factor structure. One of the hypothesized subscales ("restrictions in social activities") was dropped from the final solution as well as items from the original pool. The final scale consisted of 34 items and 6 factors. No GFI statistics are reported for the CFA model.	Cronbach's α by subscales: Financial impact of caregiving (α=0.77) Impact on schedule (α=0.84) Impact on health (α=0.81) Caregiving role responsibility (α=0.88) Negative reactions to caregiving (α=0.83) Family abandonment of CG (α=0.87)
		Social Resources Scale (SRS)	Perceptions of availability of social resources One factor: Availability of resources	6 items, 5-point Likert scale (ranging from 0=No assistance to 4=Most frequent amount of assistance)	The structural validity of the SRS was established through CFA to test a theorized one-factor structure. The single factor produced factor loadings with acceptable ranges (0.42-0.62). No GFI statistics are reported for the CFA model.	Cronbach's α, full scale=0.69
Kosberg *et al.*(1990)[6] United States	ADRD	The Cost of Care Index (CCI)	Consequences (or costs) of caretaking Five factors/components: (1) Personal and Social Restriction; (2) Physical and Emotional Health; (3) Value of Providing Care; (4) Patient as Provocateur; (5) Economic Costs	20 items, 4-point Likert scale (ranging from 1=Strongly agree to 4=Strongly disagree)	Content validity was established by first reviewing the literature related to the "costs" of providing care to dependent elderly persons and defining the dimensions needed to develop a bank of items. Twenty-seven items were initially derived from the input of professionals working with family CGs of frail and impaired elderly relatives. These 27 items were pilot tested to determine their ability to distinguish groups of carers caring for elders with different physical and mental impairment levels. The initial inspection of the structural validity of the CCI was not conducted in a sample of dementia CGs. Instead, authors recruited a sample of 137 CGs of clients seeking nursing home placement under the Florida Medicaid Program. A PCA with Varimax rotation yielded a final 20-item scale with 5-components/factors. The factorial structure of the CCI scale was not examined in the present study of dementia CGs.	Cronbach's α, full scale=0.79

					Concurrent validity was demonstrated by statistically significant Pearson's correlations between CCI scores and, for example, measures of caregiving functioning assessed by a) the Short Psychiatric Evaluation Schedule (SPES; r=0.27, p<0.01), b) self-reported mental health (r=0.36, p<0.001), and c) physical health (r=0.22, p <0.05). Significant correlations were also obtained between CCI scores and measures of "consequences of caregiving" assessed by a) the "ADL trouble due to patient" (0.24, p < 0.01) and b) "Tolerance for patient behavior" (-0.33, p < 0.001).	
Goodman (1991)[7] United States	ADRD	Perceived Social Support for Caregiving (PSSC)	Perceived social support One factor: Availability or adequacy of social support (or help)	9 items, 5-point Likert scale (ranging from 1=Not at all to 5=Extremely)	Content validity was demonstrated by a team of researchers writing an initial 21-item pool based on a review of literature and empirically-determined reasons for joining self-help groups. Structural validity was established through PCA with Varimax rotation using an initial 21-item scale. Inspection of item loading reduced the scale to 12 items. PCA analyses were repeated yielding two separate scales (each with one factor/component): The Perceived Support for Caregiving (PSSC) and the Social Conflict (SC) scale. The PSSC explained 42.8% of the variance. Concurrent validity was established by significant positive Pearson correlations between PSSC total scores and Natural Network Indices (r=0.26 to 0.39; p<0.02 to 0.001). These results were collected from a subsample of respondents (N=70 to 79).	Cronbach's α, full scale=0.84
		Social Conflict (SC)	Social conflict One factor: Lack or inadequacy of social support (or help)	3 items, 5-point Likert scale (ranging from 1=Not at all to 5=Extremely)	Structural validity. A PCA with Varimax rotation yielded the SC factor/component explaining 18% of the variance. Concurrent validity was established by a significant positive Pearson correlation between the total scores on the SC and the ZBI item “Do you feel that your relative currently affects your relationship with other family members or friends in a negative way” (r=0.34, p=0.001).	Cronbach's α, full scale =0.72
Theut *et al.* (1991)[8] United States	ADRD	Anticipatory Grief Scale (AGS)	Anticipatory grief (bereavement of wives whose spouses had been diagnosed with dementia) Seven domains: (1) Anger; (2) Guilt; (3) Anxiety; (4) Irritability; (5) Sadness; (6) Feelings of loss; (7) Decreased function	27 items, 5-point Likert scale (1=Strongly disagree, 2=Disagree, 3=Somewhat agree, 4=Agree, 5=Strongly agree)	Content validity. Authors report selecting items from previous grief scales, reviewing the literature on the dimensions of anticipatory grief and developing additional items based on clinical experience with wives of patients with dementia. Structural validity. No formal examination of the underlying structure of the scale is presented. Concurrent validity was established by positive and significant (p<0.001) correlations between total scores in the AGS scale and the depression, anxiety, and hostility dimensions of the Hopkins Symptom Checklist (SCL-90-R). Note: Coefficients were not reported.	Cronbach's α, full scale =0.84
Vitaliano *et al.* (1991)[9] United States	ADRD	The Screen for Caregiver Burden (SCB)	Burden or distress Two domains: (1) Objective burden or number of negative experiences; (2) Subjective burden or distress in response to experiences	25 items, Objective burden: 2-point scale (0=Did not Occur, 1=Occurrence) Subjective burden: 5-point scale (ranging from 0=No occurrence to 4=Occurrence with severe distress) (Each item received two ratings: one for objective burden and a second for subjective burden	Content validity was established by a review of extant literature on problems in AD caregiving and by asking a sample of spouse CGs of individuals with AD (N=68) what burden experiences were most salient to them. Items are scored based on two general "domains:" Objective and Subjective. That is, the scaling of the items was assesses both objective burden (OB) and subjective burden (SB). Structural validity. No formal examination of the underlying structure of the scale/subscales is presented. Concurrent validity: The SCB OB and SB subscales correlated significantly (p<0.05) with depression (0.54, 0.41, respectively), anxiety (0.43, 0.26, respectively), and suppressed anger (subjective=0.42). The SCB OB and SB subscales also had significant (p<0.05) negative correlations with morale (-0.51, -0.48, respectively). Depression, anxiety, suppressed anger, and morale were measured by the Beck Depression Inventory-Short (BDI-S), the Symptom Checklist--90 Anxiety Scale (SCL-90), the Suppressed Anger Subscale from the Anger Expression Scale, and the Satisfaction with Life Scale, respectively.	Cronbach's α by subscales: Objective burden (α=0.85) Subjective burden (α=0.89) Test-retest reliability (15-18 months apart) was estimated with Pearson’s correlations between scale administrations: Objective burden (r=0.64, p<0.001) Subjective burden (r=0.70, p<0.001)
Given *et al.* (1992)[10] United States	Mixed	Caregiver Reaction Assessment (CRA)	Reactions to caregiving burden Five factors: (1) Impact on health (2) Impact on CG's daily schedule; (3) Impact on CG's finances; (4) Relationship to CG’s sense of self-worth; (5) Friends/family support	24 items; 5-point scale (ranging from 1=Strongly disagree to 5=Strongly agree)	Content validity was established by requiring that all the members of a 5-group team agree that each item from a 101-item pool belonged to a particular construct. After a pilot test with 99 CGs, items were dropped due to lack of clarity, variability, or association with any other items reducing the item pool to a 40-item scale. Structural validity. An initial EFA with a sample of 377 participants (29.2% dementia CGs) led to the reduction of the 40-item scale to 35 items. A PCA with oblique rotation yielded a 5-factor/component solution. A scree plot confirmed the 5-factor structure. Further elimination of items due to low loadings resulted in a 24-item scale. A final re-run of the PCA model with the 24-item scale yielded a 5-factor solution accounting for 65.1% of the variance in items. Measurement (factorial) invariance tests. Using a second independent sample (also N=377; 26.8% dementia CGs) a multiple-group CFA was conducted to test the factorial invariance of the CRA instrument across diseases (Alzheimer’s vs. cancer), spouse vs non-spouses, and across time. Results showed that the subscale structure of the CRA remains stable across diverse groups of CGs and across time. The latter suggests suitability of the CRA to measure change in CG reactions. Concurrent validity was established by calculating Pearson correlations between CRA subscales and a)	Cronbach's α by subscales: Impact on health (α=0.80) Impact on schedule (α=0.82) Impact on finances (α=0.81) Sense of self-esteem (α=0.90) Friends/family support (α=0.85)

					the CES-D and b) the ADL Dependencies Scale using the combined independent samples (N=754). Finance (r=0.25, r=0.34), family support (r=0.20, r=0.39), and health (r=0.29, r=0.57) were significantly and positively correlated with ADL dependencies and depression respectively. CG esteem was significantly and negatively associated with depression (r=-0.23).	
Semple (1992)[11] United States	ADRD	Family Conflicts Scales (FCS)	Family conflict Three factors: (1) Definitions & strategies conflict; (2) Treatment of patient conflict; (3) Treatment of CG conflict	12 items, 4-point Likert scale (ranging from 1=No disagreement to 4=Quite a bit of disagreement)	Content validity was established by in-depth interviews with 20 CGs that resulted in the identification of three family conflict domains and the creation of 4 items per domain. Structural validity was established through CFA using the 12-item scale. After comparing competing model that conformed underlying theories, a 3-factor model yielded the best fit as measured by a GFI=0.98 and a chi-square/df ratio=2.9 (less than 3 is desirable). As evidence of concurrent validity the authors used Pearson correlations to show relationships between the three FCS factors/subscales and the Hopkins Symptom Checklist (HSC) measures of *anger* and *depression*. All correlations between the HSC-Anger measure and the FCS subscales were significant (p<0.001): Definitions & strategies conflict (r=0.23), Treatment of patient conflict (r=0.25), and Treatment of CG conflict (r=0.34). All correlations between the HSC-*depression* measure and the FCS subscales were also significant (p<0.001): Definitions & strategies conflict (r=0.23), Treatment of patient conflict (r=0.28), and Treatment of CG conflict (r=0.28).	Cronbach's α by subscales: Definitions & strategies conflict (α=0.80) Treatment of patient conflict (=0.86) Treatment of CG conflict (reported as "within the range between the two other subscales")
Teri *et al.* (1992)[12] United States	Mixed	The Revised Memory and Behavior Problem Checklist (RMPBC)	CG reactions to patient behavior problems Three factors/components: (1) Memory-related problems; (2) Depression problems; (3) Disruptive behaviors (The scale uses two scoring methods per item: frequency of patient behavior problems and CG distress or reaction to the patient behavior problems.)	24 items, Each item is rated on two scales. 1) Frequency of patient behavior: 5-point Likert scale (0=Never occurred, 1=Not in the past week, 2=1 to 2 times in the past week, 3=3 to 6 times in the past week, 4=Daily or more often) 2) Reaction of "upset" by CG: 5-point Likert scale (0=Not at all, 1=A little, 2=Moderately, 3=Very much, 4=Extremely)	Content validity was shown by raters sorting items into hypothesized content areas, rating items, and agreeing on items. This method reduced the original pool of 64 items to 47 items. A PCA approach with Varimax rotation was used to study the underlying dimensions of the scale using "frequency scorings". The analysis yielded a 24-item, 3-component/factor scale explaining 53.4% of the variance. Concurrent validity was examined calculating Pearson correlations between RMPBC subscales and *well-known* (benchmark) scales measuring similar constructs. Correlations were estimated separately by scoring method Behavior Frequency and Caregiver Reaction. For the Behavior Frequency scoring, significant positive Pearson correlations were obtained between the RMPBC Depression subscale and the HAM-D Scale (r=0.44, p<0.01) as well as between the RMPBC Memory-Related Problems subscale and the MMSE. For the Caregiver Reaction scoring, validity was demonstrated by significant positive Pearson correlations between all RMPBC subscales and the CES-D scale and the Caregiver Stress Scale (CSS) (all p-values < 0.01). Depression measured by the CES-D correlated with the RMPBC Memory-Related Problems (r=0.29), Depression (r=0.31), and Disruption (r=0.26) subscales. Burden measured by the CSS correlated with the RMPBC Memory-Related Problems (r=0.32), Depression (r=0.42), and Disruption (r=0.41) subscales. Discriminant validity was established for RMPBC Behavior Frequency by non-significant correlations between the RMPBC Depression subscale and the Mini-Mental State Exam (r=-0.04, p>0.05) as well as non-significant correlations between the RMPBC Memory-Related problems subscale and the HAM-D Scale (r=0.001, p>0.05).	Behavior *Frequency* Scoring: Cronbach's α, full scale =0.84 Cronbach's α by subscales: Depression (α=0.80) Memory-Related problems (α=0.79) Disruption (α=0.67) Caregiver *Reaction* Scoring: Cronbach's α, full scale =0.90 Cronbach's α by subscales: Depression (α=0.89) Memory-Related problems (α=0.88) Disruption (α=0.84)
Macera *et al.* (1993)[13] United States	ADRD	Caregiver Burden Scale (CBS)	Perceived burden Three domains: (1) Activity for which patient required help; (2) Activity for which CG provided help; (3) Stress by providing help	15 items, 2-point scale (0=No, 1=Yes)	Content validity and the creation of items was not addressed in the article. Authors reviewed the literature on perceived burden and state the importance of measuring burden associated with specific caregiving tasks. Results of the authors-developed CBS scale are presented as a pilot study. Structural validity. No examination of the underlying structure of the scale is presented. Concurrent validity for the CBS scale was established by a significant positive Pearson correlation with the CES-D (r=0.38, p<0.001).	Cronbach's α, full scale =0.87
Hinrichsen & Niederehe (1994)[16] United States	ADRD	The Dementia Management Strategies Scale (DMSS)	Management strategies Three factors: (1) Managing criticism; (2) Encouragement; (3) Active management	28 items, 5-point Likert scale (1=Never, 2=Seldom, 3=Sometimes, 4=Often, 5=Most of the time)	Content validity was established in a prior study by Niederehe & Scott (1987).[17] A 34-item pool was developed based on literature reviews, clinical work with dementia patients and family members, and pilot interviews. Structural validity was established through EFA using PAF extraction and Varimax rotation that yielded a 3-factor solution. The original 34-item pool was reduced to 28 items based on factor loadings.	Cronbach's α by subscales: Criticism (α=0.85) Encouragement (α=0.80) Active management (α=0.77)
Carruth (1996)[18] United States	Mixed	Caregiver Reciprocity Scale (CRS)	CG reciprocity Four factors: (1) Warmth and regard; (2) Intrinsic rewards for giving; (3) Love and affection; (4) Balance within family caregiving	26 items, 5-point Likert scale (ranging from 1=Strongly disagree to 5=Strongly agree)	Content validity was established by an initial 50-item pool developed from a literature review and interviews with family CGs. Two panels of experts rated items relevance and CVIs were computed. Items with low CVIs were eliminated reducing the pool to 32 items. A pilot test with 30 CGs provided data for further item reduction by "item-to-item", "item-to-subscale", and "item-to-total" correlations further reducing the scale to 30 items. Structural validity. Before attempting to establish validity, an inter-item analysis dropped four poorly-correlated items reduce the 30-item scale to 26 items. The sample (N=303) was randomly split into two subsamples to perform EFA (N=130) and CFA (N=173). An EFA performed by factor analysis with Varimax rotation yielded a 22-item, 4-factor solution that accounted for 62.9% of the variance. The CFA with the cross validation sample established the acceptability of the 4-factor model with adequate fit	Cronbach's α by subscales: Warmth and regard (α=0.89) Intrinsic rewards for giving (α=0.82) Love and affection (α=0.86) Balance within family caregiving (α=0.78) Test-retest reliability was estimated using Pearson's correlations with a convenience sample of N=35 who retested 2 weeks after the initial test.

					indexes (e.g., GFI=0.88; AGFI=0.85; RMR=0.05; TLI=0.95). AVE was used to assessed the convergent validity of the 4 factors extracted by CFA. AVE values ranged from 0.47 to 0.64. (Three of the four AVEs were slightly below the recommended threshold of 0.50.)	Test-retest reliability by subscales: Warmth and regard (r=0.70); Intrinsic rewards for giving (r=0.69); Love and affection (r=0.88); Balance within family caregiving (r=0.58)
Davis *et al.* (1997)[22] United States	ADRD	Caregiver Activity Survey (CAS)	Time spent in caregiving activities (One "total score" measure) Scores were the hours and minutes engaged in the activity during a 24 hour period (day and night) during weekdays.	6 items, The six items included: (1) communicating; (2) using transportation (3) dressing; (4) eating (5) looking after one's appearance; (6) supervising	Content validity. Special efforts were made to find terms that could be used with a variety of populations in different cultures. Several versions of the scale were pilot-tested. in different cultural settings. Specialists reviewed the scale drafts to develop cultural and linguistic equivalents in several languages. Specific definitions were provided, along with examples of what was meant for the various categories of assistance. As a result, an initial pool of 13-items was developed and tested. The results of the first analysis led to the reduction of the original 13-item scale to a 6-item CAS scale. The concurrent validity of CAS was established by significant (p-values < 0.05) Pearson correlations with the Alzheimer's Disease Assessment Scale Cognitive Subscale (ADAS-Cog) (r = 0.61), MMSE (r = −0.57) and Physical Self Maintenance Scale (PSMS) (r=0.43).	Test-retest reliability was established by retesting N=42 CGs within a 2-week interval (i.e., week 1 and week 3), and calculating the ICC. The ICC=0.85, p<.001.
Picot *et al.* (1997)[23] United States	ADRD	Picot Caregiver Rewards Scale (Picot-CRS)	Perceived CG rewards Two domains/subscales: (1) External rewards: communication from patient, health care professionals, or other entity regarding quality of care of the caregiving (2) Internal rewards: feelings of achievement and growth	24 items, 5-point Likert scale (0=Not at all, 1=A little, 2=Somewhat, 3=Quite a lot, 4=A great deal)	Content validity was established by interviews with eight family CGs to identify themes about positive feelings and changes (i.e., rewards) resulting from caregiving. Twenty-seven items were generated from caregiving literature and considering caregiving's external and internal rewards. A pilot test with 20 CGs led to a reduction from 27 to 24 items. The underlying factorial structure of the scale was not examined. Concurrent validity was demonstrated by a significant positive Pearson correlation between PCRS scores and "caregiving demands" (r=0.22, *p*<0.05) measured by Texas Research Institute of Mental Sciences Behavioral Problem Checklist (TRIMS BPC) as well as by a significant positive association between PCRS scores and palliative coping (r=0.26, *p*<0.05) measured by the Jalowiec Coping Scale. A hypothesized negative association between rewards and costs as measured by the Costs of Care Index (CCI) was not found (r=0.07, *p*>0.05).	Cronbach's α, full scale =0.83
Kaufer *et al.* (1998)[25] United States	ADRD	The Neuropsychiatric Inventory Caregiver Distress (NPI-D) Scale	Subjective CG distress One domain: Psychological distress Note: The NPI-D was developed as an adjunct scale of Neuropsychiatric Symptoms (NPI) scale. The NPI-D assesses the impact of neuropsychiatric symptoms in Alzheimer's disease (AD) patients on CG distress.	10 items, 6-point Likert scale (from 0=Not at all distressing to 5=Very distressing) Note: The 10 items represent symptoms obtained from one of the three subscales from the original NPI scale. Items assess AD CG distress for each of these symptoms.	Content validity. A preliminary version of the NPI-D included items from the three subscales contained in the original NPI scale (physical, social, and psychological distress). An initial field testing of the NPI-D scale revealed that AD CGs viewed the impact of neuropsychiatric disturbances primarily in terms of psychological or emotional distress. As a result, a revised version of the NPI-D scale excluded items from the physical and social subscales. No examination of the underlying factorial structure of the NPI-D scale is presented. Concurrent validity was established by correlating scores in the NPI-D scale in a subsample (N=69) with scores in an abridged version of the Relatives' Stress Scale (RSS) using Pearson's correlation (r=0.60, p<0.001). (The abridged RSS included 2 of the 3 subscales: personal distress and negative feelings.) The correlation between total NPI and NPI-D scores was 0.83 (p < 0.001).	Test-retest reliability was established by retesting 23.5% (N=20) of the CGs within an average of 4.5 days and correlating the scores using Pearson's correlation (r=0.92, p<.001). Interrater reliability was also calculated with the ICC between two raters of the NPI-D in 16 CGs (ICC=0.96, P < 0.001).
Zeiss *et al.* (1999)[26] United States	Mixed	Caregiver Self-Care Self-Efficacy	Self-care self-efficacy One domain: CG behaviors that reduce stress and enhance well being	10 items, Rating of confidence in performing item activity (ranging from 0%=No confidence to 100%=Completely confident).	Content validity was established through literature reviews and authors' own experiences working with CGs resulting in the development of items for two separate measures: Self-care self-efficacy and Problem-solving self-efficacy. The measures were pretested with ten CGs to improve the clarity of wording and to decide the best method for administration. As the result of the pretesting, the measures were administered by interview rather than a paper-and-pencil format. Both measures were field tested in the same sample of 217 CGs. No examination of the underlying structure of the scales is presented. Concurrent validity for the Self-Care Self-Efficacy scale was established by a significant positive Pearson correlation between Self-Care and the "network size" subscale of the Arizona Social Support Interview (r=0.30, p<.001).	Cronbach's α, full scale =0.76. Test-retest reliability with a subsample (N=39) retested after 11 weeks was a high Pearson coefficient (r=0.675, p<0.001).
		Problem-Solving Self-Efficacy	Problem-solving self-efficacy One domain: CG behaviors utilizing problem-solving skills shown to be positively related to psychological adjustment	4 items, Rating of confidence in performing item activity (ranging from 0%=No confidence to 100%=Completely confident).	Concurrent validity for the Problem-Solving Self-Efficacy scale was established by a significant positive Pearson's correlation between Problem Solving and the Logical Analysis subscale of the Daily Living Questionnaire (r=0.19, p<0.05).	Cronbach's α, full scale =0.83. Test-retest reliability with a subsample (18%) retested after 11 weeks was a high Pearson coefficient (r=0 .683, p<0.001).
Farran *et al.* (1999)[27] United States	ADRD	Finding Meaning Through Caregiving Scale (FMTCS)	Positive aspects of caregiving Three factors: (1) Loss/Powerlessness (LP); (2) Provisional meaning	43 items, 5-point Likert scale (ranging from 1=Strongly disagree to 5=Strongly agree)	Content validity was demonstrated through a preliminary qualitative study of family CGs of demented patients. Their answers to open-ended questions became six major themes from which the authors developed an initial 135-item pool for the FMTCS measure. A pilot study (N=46) established preliminary psychometric properties for a shortened 43-item FMTCS. The shortened FMTCS measure resulted from the examination of item-to-scale, item-to-item, and item-	Cronbach's α, full scale =0.91 Cronbach's α by subscales: Loss/Powerlessness (LP) (α=0.89) Provisional Meaning (PM) (α =0.88) Ultimate Meaning (UM) (α = 0.91)

			(PM); (3) Ultimate meaning (UM)		to-total correlations. This pilot produced "acceptable" reliability estimates both by three factors/subscales (0.88 to 0.95) and total scale (0.91). The pilot test-retest reliability (one-month interval), estimated with Spearman correlation, ranged from 0.85-0.89 for the three subscales and 0.80 for the full FMTCS. Given that the three original subscales had a "strong" theoretical base, the authors used CFA to establish the factorial validity of the FMTCS using an independent sample of N=215 caretakers (only N=208 had available data on the FMTCS). The CFA model confirmed the 3-factor structure identified in the previous pilot study and provided an adequate overall fit (e.g., GFI=0.998 and a coefficient of determination=0.763). Concurrent validity was established by Pearson's correlations between FMTCS scores and existing measures hypothesized to be related. Scores on the LP subscale were significantly (all p-values < 0.01) correlated with scores on a) Patient Problem behaviors (r=0.44), b) Marital tension (r=0.38), c) Global role strain (r=0.70), and d) depression (r=0.61), as measured by the CES-D. Scores on the PM subscale were significantly correlated with a) Marital satisfaction (r=0.24), b) Caregiver Satisfaction (r=0.64), and c) Personal gain(r=0.57). Scores on the UM subscale were significantly correlated with a) Religious participation (r=0.53), b) Personal religion beliefs (r=0.61), and c) Religious support satisfaction (r=0.24). Total FMTSC scores revealed similar relationships. FMTSC total scores were positively associated with measures of a) Marital satisfaction (r=0.46), b) Caregiver satisfaction (r=0.58), c) Personal gain (r=0.39), c) Religious participation (r=0.37), d) Religious beliefs (r=0.54), and e) Religious support satisfaction (r=0.21). Total FMTSC scores, however, were *negatively associated* with Patient Problem behaviors (r=-0.35), Marital tension (r=-0.49), Role strain (-0.64), and Depression (-0.60).	
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-	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

					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.	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.
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 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.	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. 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).	Cronbach's α, full scale =0.66.
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 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,	Cronbach's α, full scale =0.80.

					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 disengagement (3) Conflict resolution (4) Expressed positive affect (Freedom from Negativity) (5) Identified Patienthood	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 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	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.

			(6) Expressed anger		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)
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 4. Hardening of the arteries 5. Benign senescent forgetfulness	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 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).	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.
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	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	Cronbach's α estimates were high for the sample of spouses (0.95) and non-spouses (0.96).

				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.	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).	
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)	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	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).

			Caregiving; (7) Treatment and management		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).	
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.
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 Psychosocial (5) Family involvement (6) Caregiving demands (7) Worry (8) CG feelings Benefits/Faith (9) Spirituality and faith (10) Benefits of 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 response categories into a 0-100 rating where higher is better quality-of-life.	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. 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).	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). 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]	ADRD	Intrinsic Spirituality Scale	Spirituality One factor: Intrinsic	6 items, 11-point scale (ranging	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	Cronbach's α, full scale =0.919.

United States		(ISS)	spirituality	from 0=Spirituality answers no questions about life to 10=Spirituality answers absolutely all questions about life)	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).	Guttman's split-half reliability estimate for the full scale=0.914.
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)
Erder *et al.* (2012)[81] United States	ADRD	Caregiver-Perceived Burden Questionnaire (CPBQ): Scale 1: Caregivers' Assessment of the Patient (CAP)	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.)	20 items, Likert scale (cut-points or thresholds not provided)	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: 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).	Cronbach's α, full scale =0.88. Test-retest reliability after a 4-week 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.)
Lopez & Guarino (2013)[84]	ADRD	Surrogate Decision Making Self-Efficacy	Self-efficacy for decision making One factor: Self-efficacy	5 items, 4-point Likert scale (ranging from 1=Strongly	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	Cronbach's α, full scale =0.87

United States		Scale (SDM-SES)		disagree to 4=Strongly agree)	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).	
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.
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).
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-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.	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).
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 scale was reduced to 13 items with high loadings on a single factor explaining 50% of the total variance.	Cronbach's α estimate for the 13-item scale=0.74.
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	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).

			changes in person's self-care			
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).
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)
Weisman de Mamani et al. (2018)[121] United States	ADRD	Stigma Impact Scale (SIS)	Stigma Four domains: (1) Social Rejection; (2) Financial Insecurity; (3) Internalized Shame; (4) Social Isolation	24-items, 4-point Likert scale (ranging from 1=Strongly disagree to 4=Strongly agree)	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 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).	Cronbach's α, full scale =0.93.
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.
Davis et al. (2019)[129] United States	ADRD	Guilt After Placement Questionnaire (GAP-Q)	Guilt and ambivalence following nursing home placement One factor: Decisional guilt-reflecting guilt associated with making the decision to place	10-items, 5-point Likert scale (ranging from 0=Never to 4=Always)	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. 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).	Cronbach's α, full scale =0.92.
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)

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)
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).
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.

		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)
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
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.

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.