Supplementary Table S1. PRISMA Checklist.

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	p.1 line 3
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	p. 2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	p. 3-4 line 47-77
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	PICOS n.a. (no intervention, no comparisons), for questions see p. 4 line 79-91
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	Abstract: p. 2 line 43 & p. 5 line 98
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	RQ1: p. 5-6 line 112-136 RQ2: p. 7-8 line 167-174
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	RQ1: p. 5 line 102-103 RQ2: p. 7 line 159 & 170 & p. 8 line 173-174
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	RQ1: appendix 1 RQ2: appendix 2
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	RQ1: p. 6-7 line 138-150 RQ2: p. 8 line 176-182
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	RQ1: p. 7 line 152-154 RQ2: p. 8 line 184-192
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	RQ1: p. 7 line 152-154 RQ2: p. 8 line 184-192
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	RQ1: n.a. RQ2: p. 9 line 197-204
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	RQ1: only descriptive, so n.a. RQ2: p. 9 line 209-210
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	RQ1: only descriptive, so n.a. RQ2: p. 9-10 line 210-228

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	p. 9 line 197-199; p. 9-10 line 212-219
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	N.a.
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	RQ1: figure 1 RQ2: figure 2
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	RQ1: table 1 RQ2: table 2 (supplement 2 specified at publication level); table 3 (supplement 3 specified at publication level)
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	RQ1: n.a. RQ2: table 4 and table 5 (supplement 4 is more specified at publication level)
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Specified version table 2 reported in supplement 2 and specified version of table 5 reported in supplement 4
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	RQ1: n.a. RQ2: table 5
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	RQ1: n.a. RQ2: table 5
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	N.a.
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	p. 17 line 402-415
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	p. 20-21 line 480-499
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	p. 21 line 511-515
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	p. 22 line 529-533

For more information, visit: **www.prisma-statement.org**.

Supplementary Table S2. Search string RQ1: Caregiver-reported negative and positive impact of caregiving.

Database	Search component	Search string
Embase	1. Caregiving	'care giv*':ab,ti OR 'care taker*':ab,ti OR 'caregiv*':ab,ti OR 'carer*':ab,ti OR 'caretaker*':ab,ti
	2. Impact	('burden':ab,ti OR 'burnout':ab,ti OR 'distress*':ab,ti OR 'load':ab,ti OR 'loss':ab,ti OR 'overload':ab,ti OR 'strain':ab,ti OR 'stress':ab,ti) OR ('benefits':ab,ti OR 'competence*':ab,ti OR 'life change*':ab,ti OR 'satisfaction':ab,ti OR 'self-efficacy':ab,ti OR 'meaning*':ab,ti OR 'positive aspect*':ab,ti OR 'mastery':ab,ti OR 'reward*':ab,ti OR 'skill*':ab,ti)
	3. Diagnosis	('cerebral':ab,ti OR 'cerebrovascular':ab,ti OR 'cva*':ab,ti OR 'poststroke':ab,ti OR 'stroke':ab,ti) OR ('myelopath*':ab,ti OR 'paraplegia*':ab,ti OR 'quadriplegia*':ab,ti OR 'spinal cord':ab,ti OR 'tetraplegia*':ab,ti OR 'transverse lesion':ab,ti OR 'transverse myelitis':ab,ti) OR ('amputat*':ab,ti OR 'ampute*':ab,ti OR 'dismembered':ab,ti)
	4. Total	1 AND 2 AND 3
PsychINFO	1. Caregiving	'care giv*'.ab,ti. OR 'care taker*'.ab,ti. OR 'caregiv*'.ab,ti. OR 'carer*'.ab,ti. OR 'caretaker*'.ab,ti.
	2. Impact	('burden'.ab,ti. OR 'burnout'.ab,ti. OR 'distress*'.ab,ti. OR 'load'.ab,ti. OR 'loss'.ab,ti. OR 'overload'.ab,ti. OR 'strain'.ab,ti. OR 'stress'.ab,ti.) OR ('benefits'.ab,ti. OR 'competence*'.ab,ti. OR 'life change*'.ab,ti. OR 'satisfaction'.ab,ti. OR 'self-efficacy'.ab,ti. OR 'meaning*'.ab,ti. OR 'positive aspect*'.ab,ti. OR 'mastery'.ab,ti. OR 'reward*'.ab,ti. OR 'skill*'.ab,ti.)
	3. Diagnosis	('cerebral'.ab,ti. OR 'cerebrovascular'.ab,ti. OR 'CVA*'.ab,ti. OR 'poststroke'.ab,ti. OR 'stroke'.ab,ti.) OR ('myelopath*'.ab,ti. OR 'paraplegia*'.ab,ti. OR 'quadriplegia*'.ab,ti. OR 'spinal cord'.ab,ti. OR 'tetraplegia*'.ab,ti. OR 'transverse lesion'.ab,ti. OR 'transverse myelitis'.ab,ti.) OR ('amputat*'.ab,ti. OR 'ampute*'.ab,ti. OR 'dismembered'.ab,ti.)
	4. Total	1 AND 2 AND 3
CINAHL	1. Caregiving	(TI ("care giv" OR "care taker*" OR "caregiv*" OR "carer*" OR "caretaker*")) OR (AB ("care giv" OR "care taker*" OR "caregiv*" OR "carer*" OR "caretaker*")) OR (MH ("Caregivers"))
	2. Impact	(TI ("burden" OR "burnout" OR "distress*" OR "load" OR "loss" OR "overload" OR "strain" OR "stress")) OR (AB ("burden" OR "burnout" OR "distress*" OR "load" OR "loss" OR "overload" OR "strain" OR "stress")) OR (TI ("benefits" OR "competence*" OR "life change*" OR "satisfaction" OR "self-efficacy" OR "meaning*" OR "positive aspect*" OR "mastery" OR "reward*" OR "skill*")) OR (AB ("benefits" OR "competence*" OR "life change*" OR "satisfaction" OR "self-efficacy" OR "meaning*" OR "positive aspect*" OR "mastery" OR "reward*" OR "skill*"))
	3. Diagnosis	(TI ("cerebral" OR "cerebrovascular" OR "CVA*" OR "poststroke" OR "stroke")) OR (AB ("cerebral" OR "cerebrovascular" OR "CVA*" OR "poststroke" OR "stroke")) OR (MH ("Stroke")) OR (TI ("myelopath*" OR "paraplegia*" OR "quadriplegia*" OR "spinal cord" OR "tetraplegia*" OR "transverse lesion" OR "transverse myelitis")) OR (AB ("myelopath*" OR "paraplegia*" OR "quadriplegia*" OR "spinal cord" OR "tetraplegia*" OR "transverse lesion" OR "transverse myelitis")) OR (TI ("amputat*" OR "ampute*" OR "dismembered")) OR (AB ("amputat*" OR "ampute*" OR "dismembered")) OR (MH ("amputation"))
	4. Total	1 AND 2 AND 3
Pubmed (Medline)	1. Caregiving	care giv*[tiab] OR care taker*[tiab] OR caregiv*[tiab] OR carer*[tiab] OR caretaker*[tiab] OR "Caregivers"[Mesh]
	2. Impact	(burden[tiab] OR burnout[tiab] OR distress*[tiab] OR load[tiab] OR loss[tiab] OR overload[tiab] OR strain[tiab] OR stress[tiab]) OR (benefits[tiab] OR competence*[tiab] OR life change*[tiab] OR satisfaction[tiab] OR self-efficacy[tiab] OR meaning*[tiab] OR positive aspect*[tiab] OR mastery[tiab] OR reward*[tiab] OR skill*[tiab])
	3. Diagnosis	(cerebral[tiab] OR cerebrovascular[tiab] OR CVA*[tiab] OR poststroke[tiab] OR stroke[tiab] OR "Stroke"[Mesh]) OR (myelopath*[tiab] OR paraplegia*[tiab] OR quadriplegia*[tiab] OR spinal cord[tiab] OR tetraplegia*[tiab] OR transverse lesion[tiab] OR transverse myelitis[tiab]) OR (amputat*[tiab] OR ampute*[tiab] OR dismembered[tiab] OR "amputation"[Mesh])
	4. Total	1 AND 2 AND 3

Supplementary Table S3. Search string RQ2: Clinimetric properties measures.

Database	Search component	Search string
Embase	1. Names measures	'Caregiver Strain Index':ab,ti OR 'Carer Strain Index':ab,ti OR 'Relative Stress Scale':ab,ti OR 'Zarit Burden Interview':ab,ti OR 'Zarit Caregiver Burden Interview':ab,ti OR 'Zarit Caregiving Burden Scale':ab,ti OR 'Zarit Burden Inventory':ab,ti OR 'Zarit Care Burden Scale':ab,ti OR 'Burden Interview':ab,ti OR 'Caregiver Burden Interview':ab,ti OR 'Caregiver Burden Scale':ab,ti OR 'Caregivers Burden Scale':ab,ti OR 'Caregiver Burden Inventory':ab,ti OR 'Oberst Caregiving Burden Scale':ab,ti OR 'Oberst Caregiver Burden Scale':ab,ti OR 'Pearlins Caregiving and Stress Process Tool':ab,ti OR 'Caregiver Stress Scale':ab,ti OR 'Pearlin Burden Scale':ab,ti OR 'Caregiver Burden Scale':ab,ti OR 'Burden Assessment Scale':ab,ti OR 'Burden Assessment Schedule':ab,ti OR 'Bakas Caregiving Outcomes Scale':ab,ti OR 'Bakas Caregiving Outcome Scale':ab,ti OR 'Bakas Caregiver Outcomes Scale':ab,ti OR 'Bakas Caregiver Outcome Scale':ab,ti OR 'Bakas Caregiving Outcomes Score':ab,ti OR 'Bakas Caregiving Outcome Score':ab,ti OR 'Bakas Caregiver Outcomes Score':ab,ti OR 'Bakas Caregiver Outcome Score':ab,ti OR 'Caregiver Reaction Assessment Instrument':ab,ti OR 'Sense of Competence Questionnaire':ab,ti OR 'Carer Assessment Scale':ab,ti OR 'Carers Assessment of Managing Index':ab,ti OR 'Positive Aspects of Caregiving Questionnaire':ab,ti OR 'Positive Aspects of Caregiving Scale':ab,ti OR 'Appraisal of Caregiving Scale':ab,ti OR 'Caregiving Appraisal Scale':ab,ti
	2. Terms indicating clinimetric properties	'intermethod comparison'/exp OR 'data collection method'/exp OR 'validation study'/exp OR 'feasibility study'/exp OR 'pilot study'/exp OR 'psychometry'/exp OR 'reproducibility'/exp OR reproducib*:ab,ti OR 'audit':ab,ti OR psychometr*:ab,ti OR clinimetr*:ab,ti OR clinometr*:ab,ti OR 'observer variation'/exp OR 'observer variation':ab,ti OR 'discriminant analysis'/exp OR 'validity'/exp OR reliab*:ab,ti OR valid*:ab,ti OR 'coefficient':ab,ti OR 'internal consistency':ab,ti OR (cronbach*:ab,ti AND ('alpha':ab,ti OR 'alphas':ab,ti)) OR 'item correlation':ab,ti OR 'item correlations':ab,ti OR 'item selection':ab,ti OR 'item selections':ab,ti OR 'item reduction':ab,ti OR 'item reductions':ab,ti OR 'agreement':ab,ti OR 'precision':ab,ti OR 'imprecision':ab,ti OR 'precise values':ab,ti OR 'test-retest':ab,ti OR ('test':ab,ti AND 'retest':ab,ti) OR (reliab*:ab,ti AND ('test':ab,ti OR 'retest':ab,ti)) OR 'stability':ab,ti OR 'interrater':ab,ti OR 'inter-rater':ab,ti OR 'intrarater':ab,ti OR 'intra-rater':ab,ti OR 'intertester':ab,ti OR 'inter-tester':ab,ti OR 'intratester':ab,ti OR 'intra-tester':ab,ti OR 'interobeserver':ab,ti OR 'inter-observer':ab,ti OR 'intraobserver':ab,ti OR 'intra-observer':ab,ti OR 'intertechnician':ab,ti OR 'inter-technician':ab,ti OR 'intratechnician':ab,ti OR 'intra-technician':ab,ti OR 'interexaminer':ab,ti OR 'inter-examiner':ab,ti OR 'intraexaminer':ab,ti OR 'intra-examiner':ab,ti OR 'interassay':ab,ti OR 'inter-assay':ab,ti OR 'intraassay':ab,ti OR 'intra-assay':ab,ti OR 'interindividual':ab,ti OR 'inter-individual':ab,ti OR 'intraindividual':ab,ti OR 'intra-individual':ab,ti OR 'interparticipant':ab,ti OR 'inter-participant':ab,ti OR 'intraparticipant':ab,ti OR 'intra-participant':ab,ti OR 'kappa':ab,ti OR 'kappas':ab,ti OR 'coefficient of variation':ab,ti OR repeatab*:ab,ti OR (replicab*:ab,ti OR 'repeated':ab,ti AND ('measure':ab,ti OR 'measures':ab,ti OR 'findings':ab,ti OR 'result':ab,ti OR 'results':ab,ti OR 'test':ab,ti OR 'tests':ab,ti)) OR generaliza*:ab,ti OR generalisa*:ab,ti OR 'concordance':ab,ti OR ('intraclass':ab,ti AND correlation*:ab,ti) OR 'discriminative':ab,ti OR 'known group':ab,ti OR 'factor analysis':ab,ti OR 'factor analyses':ab,ti OR 'factor structure':ab,ti OR 'factor structures':ab,ti OR 'dimensionality':ab,ti OR subscale*:ab,ti OR 'multitrait scaling analysis':ab,ti OR 'multitrait scaling analyses':ab,ti OR 'item discriminant':ab,ti OR 'interscale correlation':ab,ti OR 'interscale correlations':ab,ti OR ('error':ab,ti OR 'errors':ab,ti AND (measure*:ab,ti OR correlat*:ab,ti OR evaluat*:ab,ti OR 'accuracy':ab,ti OR 'accurate':ab,ti OR 'precision':ab,ti OR 'mean':ab,ti)) OR 'individual variability':ab,ti OR 'interval variability':ab,ti OR 'rate variability':ab,ti OR 'variability analysis':ab,ti OR ('uncertainty':ab,ti AND ('measurement':ab,ti OR 'measuring':ab,ti)) OR 'standard error of measurement':ab,ti OR sensitiv*:ab,ti OR responsive*:ab,ti OR ('limit':ab,ti AND 'detection':ab,ti) OR 'minimal detectable concentration':ab,ti OR interpretab*:ab,ti OR (small*:ab,ti AND ('real':ab,ti OR 'detectable':ab,ti) AND ('change':ab,ti OR 'difference':ab,ti)) OR 'meaningful change':ab,ti OR 'minimal important change':ab,ti OR 'minimal important difference':ab,ti OR 'minimally important change':ab,ti OR 'minimally important difference':ab,ti OR 'minimal detectable change':ab,ti OR 'minimal detectable difference':ab,ti OR 'minimally detectable change':ab,ti OR 'minimally detectable difference':ab,ti OR 'minimal real change':ab,ti OR 'minimal real difference':ab,ti OR 'minimally real change':ab,ti OR 'minimally real difference':ab,ti OR 'ceiling effect':ab,ti OR 'floor effect':ab,ti OR 'item response model':ab,ti OR 'irt':ab,ti OR 'rasch':ab,ti OR 'differential item functioning':ab,ti OR 'dif':ab,ti OR 'computer adaptive testing':ab,ti OR 'item bank':ab,ti OR 'cross-cultural equivalence':ab,ti
	3. Exclusion	'addresses':it OR 'biography':it OR 'case reports':it OR 'comment':it OR 'directory':it OR 'editorial':it OR 'festschrift':it OR 'interview':it OR 'lectures':it OR 'legal cases':it OR 'legislation':it OR 'letter':it OR 'news':it OR 'newspaper article':it OR 'patient education handout':it OR 'popular works':it OR 'congresses':it OR 'consensus development conference':it OR 'consensus development conference, nih':it OR 'practice guideline':it NOT ('animals'/exp NOT 'humans'/exp)
	4. Total	(1 AND 2) NOT 3

PsychINFO	1. Names measures	'Caregiver Strain Index'.ab,ti. OR 'Carer Strain Index'.ab,ti. OR 'Relative Stress Scale'.ab,ti. OR 'Zarit Burden Interview'.ab,ti. OR 'Zarit Caregiver Burden Interview'.ab,ti. OR 'Zarit Caregiving Burden Scale'.ab,ti. OR 'Zarit Burden Inventory'.ab,ti. OR 'Zarit Care Burden Scale'.ab,ti. OR 'Burden Interview'.ab,ti. OR 'Caregiver Burden Interview'.ab,ti. OR 'Caregiver Burden Scale'.ab,ti. OR 'Caregivers Burden Scale'.ab,ti. OR 'Caregiver Burden Inventory'.ab,ti. OR 'Oberst Caregiving Burden Scale'.ab,ti. OR 'Oberst Caregiver Burden Scale'.ab,ti. OR 'Pearlins Caregiving and Stress Process Tool'.ab,ti. OR 'Caregiver Stress Scale'.ab,ti. OR 'Pearlin Burden Scale'.ab,ti. OR 'Caregiver Burden Scale'.ab,ti. OR 'Burden Assessment Scale'.ab,ti. OR 'Burden Assessment Schedule'.ab,ti. OR 'Bakas Caregiving Outcomes Scale'.ab,ti. OR 'Bakas Caregiving Outcome Scale'.ab,ti. OR 'Bakas Caregiver Outcomes Scale'.ab,ti. OR 'Bakas Caregiver Outcome Scale'.ab,ti. OR 'Bakas Caregiving Outcomes Score'.ab,ti. OR 'Bakas Caregiving Outcome Score'.ab,ti. OR 'Bakas Caregiver Outcomes Score'.ab,ti. OR 'Bakas Caregiver Outcome Score'.ab,ti. OR 'Caregiver Reaction Assessment Instrument'.ab,ti. OR 'Sense of Competence Questionnaire'.ab,ti. OR 'Carer Assessment Scale'.ab,ti. OR 'Carers Assessment of Managing Index'.ab,ti. OR 'Positive Aspects of Caregiving Questionnaire'.ab,ti. OR 'Positive Aspects of Caregiving Scale'.ab,ti. OR 'Appraisal of Caregiving Scale'.ab,ti. OR 'Caregiving Appraisal Scale'.ab,ti.
	2. Terms indicating clinimetric properties	'intermethod comparison'.sh. OR 'data collection method'.sh. OR 'validation study'.sh. OR 'feasibility study'.sh. OR 'pilot study'.sh. OR 'psychometry'.sh. OR 'reproducibility'.sh. OR 'reproducib*'.ab,ti. OR 'audit'.ab,ti. OR 'psychometr*'.ab,ti. OR 'clinimetr*'.ab,ti. OR 'clinometr*'.ab,ti. OR 'observer variation'.sh. OR 'observer variation'.ab,ti. OR 'discriminant analysis'.sh. OR 'validity'.sh. OR 'reliab*'.ab,ti. OR 'valid*'.ab,ti. OR 'coefficient'.ab,ti. OR 'internal consistency'.ab,ti. OR ('cronbach*'.ab,ti. AND ('alpha'.ab,ti. OR 'alphas'.ab,ti.)) OR 'item correlation'.ab,ti. OR 'item correlations'.ab,ti. OR 'item selection'.ab,ti. OR 'item selections'.ab,ti. OR 'item reduction'.ab,ti. OR 'item reductions'.ab,ti. OR 'agreement'.ab,ti. OR 'precision'.ab,ti. OR 'imprecision'.ab,ti. OR 'precise values'.ab,ti. OR 'test-retest'.ab,ti. OR ('test'.ab,ti. AND 'retest'.ab,ti.) OR ('reliab*'.ab,ti. AND ('test'.ab,ti. OR 'retest'.ab,ti.)) OR 'stability'.ab,ti. OR 'interrater'.ab,ti. OR 'inter-rater'.ab,ti. OR 'intrarater'.ab,ti. OR 'intra-rater'.ab,ti. OR 'intertester'.ab,ti. OR 'inter-tester'.ab,ti. OR 'intratester'.ab,ti. OR 'intra-tester'.ab,ti. OR 'interobeserver'.ab,ti. OR 'inter-observer'.ab,ti. OR 'intraobserver'.ab,ti. OR 'intra-observer'.ab,ti. OR 'intertechnician'.ab,ti. OR 'inter-technician'.ab,ti. OR 'intratechnician'.ab,ti. OR 'intra-technician'.ab,ti. OR 'interexaminer'.ab,ti. OR 'inter-examiner'.ab,ti. OR 'intraexaminer'.ab,ti. OR 'intra-examiner'.ab,ti. OR 'interassay'.ab,ti. OR 'inter-assay'.ab,ti. OR 'intraassay'.ab,ti. OR 'intra-assay'.ab,ti. OR 'interindividual'.ab,ti. OR 'inter-individual'.ab,ti. OR 'intraindividual'.ab,ti. OR 'intra-individual'.ab,ti. OR 'interparticipant'.ab,ti. OR 'inter-participant'.ab,ti. OR 'intraparticipant'.ab,ti. OR 'intra-participant'.ab,ti. OR 'kappa'.ab,ti. OR 'kappas'.ab,ti. OR 'coefficient of variation'.ab,ti. OR repeatab*'.ab,ti. OR ('replicab*'.ab,ti. OR 'repeated'.ab,ti. AND ('measure'.ab,ti. OR 'measures'.ab,ti. OR 'findings'.ab,ti. OR 'result'.ab,ti. OR 'results'.ab,ti. OR 'test'.ab,ti. OR 'tests'.ab,ti.)) OR 'generaliza*'.ab,ti. OR 'generalisa*'.ab,ti. OR 'concordance'.ab,ti. OR ('intraclass'.ab,ti. AND 'correlation*'.ab,ti.) OR 'discriminative'.ab,ti. OR 'known group'.ab,ti. OR 'factor analysis'.ab,ti. OR 'factor analyses'.ab,ti. OR 'factor structure'.ab,ti. OR 'factor structures'.ab,ti. OR 'dimensionality'.ab,ti. OR 'subscale*'.ab,ti. OR 'multitrait scaling analysis'.ab,ti. OR 'multitrait scaling analyses'.ab,ti. OR 'item discriminant'.ab,ti. OR 'interscale correlation'.ab,ti. OR 'interscale correlations'.ab,ti. OR ('error'.ab,ti. OR 'errors'.ab,ti. AND ('measure*'.ab,ti. OR 'correlat*'.ab,ti. OR 'evaluat*'.ab,ti. OR 'accuracy''.ab,ti. OR 'accurate'.ab,ti. OR 'precision'.ab,ti. OR 'mean'.ab,ti.)) OR 'individual variability'.ab,ti. OR 'interval variability'.ab,ti. OR 'rate variability'.ab,ti. OR 'variability analysis'.ab,ti. OR ('uncertainty'.ab,ti. AND ('measurement'.ab,ti. OR 'measuring'.ab,ti.)) OR 'standard error of measurement'.ab,ti. OR 'sensitiv*'.ab,ti. OR 'responsive*'.ab,ti. OR ('limit'.ab,ti. AND 'detection'.ab,ti.) OR 'minimal detectable concentration'.ab,ti. OR 'interpretab*'.ab,ti. OR ('small*'.ab,ti. AND ('real'.ab,ti. OR 'detectable'.ab,ti.) AND ('change'.ab,ti. OR 'difference'.ab,ti.)) OR 'meaningful change'.ab,ti. OR 'minimal important change'.ab,ti. OR 'minimal important difference'.ab,ti. OR 'minimally important change'.ab,ti. OR 'minimally important difference'.ab,ti. OR 'minimal detectable change'.ab,ti. OR 'minimal detectable difference'.ab,ti. OR 'minimally detectable change'.ab,ti. OR 'minimally detectable difference'.ab,ti. OR 'minimal real change'.ab,ti. OR 'minimal real difference'.ab,ti. OR 'minimally real change'.ab,ti. OR 'minimally real difference'.ab,ti. OR 'ceiling effect'.ab,ti. OR 'floor effect'.ab,ti. OR 'item response model'.ab,ti. OR 'irt'.ab,ti. OR 'rasch'.ab,ti. OR 'differential item functioning'.ab,ti. OR 'dif'.ab,ti. OR 'computer adaptive testing'.ab,ti. OR 'item bank''.ab,ti. OR 'cross-cultural equivalence'.ab,ti.
	3. Exclusion	'addresses'.pt. OR 'biography'.pt. OR 'case reports'.pt. OR 'comment'.pt. OR 'directory'.pt. OR 'editorial'.pt. OR 'festschrift'.pt. OR 'interview'.pt. OR 'lectures'.pt. OR 'legal cases'.pt. OR 'legislation'.pt. OR 'letter'.pt. OR 'news'.pt. OR 'newspaper article'.pt. OR 'patient education handout'.pt. OR 'popular works'.pt. OR 'congresses'.pt. OR 'consensus development conference'.pt. OR 'consensus development conference, nih'.pt. OR 'practice guideline'.pt. NOT ('animals'.sh. NOT 'humans'.sh.)
	4. Total	(1 AND 2) NOT 3

CINAHL	1. Names measures	(TI ("Caregiver Strain Index" OR "Carer Strain Index" OR "Relative Stress Scale" OR "Zarit Burden Interview" OR "Zarit Caregiver Burden Interview" OR "Zarit Caregiving Burden Scale" OR "Zarit Burden Inventory" OR "Zarit Care Burden Scale" OR "Burden Interview" OR "Caregiver Burden Interview" OR "Caregiver Burden Scale" OR "Caregivers Burden Scale" OR "Caregiver Burden Inventory" OR "Oberst Caregiving Burden Scale" OR "Oberst Caregiver Burden Scale" OR "Pearlin's Caregiving and Stress Process Tool" OR "Caregiver Stress Scale" OR "Pearlin Burden Scale" OR "Caregiver Burden Scale" OR "Burden Assessment Scale" OR "Burden Assessment Schedule" OR "Bakas Caregiving Outcomes Scale" OR "Bakas Caregiving Outcome Scale" OR "Bakas Caregiver Outcomes Scale" OR "Bakas Caregiver Outcome Scale" OR "Bakas Caregiving Outcomes Score" OR "Bakas Caregiving Outcome Score" OR "Bakas Caregiver Outcomes Score" OR "Bakas Caregiver Outcome Score" OR "Caregiver Reaction Assessment Instrument" OR "Sense of Competence Questionnaire" OR "Carer Assessment Scale" OR "Carer's Assessment of Managing Index" OR "Positive Aspects of Caregiving Questionnaire" OR "Positive Aspects of Caregiving Scale" OR "Appraisal of Caregiving Scale" OR "Caregiving Appraisal Scale") OR (AB ("Caregiver Strain Index" OR "Carer Strain Index" OR "Relative Stress Scale" OR "Zarit Burden Interview" OR "Zarit Caregiver Burden Interview" OR "Zarit Caregiving Burden Scale" OR "Zarit Burden Inventory" OR "Zarit Care Burden Scale" OR "Burden Interview" OR "Caregiver Burden Interview" OR "Caregiver Burden Scale" OR "Caregivers Burden Scale" OR "Caregiver Burden Inventory" OR "Oberst Caregiving Burden Scale" OR "Oberst Caregiver Burden Scale" OR "Pearlin's Caregiving and Stress Process Tool" OR "Caregiver Stress Scale" OR "Pearlin Burden Scale" OR "Caregiver Burden Scale" OR "Burden Assessment Scale" OR "Burden Assessment Schedule" OR "Bakas Caregiving Outcomes Scale" OR "Bakas Caregiving Outcome Scale" OR "Bakas Caregiver Outcomes Scale" OR "Bakas Caregiver Outcome Scale" OR "Bakas Caregiving Outcomes Score" OR "Bakas Caregiving Outcome Score" OR "Bakas Caregiver Outcomes Score" OR "Bakas Caregiver Outcome Score" OR "Caregiver Reaction Assessment Instrument" OR "Sense of Competence Questionnaire" OR "Carer Assessment Scale" OR "Carer's Assessment of Managing Index" OR "Positive Aspects of Caregiving Questionnaire" OR "Positive Aspects of Caregiving Scale" OR "Appraisal of Caregiving Scale" OR "Caregiving Appraisal Scale")
	2. Terms indicating clinimetric properties	(MH "Psychometrics") or (TI psychometr* or AB psychometr*) or (TI clinimetr* or AB clinimetr*) or (TI clinometr* or AB clinometr*) or (MH "Outcome Assessment") or (TI outcome assessment or AB outcome assessment) or (TI outcome measure* or AB outcome measure*) or (MH "Health Status Indicators") or (MH "Reproducibility of Results") or (MH "Discriminant Analysis") or ((TI reproducib* or AB reproducib*) or (TI reliab* or AB reliab*) or (TI unreliab* or AB unreliab*)) or ((TI valid* or AB valid*) or (TI coefficient or AB coefficient) or (TI homogeneity or AB homogeneity)) or (TI homogeneous or AB homogeneous) or (TI "coefficient of variation" or AB "coefficient of variation") or (TI "internal consistency" or AB "internal consistency") or (MH "Internal Consistency+") or (MH "Reliability+") or (MH "Measurement Error+") or (MH "Content Validity+") or "hypothesis testing" or "structural validity" or "cross-cultural validity" or (MH "Criterion-Related Validity+") or "responsiveness" or "interpretability" or (TI reliab* or AB reliab*) or ((TI test or AB test) or (TI retest or AB retest)) or (TI stability or AB stability) or (TI interrater or AB interrater) or (TI inter-rater or AB inter-rater) or (TI intrarater or AB intrarater) or (TI intra-rater or AB intrarater) or (TI intertester or AB intertester) or (TI inter-tester or AB inter-tester) or (TI intratester or AB intratester) or (TI intra-tester or AB intra-tester) or (TI interobserver or AB interobserver) or (TI inter-observer or AB inter-observer) or (TI intraobserver or AB intraobserver) or (TI intra-observer or AB intra-observer) or (TI intertechnician or AB intertechnician) or (TI inter-technician or AB inter-technician) or (TI intratechnician or AB intratechnician) or (TI intra-technician or AB intra-technician) or (TI interexaminer or AB interexaminer) or (TI inter-examiner or AB inter-examiner) or (TI intraexaminer or AB intraexaminer) or (TI intra-examiner or AB intra-examiner) or (TI intra-examiner or AB intraexaminer) or (TI interassay or AB interassay) or (TI inter-assay or AB inter-assay) or (TI intraassay or AB intraassay) or (TI intra-assay or AB intra-assay) or (TI interindividual or AB interindividual) or (TI inter-individual or AB inter-individual) or (TI intraindividual or AB intraindividual) or (TI intra-individual or AB intra-individual) or (TI interparticipant or AB interparticipant) or (TI inter-participant or AB inter-participant) or (TI intraparticipant or AB intraparticipant) or (TI intra-participant or AB intra-participant) or (TI kappa or AB kappa) or (TI kappa's or AB kappa's) or (TI kappas or AB kappas) or (TI repeatab* or AB repeatab*) or (TI responsive* or AB responsive*) or (TI interpretab* or AB interpretab*)
	3. Exclusion	(PT ("adressess" OR "biography" OR "case reports" OR "comment" OR "directory" OR "editorial" OR "festschrift" OR "interview" OR "lectures" OR "legal cases" OR "legislation" OR "letter" OR "news" OR "newspaper article" OR "patient education handout" OR "popular works" OR "congresses" OR "consensus development conference" OR "consensus development conference, nih" OR "practice guideline")) NOT (MH ("animals" NOT "humans"))
	4. Total	(1 AND 2) NOT 3

Pubmed (Medline)	1. Names measures	Caregiver Strain Index[tiab] OR Carer Strain Index[tiab] OR Relative Stress Scale[tiab] OR Zarit Burden Interview[tiab] OR Zarit Caregiver Burden Interview[tiab] OR Zarit Caregiving Burden Scale[tiab] OR Zarit Burden Inventory[tiab] OR Zarit Care Burden Scale[tiab] OR Burden Interview[tiab] OR Caregiver Burden Interview[tiab] OR Caregiver Burden Scale[tiab] OR Caregivers Burden Scale[tiab] OR Caregiver Burden Inventory[tiab] OR Oberst Caregiving Burden Scale[tiab] OR Oberst Caregiver Burden Scale[tiab] OR Pearlin's Caregiving and Stress Process Tool[tiab] OR Caregiver Stress Scale[tiab] OR Pearlin Burden Scale[tiab] OR Caregiver Burden Scale[tiab] OR Burden Assessment Scale[tiab] OR Burden Assessment Schedule[tiab] OR Bakas Caregiving Outcomes Scale[tiab] OR Bakas Caregiving Outcome Scale[tiab] OR Bakas Caregiver Outcomes Scale[tiab] OR Bakas Caregiver Outcome Scale[tiab] OR Bakas Caregiving Outcomes Score[tiab] OR Bakas Caregiving Outcome Score[tiab] OR Bakas Caregiver Outcomes Score[tiab] OR Bakas Caregiver Outcome Score[tiab] OR Caregiver Reaction Assessment Instrument[tiab] OR Sense of Competence Questionnaire[tiab] OR Carer Assessment Scale[tiab] OR Carer's Assessment of Managing Index[tiab] OR Positive Aspects of Caregiving Questionnaire[tiab] OR Positive Aspects of Caregiving Scale[tiab] OR Appraisal of Caregiving Scale[tiab] OR Caregiving Appraisal Scale[tiab]
	2. Terms indicating clinimetric properties	instrumentation[sh] OR methods[sh] OR Validation Studies[pt] OR Comparative Study[pt] OR "psychometrics"[MeSH] OR psychometr*[tiab] OR clinimetr*[tw] OR clinometr*[tw] OR "outcome assessment (health care)"[MeSH] OR outcome assessment[tiab] OR outcome measure*[tw] OR "observer variation"[MeSH] OR observer variation[tiab] OR "Health Status Indicators"[Mesh] OR "reproducibility of results"[MeSH] OR reproducib*[tiab] OR "discriminant analysis"[MeSH] OR reliab*[tiab] OR unreliab*[tiab] OR valid*[tiab] OR coefficient[tiab] OR homogeneity[tiab] OR homogeneous[tiab] OR "internal consistency"[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR (item[tiab] AND (correlation*[tiab] OR selection*[tiab] OR reduction*[tiab])) OR agreement[tiab] OR precision[tiab] OR imprecision[tiab] OR "precise values"[tiab] OR test-retest[tiab] OR (test[tiab] AND retest[tiab]) OR (reliab* [tiab] AND (test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] OR intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-tester[tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intra-observer[tiab] OR intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intraassay[tiab] OR intra-assay[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[tiab] OR kappa's[tiab] OR kappas[tiab] OR repeatab*[tiab] OR ((replicab*[tiab] OR repeated[tiab]) AND (measure[tiab] OR measures[tiab] OR findings[tiab] OR result[tiab] OR results[tiab] OR test[tiab] OR tests[tiab])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR "known group"[tiab] OR factor analysis[tiab] OR factor analyses[tiab] OR dimension*[tiab] OR subscale*[tiab] OR (multitrait[tiab] AND scaling[tiab] AND (analysis[tiab] OR analyses[tiab])) OR item discriminant[tiab] OR interscale correlation*[tiab] OR error[tiab] OR errors[tiab] OR "individual variability"[tiab] OR (variability[tiab] AND (analysis[tiab] OR values[tiab])) OR (uncertainty[tiab] AND (measurement[tiab]OR measuring[tiab])) OR "standard error of measurement"[tiab] OR sensitiv*[tiab] OR responsive*[tiab] OR ((minimal[tiab] OR minimally[tiab] OR clinical[tiab] OR clinically[tiab]) AND (important[tiab] OR significant[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR (small*[tiab] AND (real[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR meaningful change[tiab] OR "ceiling effect"[tiab] OR "floor effect"[tiab] OR "Item response model"[tiab] OR IRT[tiab] OR Rasch[tiab] OR "Differential item functioning"[tiab] OR DIF[tiab] OR "computer adaptive testing"[tiab] OR "item bank"[tiab] OR "cross-cultural equivalence"[tiab])
	3. Exclusion	"addresses"[Publication Type] OR "biography"[Publication Type] OR "case reports"[Publication Type] OR "comment"[Publication Type] OR "directory"[Publication Type] OR "editorial"[Publication Type] OR "festschrift"[Publication Type] OR "interview"[Publication Type] OR "lectures"[Publication Type] OR "legal cases"[Publication Type] OR "legislation"[Publication Type] OR "letter"[Publication Type] OR "news"[Publication Type] OR "newspaper article"[Publication Type] OR "patient education handout"[Publication Type] OR "popular works"[Publication Type] OR "congresses"[Publication Type] OR "consensus development conference"[Publication Type] OR "consensus development conference, nih"[Publication Type] OR "practice guideline"[Publication Type]) NOT ("animals"[MeSH Terms] NOT "humans"[MeSH Terms])
	4. Total	(1 AND 2) NOT 3

Supplementary Table S4. Definitions of measurement properties.

Measurement properties	Definition
Content validity	The degree to which the content of the measure is an adequate reflection of the construct to be measured
Structural validity	The degree to which scores of the measure are an adequate reflection of the dimensionality of the construct to be measured
Internal consistency	The degree of the interrelatedness among the items
Cross-cultural validity	The degree to which the performance of the items on a translated or culturally adapted measure are an adequate reflection of the performance of the items of the original version of the measure
Reliability	The extent to which scores for patients who have not changed are the same for repeated measurement under several conditions: e.g. using different sets of items from the same measure (internal consistency); over time (test-retest); by different persons on the same occasion (inter-rater); or by the same persons (i.e. raters or responders) on different occasions (intra-rater)
Measurement error	The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured
Criterion validity	The degree to which the scores of a measure are an adequate reflection of a "gold standard"
Hypotheses testing for construct validity	The degree to which the scores of a measure are consistent with hypotheses (for instance with regard to internal relationships, relationships to scores of other instruments, or differences between relevant groups) based on the assumption that the measure validly measures the construct to be measured
Responsiveness	The ability of a measure to detect change over time in the construct to be measured

Definitions as described in COSMIN guidelines [15–17].

Supplementary Table S5. Characteristics of the 18 measures used to assess negative or positive caregiver impact.

Measures (year)	Construct negative/ positive impact of caregiving	Original target population	Original mode of administration	# items	Completion time (min)	Question example	Response categories	Subscales (# items)	Score (min-max)	Score interpretation	Original language	Available translations (based on found records)	Copyright
Appraisal of Caregiving Scale (1991) [245]	Benefit, benign, threat, caregiving appraisal	Caregivers of patients receiving radiotherapy for cancer	Self-administered	27	20	This situation does not affect how I feel about myself	5-point-scale (very true to very untrue)	Threat, general stress, benefit (# items per subscale not found)	Mean score per subscale (1-5)	Higher scores indicate greater threat, more stress, and higher perceived benefits	English	Not found	Not found
Burden Assessment Scale (1994) [174]	Burden	Family related to individuals with severe mental disorders	Self-administered or interview-based	19	Not reported	Would you tell me to what extent you have had any of the following experiences in the past 6 months? Felt trapped by your caregiving role	4-point-scale (not at all to a lot)	Objective (10) and subjective burden (9)	Sum per subscale (10-40 and 9-36) and total score (19-76)	Higher scores indicate higher burden	English	Spanish, Puerto Rican, Swedish, Taiwanese, Turkish, German	Not found
Burden Assessment Schedule Modified (2010) [177]	Burden	Caregivers of persons with stroke	Self-administered	20	Not reported	Has your workload increased after patient's illness	3-point-scale (not at all to very much)	Financial situation, physical and mental stress, family and social relationships (# items per subscale not found)	No scale scores, interpretation items	Individual interpretation of the items	Indian	Not found	Not found
Caregiver Burden Inventory (1989) [150]	Burden	Caregivers of patients with Alzheimer	Self-administered	24	10-15	My care receiver needs my help to perform many daily tasks	5-point-scale (never to nearly always)	Time-dependence burden (5), developmental burden (5), physical burden (4), social burden (5), emotional burden (5)	Sum per sub-dimension (0-20; physical burden weighted by a factor of 1.25), no total score	Higher scores (on subscales) indicate higher burden	English	Chinese, Italian, Brazilian-Portuguese	Not found
Caregiver Burden Scale (1996) [130]	Burden	Caregivers of persons with stroke	Self-administered	22	Not reported	Do you find yourself facing purely practical problems in the care of your relative that you think are difficult to solve?	4-point-scale (not at all to frequently)	General strain (8), isolation (3), disappointment (5), emotional involvement (3), environment (3)	Average of sub-dimensions, no total score	Higher scores indicate higher burden	Swedish	English, Turkish, Persian	Not found
Caregiver Reaction Assessment (1992) [210]	Experiences with providing care	Caregivers of persons with physical impairments and Alzheimer's disease	Interview-based	24	7-15	My activities are centered around care for…	5-point-scale (strongly disagree to strongly agree)	Caregivers' self-esteem (7), lack of family support (5), impact on finances (3), impact on schedule (5), impact on health (4)	Mean subscale scores (1-5), no total score	Higher scores indicate stronger impact	English	French, Dutch, Chinese, Swedish, Japanese, Malay, Tamil, German	Not found
Caregiver Strain Index (1983) [19]	Strain	Caregivers of patients recently hospitalized for hip surgery or heart problems	Interview-based	13	<5	Feeling completely overwhelmed (e.g., because of worry about … concerns about how you will manage)	Yes/no	N.a.	Sum (0-13)	Higher scores indicate higher burden; ≥7 indicates a greater level of stress (based on mean + 1SD)	English	Dutch, French, Malay, Spanish, Taiwanese, Turkish	The Gerontological Society of America. Allowed to reproduce by permission of the publisher. Permission is granted for non-for-profit educational purposes, provided The Hartford Institute for Geriatric Nursing, Division of Nursing, New York University is cited as the source
Modified Caregiver Strain Index (modernization) (2003) [77]	Strain	Long-term family caregivers	Self-administered	13	<5	I feel completely overwhelmed	Yes/on a regular basis/ yes, sometimes/or no	N.a.	Sum (0-26)	Higher scores indicate higher burden	English	Chinese	Not found

Instrument	Construct	Target population	Administration	Items	Time	Example item	Response scale	Subscales	Scoring	Interpretation	Original language	Translations	Publisher
Carer Assessment Scale (1998) [234]	Level of difficulty in caring	Caregivers of persons with stroke	Self-administered	14	10	Indicate the extent to which the following areas cause difficulty in caring for a relative with stroke: Inner conflict caused by responsibilities	4-point-scale (no problem to great problem)	N.a.	Sum (0-42)	Higher scores indicate greater caregiver needs	Chinese and English	Not found	Not found
Carer's Assessment of Managing Index (1998) [235]	Carer ratings of helpfulness of management strategies and their own effectiveness	Carers involved in chronic care	Self-administered	38	Not reported	Please circle the number that best describes your experience: Talking over your problems with someone you trust	4-point-scale (I do not use this, to very helpful)	N.a.	No scale score, profile	Profile that captures caregiving experience	English	Italian, Polish	Not found
Modified Pearlin Burden Scale (1990) [170]	Burden	Caregivers of community-dwelling individuals with Alzheimer's disease	Self-administered	6	Not reported	Exact formulation not found	5-point-scale (not at all to to a great extent)	N.a.	Sum (6-30)	Higher scores indicate higher burden	English	Not found	Not found
Oberst Caregiving Burden Scale (1991) [23]	Difficulty associated with caregiving	Caregivers of patients with cancer	Self-administered	15	Not reported	Exact formulation not found	5-point-scale (not difficult to extremely difficult (first subscale) or none to a great amount (second subscale))	Perceptions of the difficulty (15) and time spent (15) (same items for both subscales)	Sum per subscale (15-75)	Higher scores indicate great task difficulty and more time spent on each task	English	Not found	Not found
Positive Aspects of Caregiving Questionnaire (2004) [241]	Positive aspects of caregiving	Caregivers of patients with Alzheimer's disease	Self-administered	9	Not reported	Helping your relative … makes you feel more useful	5-point-scale (disagree a lot to agree a lot)	Self-affirmation (6), outlook on life (3)	Sum (9-45), subscale sum (6-30 and 3-15)	Higher scores indicate greater caregivers' sense that caregiving generally satisfying and rewarding	English and Spanish	Chinese	Not found
Relative Stress Scale (1982) [84]	Stress	Relatives supporting elderly psychogeriatric patients living in the community	Self-administered	15	5-10	Do you ever feel you can no longer cope with the situation?	5-point-scale (never/not at all to always/considerably)	Personal distress (6), life upset (5), negative feelings (4)	Sum per subscale (0-24; 0-20; 0-16) and total sum (0-60)	Higher scores indicate higher stress	English	Italian, Norwegian	John Wiley & Sons, Ltd.
Revised 15-item Bakas Caregiving Outcomes Scale (2006) [197]	Life changes resulting from providing care	Caregivers of persons with stroke	Self-administered or interview-based	15	5-10	As a result of providing care for the person with stroke: My self-esteem…	7-point-scale (changed for the worst to changed for the best)	N.a.	Sum (15-105)	Higher scores indicate more positive caregiver outcomes	English	Greek	Not found
Sense of Competence Questionnaire (1993) [225]	Feelings of being capable of caring	Caregivers of patients with dementia	Interview-based (preferred) or self-administered	27	15-20	I feel that I cannot leave my ... alone, he/she needs me continuously	5-point-scale (agree to disagree)	Satisfaction with the patient as recipient of care (7), satisfaction with own performance as caregiver (12) and consequences of involvement in care for the personal life (8)	Sum per subscale (7-35, 12-60, 8-40) and total sum (27-135)	Higher scores indicate higher competence and lower burden	Dutch	English, German	Not found

Instrument	Construct	Target population	Administration	Items	Time (min)	Example item	Response scale	Subscales	Scoring	Interpretation	Original language	Available languages	Costs
Zarit Burden Interview (revised) (1980) [89]	Burden	Caregivers of elderly persons with senile dementia	Self-administered	22	10-15	Do you feel stressed between caring for your relative and trying to meet other responsibilities for your family or work?	5-point-scale (never to nearly always)	N.a.	Sum (0-88)	Higher scores indicate higher burden; 0-20 little/no burden; 21-40 mild/moderate; 41-60 moderate/severe; 61-88 severe (almost equal groups)	English	African, Arabic, Bengali, Bulgarian, Czech, Danish, Dutch, English, Finnish, Flemish, French, Georgian, German, Greek, Gujarati, Hebrew, Hindi, Hungarian, Indonesian, Italian, Japanese, Kannada, Kinyarwanda, Korean, Latvian, Lithuanian, Malay, Malayalam, Mandarin, Marathi, Nepali, Norwegian, Polish, Portuguese, Punjabi, Romanian, Russian, Serbian, Slovak, Spanish, Swedish, Tamil, Telugu, Thai, Turkish, Ukrainian, Urdu, Vietnamese	Free for not funded academic users, costs for funded academic research are €300 per study, and €50 per language. 1980 Steven H Zarit and Judy M Zarit
Zarit Burden Interview Short Form (2001) [126]	Burden	Caregivers of patients with a memory disorder	Self-administered	12	Not reported	Do you feel that because of the time you spend with your relative that you don't have enough time for yourself?	5-point-scale (never to nearly always)	Personal strain (9) and role strain (3)	Sum (0-48), no information found about subscale score calculations	Higher scores indicate higher burden; ≥17-48 severe/high burden	English	Hebrew, Arabic, Persian, Cantonese, Swedish	Not found

Supplementary Table S6. Characteristics of the clinimetric publications.

Measure (year of first publication)	Publication	Sample			Study design	Admin. mode	Disease characteristics			Background						
		n	Age mean in years (SD; range)	Gender % female			Disease	Disease duration	Disease severity	Caregiving setting	Country	Language	Response rate %	Distribution scores, mean (SD and range)	Missings	Floor and ceiling effects
Appraisal of Caregiving Scale (1991) [215]	Oberst, Thomas, Gass, & Ward (1989) Caregiving demands and appraisal of stress among family caregivers [256].	47	53.3 (SD not reported; 23-74)	77.0	Cross-sectional	Self-report	Cancer	Mean duration 17 months (SD/range not reported)	Moderate-extremely	Persons involved in the care of patients receiving radiotherapy for cancer, outpatient	United States of America	English	74.0	Not reported in correct subscales	3 cases excluded based on missings	Not reported
	Lambert, Yoon, Ellis, & Northouse (2015) Measuring appraisal during advanced cancer: Psychometric testing of the Appraisal of Caregiving Scale [257].	484	56.5 (13.4; range not reported)	56.8	Repeated administration (part of randomized clinical trial)	Self-report	Advanced cancer	Not reported	Not reported	Caregivers of patients with advanced cancer	United States of America	English	68.6	General stress: 3.1 (0.8; 1.0-4.9); threat: 2.8 (0.8; 1.0-4.7); benefit: 3.5 (0.6; 1.3-5.0)	Not reported	Not reported
Burden Assessment Scale (1994) [153]	Reinhard, Gubman, Horwitz, & Minsky (1994) Burden Assessment Scale for families of the seriously mentally ill [174].	Two separate samples: 94 and 94	Sample 1: 56.5; sample 2: 58.0 (SD and ranges not reported)	Sample 1: 68.1; sample 2:81.9	Sample 1: cross-sectional; sample 2: repeated administration	Self-report and interview based	Seriously mentally ill	Not reported	Not reported	Family members of severe mentally ill adults participating in 'The Club' (sample 1) and family members in an new initiative by New Jersey's Division of Mental Health Hospitals (sample 2)	United States of America	English	Not reported	Not reported	Not reported	Not reported
	Ivarsson, Sidenvall, & Carlsson (2004) The factor structure of the Burden Assessment Scale and the perceived burden of caregivers for individuals with severe mental disorders [258].	256	Not reported	Percent female not reported for caregivers (only for clients)	Cross-sectional	Self-report	Severe mental disorders	Not reported	Not reported	Caregivers for individuals with severe mental disorders	Sweden	Swedish	39.0	No total scores reported, scores were subdivided in factor scores in different age groups, gender, educational level and civil status	24 cases excluded due to missing scores	Not reported
	Guada, Land, & Han (2011) An exploratory factor analysis of the Burden Assessment Scale with a sample of African-American families [259].	94	47.0 (SD not reported; 18-80)	86.2	Cross-sectional (baseline of intervention study)	Self-report	Schizophrenia	Not reported	Not reported	Low-income African American families caring for a family member with schizophrenia	United States of America	English	Not reported	40.73 (16.51; range not reported)	Not reported	Not reported
	Murdoch, Rahman, Barsky, Maunula, & Cawthorpe (2014) The use of the Burden Assessment Scale with families of a paediatric population [260].	300	Not reported	88.0	Cross-sectional	Self-report	Paediatric psychiatric, behavioural or emotional disorder	Not reported	Not reported	Caregivers of children and youth with psychiatric, behavioural or emotional disorders	Canada	English	Not reported	49.0 (12.2; 20-75)	Not reported	No floor/ceiling effects
	Hunger, Krause, Hilzinger, Ditzen, & Schweitzer (2016) When significant others suffer: German Validation of the Burden Assessment Scale (BAS) [261].	215	32.0 (14.0; 18-77)	72.0	Cross-sectional	Self-report	Mentally ill	Not reported	Not reported	Caregivers of persons with a mental disorder currently, inpatient or outpatient psychotherapy	Germany	German	Not reported	2.3 (0.7; range not reported); mean instead of sum score	23 cases were excluded due to missings (at least 20% missing)	Not reported
Burden Assessment Schedule Modified (2010) [156]	Das, Hazra, Ray, Ghosal, Banerjee, Roy, Chaudhuri, Raut, & Das (2010) Burden among stroke caregivers: Results of a community-based study from Kolkata, India [177].	199	42.5 (14.6; range not reported)	76.0	Cross-sectional	Self-report	Stroke	Not reported	Not reported	Stroke caregivers	India	Bengali	Not reported	Distribution reported for all items separated	Not reported	Not reported
Caregiver Burden Inventory (1989) [133]	Novak, & Guest (1989) Application of a multidimensional Caregiver Burden Inventory [150].	171	60.1 (13.1; 25-87)	73.8	Cross-sectional	Interview based	Alzheimer's disease, senile dementia, organic brain syndrome	Mean care duration 63.1 months (49.3; range not reported)	Not reported	Caregivers of cognitively impaired older people	United States of America	English	Not reported	Means on the 5 different factors ranged: 2.0-7.1 (range SD: 3.0-6.0)	Not reported	Not reported

Scale	Study	N	Age	% female	Design	Method	Population	Care duration	Severity	Sample	Country	Language	Response rate	Score	Missing	Notes
	Marvardi, Mattioli, Spazzafumo, Martriforti, et al. (2005) The Caregiver Burden Inventory in evaluating the burden of caregivers of elderly demented patients: Results from a multicentre study [262].	419	58.2 (14.1; range not reported)	71.0	Cross-sectional	Interview based	Dementia	Mean duration 36.8 months (26.8; range not reported)	13.8% severe dementia; 51.0% moderate; 35.2% mild	Caregivers of elderly demented patients, outpatient	Italy	Italian	Not reported	Total score: 32.5 (18.0; range not reported); scores on subscales also reported	Not reported	Not reported
	McCleery, Addington, & Addington (2007) Family assessment in early psychosis [263].	113	Not reported	Not reported	Repeated administration (baseline, 6, 12 and 24 months)	Self-report	First-episode psychosis	Not reported	Not reported	Relatives of individuals with a first-episode of psychosis at admission to an early psychosis and follow-up	Canada	English	Not reported	Total score at baseline: 16.9 (14.2; range not reported); also scores reported for other time points and subscales	Only completely completed cases were included, no information about missings	Not reported
	Valer, Aires, Fengler, & Paskulin (2015) Adaptation and validation of the Caregiver Burden Inventory for use with caregivers of elderly individuals [264].	120	58.6 (13.7; range not reported)	73.3	Repeated administration (repeated measure after 2 weeks)	Self-report	Elderly	Median care duration 7.6 years	Not reported	Caregivers of older adults dependent on assistance to perform activities of daily living	Brazil	Brazilian Portuguese	Not reported	41.8 (21.0; range not reported)	Not reported	Not reported
	Greco, Pancani, Sala, Annoni, Steca, Paturzo, D'Agostino, Alvaro, & Vellone (2017) Psychometric characteristics of the Caregiver Burden Inventory in caregivers of adults with heart failure [265].	505	56.6 (14.9; range not reported)	52.2	Cross-sectional	Self-report	Heart failure	Mean duration 57.8 months (47.3; range not reported)	Not reported	Informal caregivers of patients with heart failure	Italy	Italian	Not reported	Only item means/SD were reported, no (sub)scale mean	Not reported	Not reported
	Farmer, Thienemann, Leibold, Kamalani, Sauls, & Frankovich (2018) Psychometric evaluation of the Caregiver Burden Inventory in children and adolescents with PANS [266].	104	Not reported	76.0	Cross-sectional	Self-report	Paediatric acute-onset neuropsychiatric syndrome	Not reported	Not reported	Parental caregivers of child/adolescent patients of a paediatric acute-onset neuropsychiatric syndrome clinic, burden measurement during active disease flare	United States of America	English	Not reported	36.7 (19.8; range not reported)	10 cases excluded due to missings, of the remaining 104 cases, 12 cases (12%) had at least one missing (which was imputed)	Not reported
	Vázquez, Otero, Simón, Bueno, & Blanco (2019) Psychometric properties of the Spanish version of the caregiver burden inventory [267].	201	56.2 (10.1; range not reported)	87.1	Cross-sectional	Interview based	Various diseases (intellectual, mental, physical or cognitive)	Mean care duration 14.5 years (11.7; range not reported)	Not reported	Non-proffessional caregivers of dependent persons	Spain	Spanish	95.7%	42.0 (15.9; 6-93)	Not reported	36.4% of the items scored as 0; 19.5% as 4
Caregiver Burden Scale (1996) [115]	Elmstahl, Malmberg, & Annerstedt (1996) Caregiver's burden of patients 3 years after stroke assessed by a novel Caregiver Burden Scale [130].	150 (Dementia: 83; stroke: 67)	Dementia: 83.3 (5.5); stroke: 69.8 (19.3); ranges not reported	Dementia: 89.2; stroke: 46.3	Repeated administration (repeated measure after 2 weeks)	Self-report	Dementia and stroke	Not reported	Not reported	Family caregivers of patients with dementia or stroke	Sweden	Swedish	Not reported	Not reported	Not reported	Not reported
	Cil Akinci, & Pinar (2014) Validity and reliability of Turkish Caregiver Burden Scale among family caregivers of haemodialysis patients [268].	161	45.4 (15.3; 18-80)	65.2	Cross-sectional	Self-report	Haemodialysis patients	Mean care duration 4.9 years (4.0; 0.5-20)	91.9% have 3 dialysis per week	Family members who provide primary care for haemodialysis patients	Turkey	Turkish	Not reported	Not reported	Not reported	Not reported
	Farajzadeh, Akbarfahimi, Maroufizadeh, Rostami, & Kohan (2018) Psychometric properties of Persian version of the Caregiver Burden Scale in Iranian caregivers of patients with spinal cord injury [269].	110	37.6 (12.1; 18-60)	45.5	Repeated administration (repeated measure after 2 weeks)	Self-report	Spinal cord injury	Mean care duration 3.2 years (1.4; range not reported)	Not reported	Family caregivers of community dwelling individuals with spinal cord injury	Iran	Persian	Not reported	General strain: 19.7 (4.5; sum score); isolation: 6.1 (2.1); disappointment: 12.7 (2.9); emotional involvement: 5.7 (1.7); environment: 7.3 (1.9); ranges not reported	9 cases excluded due to missing scores	Not reported
Caregiver Reaction Assessment (1992) [182]	Given, Given, Stommel, Collins, King, & Franklin (1992) The Caregiver Reaction Assessment (CRA) for Caregivers to persons with chronic physical and mental impairments [210].	377	61.1 (12.0; range not reported)	81.4	Cross-sectional and (partly) repeated administration	Self-report	Physical impairments, Alzheimer's disease and cancer	Not reported	Not reported	Caregivers to persons with chronic physical and mental impairments	United States of America	English	Not reported	Not reported	Not reported	Not reported

Study	N	Age	%	Design	Method	Diagnosis	Duration	Severity	Sample	Country	Language	%	Results	Missing values	Distribution
Nijboer, Triemstra, Tempelaar, Sanderman, Van den Bos (1999) Measuring both negative and positive reactions to giving care to cancer patients: Psychometric qualities of the Caregiver Reaction Assessment (CRA) [270].	181	63.4 (10.8; range not reported)	65.0	Cross-sectional (part of longitudinal study)	Self-report and interview based	Cancer	Mean duration: 32% 0-3 months; 23% 7-12 months; 14% >12 months	Not reported	Partners of colorectal cancer patients	The Netherlands	Dutch	76.0	Disrupted schedule: 2.4 (0.8); financial problems: 1.9 (0.6); lack of family support: 2.1 (0.6); health problems: 2.0 (0.6); self-esteem: 4.2 (0.4); ranges not reported	97% of the cases had no missing values; 1 case had >10% missing(3 items); range of missing values per item: 0-1.7%; standardized index of missings: 0.2	Not reported
Van Exel, Scholte Op Reimer, Brouwer, Van den Berg, Koopmanschap, & Van den Bos (2004) Instruments for assessing the burden of informal caregiving for stroke patients in clinical practice: A comparison of CSI, CRA, SCO and self-rated burden [271].	148	60.6 (13.7; range not reported)	69.6	Cross-sectional	Self-report	Stroke	Mean duration 0.5 years (SD/range not reported)	59.5% mild/severe disability; 40.5% independent	Primary informal caregivers of stroke patients	The Netherlands	Dutch	35.9	Caregivers' self-esteem: 3.8 (0.6; 2.1-5.0); lack of family support: 2.4 (0.9; 1.0-4.5); impact on finances: 3.0 (1.2; 1.0-5.0); impact on daily schedule 2.8 (0.7; 1.3-5.0); impact on health 2.9 (0.9; 1.0-4.5)	73.6% of the scores no missing values; 85.1% <10% missing values; range of missing values per item: 4.7-17.6%; standardized index missings: 9.7	The self-esteem subscale indicate a somewhat negatively skewed distribution
Post, Festen, Van de Port, & Visser-Meily (2007) Reproducibility of the Caregiver Strain Index and the Caregiver Reaction Assessment in partners of stroke patients living in the Dutch community [272].	21	55.9 (10.3; range not reported)	57.7	Repeated administration (repeated measure after 1 week)	Self-report	Stroke	Mean duration 3.0 years (0.1; range not reported)	One-sided supratentorial lesion, no disabling comorbidity, mostly ischaemic stroke (65.4%)	Partners of stroke patients living in the Dutch community	The Netherlands	Dutch	Not reported	Means at T1: disrupted schedule=2.7 (1.0); financial problems=2.4 (0.8); lack of family support=2.5 (0.9); health problems=2.3 (0.8); self-esteem=3.9 (0.5); ranges not reported	Not reported	Not reported
Persson, Wennman-Larsen, Sundin, & Gustavsson (2008) Assessing informal caregivers' experiences: A qualitative and psychometric evaluation of the Caregiver Reaction Assessment Scale [273].	209	60.0 (SD not reported; 22-86)	55.5	Cross-sectional	Self-report	Malignant disease, dementia, physical impairment	Not reported	Not reported	Informal caregivers to individuals with a malignant disease, dementia or a physical impairment	Sweden	Swedish	Not reported	Not reported	Missing data were rare (n=0-4), with the exception of question 12 (n=8)	3 of the 5 subscales were skewed (answers corresponding to a positive and non-burdensome experience), subscales Impact on health and Impact on schedule were approximately normal
Misawa, Miyashita, Kawa, Abe, Abe, Nakayama; & Given (2009) Validity and reliability of the Japanese version of the Caregiver Reaction Assessment Scale (CRA-J) for community-dwelling cancer patients [274].	57	57.0 (13.0; range not reported)	77.0	Cross-sectional	Self-report	Advanced cancer	Mean care duration 22 months (38; range not reported)	72% had metastasis	Caregivers of community-dwelling advanced cancer patients	Japan	Japanese	Not reported	Not reported	Not reported	No items demonstrated a floor of ceiling effect

Scale	Study	N	Age (mean, SD, range)	%	Design	Method	Population	Duration	Severity	Sample description	Country	Language	Response rate	Scores	Missing values	Floor/ceiling effect
	Malhotra, Chan, Malhotra, & Ostbye (2012) Validity and reliability of the Caregiver Reaction Assessment scale among primary informal caregivers for older persons in Singapore [275].	1190	55.6 (SD not reported; 20-95)	60.2	Cross-sectional	Self-report	Elderly	Not reported	Not reported	Primary informal caregivers for older persons (≥75 years) receiving human assistance for at least one activity of daily living	Singapore	Chinese, English, Malay, Tamil	Not reported	Mean impact on schedule=2.7 (0.7); finance=2.8 (0.7); lack of family support=2.1 (0.6); health=2.4 (0.6); caregiver esteem=3.9 (0.6). Modified Caregiver Reaction Assessment: impact on schedule and health=2.6 (0.6); finances=2.7 (0.9); lack of family support=2.1 (0.6); caregiver esteem=3.8 (0.6)	No missing values	Not reported
	Stephan, Mayer, Renom Guiteras, & Meyer (2013) Validity, reliability, and feasibility of the German version of the Caregiver Reaction Assessment scale (G-CRA): A validation study [276].	234	61.1 (11.8; range not reported)	69.7	Cross-sectional (part of longitudinal study)	Interview based	Dementia	Not reported	66.3% high dependency in ADL; 27.8% moderate; 5.6% low	Informal caregivers of persons with dementia recently admitted to institutional long-term care or living at home	Germany	German	Not reported	Mean caregiver esteem=26.1 (4.4; 11-35); Impact in finance=7.1 (2.9; 3-15); lack of family support=11.1 (4.4; 5-24); disrupted schedule=14.0 (4.7; 5-25); impact on health=9.8 (3.5; 4-19)	<5% of the subscales' sum-scores were missing, missing values were equally distributed throughout all items	Not reported
Caregiver Strain Index (1983) [19]	Robinson (1983) Validation of a Caregiver Strain Index [19].	81	Mean and SD not reported (range: 22-83)	Not reported	Cross-sectional	Interview based	Hip surgery and heart problems	Period hospitalized plus 2 months	Not reported	Spouses, family, friends and neighbours who provided varying degrees of care to recently hospitalized hip surgery and heart patients aged 65 and over	United States of America	English	97.8	3.5 (3.5; 0-12)	4 cases were excluded based on missing items	30.9% had the minimal score (0), 0% had the maximum score (13)
	McGartland Rubio, Berg-Weger, & Tebb (1999) Assessing the validity and reliability of well-being and stress in family caregivers [277].	143	61.0 (SD and range not reported)	83.0	Cross-sectional	Self-report	Not reported	Not reported	Not reported	Caregivers from support groups	United States of America	English	Not reported	Not reported	Not reported	Not reported
	Chen, & Hu (2002) The generalizability of Caregiver Strain Index in family caregivers of cancer patients [278].	14	44.7 (SD not reported; 25-70)	71.4	Repeated administration (repeated measures after 1, 2 and 4 weeks)	Interview based	Cancer	Mean duration 2.3 years (SD/range not reported)	Metastasis and hospitalized	Family caregivers of hospitalized cancer patients	Taiwan	Taiwanese	Not reported	T1-T4: 0.48 (0.04); 0.34 (0.04); 0.33 (0.03); 0.33 (0.05) (mean instead of sum scores), ranges not reported	T2: 21 missing total scores, T3: 16, T4: 13 . 14 caregivers had complete data on al four occasions	Not reported
	Van Exel, Scholte Op Reimer, Brouwer, Van den Berg, Koopmanschap, & Van den Bos (2004) Instruments for assessing the burden of informal caregiving for stroke patients in clinical practice: A comparison of CSI, CRA, SCQ and self-rated burden [271].	148	60.6 (13.7; range not reported)	69.6	Cross-sectional	Self-report	Stroke	Mean duration 0.5 years (SD/range not reported)	59.5% mild/severe disability; 40.5% were independent	Primary informal caregivers of stroke patients	The Netherlands	Dutch	35.9	4.3 (3.1; 0-10)	78.4% scores no missing values; 88.5% <10%; range missing values per item: 4.1-8.8%. standardized index missing values: 6.7	No indication for floor or ceiling effect
	Post, Festen, Van de Port, & Visser-Meily (2007) Reproducibility of the Caregiver Strain Index and the Caregiver Reaction Assessment in partners of stroke patients living in the Dutch community [272].	26	55.9 (10.3; range not reported)	57.7	Repeated administration (repeated measure after 1 week)	Self-report	Stroke	Mean duration 3.0 years (0.1; range not reported)	One-sided supratentorial lesion, no disabling comorbidity, mostly ischaemic stroke (65.4%)	Partners of stroke patients living in the Dutch community	The Netherlands	Dutch	Not reported	T1: 5.6 (3.4); T2: 5.6 (3.9); ranges not reported	Not reported	Not reported

Instrument	Study	N	Age	%	Design	Method	Condition	Care duration	Dependency	Sample	Country	Language				
	Ugur, & Fadiloglu (2010) "Caregiver strain index" validity and reliability in Turkish society [279].	132	Age categories: 18-30 (3.0%); 31-42 (20.5%); 43-55 (37.9%); 56-68 (33.3%); 69 and above (5.3%)	45.5	Repeated administration (repeated measure after 2-3 weeks)	Interview based	Cancer	Care duration: <6 months (34.8%); 6-12 months (26.5%); 13-24 months (18.9%); 25-37 months (7.6%) and ≥38 months (12.1%)	Not reported	Family members providing care for cancer patients treated at the chemotherapy unit of a university hospital in Izmir	Turkey	Turkish	Not reported	Not reported	Not reported	Not reported
	Othman, & Siongteck (2014) Validation of Malay Caregiver Strain Index [280].	50	42.7 (SD not reported; 21-68)	64.0	Cross-sectional	Self-report	Stroke	Not reported	Not reported	Caregivers of post-stroke patients attending the medical clinic	Malaysia	Malay	Not reported	Not reported	No missings at all	Not reported
	García-Domíngue, Martínez-Ginés, Carmona, Caminero, Prefasi, Maurino, et al. (2019) Measuring burden in caregivers of people with multiple sclerosis: Psychometric properties of the CSI questionnaire [281].	72	44.9 (11.7; range not reported)	59.7	Cross-sectional	Self-report	Multiple sclerosis	Not reported	73.6% mild	Caregivers of patients with multiple scerosis	Spain	Spanish	Not reported	3.9 (3.4; range not reported)	Not reported	Not reported
Modified Caregiver Strain Index (2003) [69]	Thornton, & Travis (2003) Analysis of the reliability of the Modified Caregiver Strain Index [77].	158	61.0 (SD not reported; 18-86)	Not reported	Repeated administration (repeated measure after 2 weeks)	Interview based	Patients who took medications and receive informal care	Mean care duration 62.5 months (3-360 months; SD not reported)	Mean mental functioning of 3.8 on a scale ranging 0-8 (higher score indicates more impairment); mean physical functioning was 11 on a scale ranging 0-32 (higher score indicates more impairment)	Caregivers of family members or friends aged ≥53 who took medications on regular basis and receive informal care.	United States of America	English	18.0	Not reported	98.0% completion rate	Not reported
	Chan, Chan, & Suen (2013) Validation of the Chinese version of the Modified Caregivers Strain Index among Hong Kong caregivers: An initiative of medical social workers [282].	219	54.7 (SD not reported; 18-87)	71.7	Cross-sectional	Self-report	Chronic illness (stroke, renal, palliative care, hip fracture)	Care duration at least 2 months	Not specified	Primary caregivers of patients with various chronic illnesses in a home stetting	Hong Kong	Chinese	Not reported	13.1 (6.9; range not reported)	98.3% completion rate	Not reported
Carer Assessment Scale (1998) [205]	Mackenzie, Holroyd, & Lui (1998) Community nurses' assessment of the needs of Hong Kong family carers who are looking after stroke patients [234].	14	Not reported	Not reported	Cross-sectional	Self-report	Stroke	Not reported	Not reported	Main carer of a person who had suffered a stroke	Hong Kong	Chinese and English	Not reported	Not reported	Not reported	Not reported
Carer's Assessment of Managing Index (1998) [206]	McKee, Spazzafumo, Nolan, Wojszel, Lamura, & Bien (2009) Components of the difficulties, satisfactions and management strategies of carers of older people: A principal component analysis of CADI-CASI-CAMI [283].	295 (35.9% United Kingdom, 29.8% Italy and 34.3% Poland)	United Kingdom: 62% ≥65; Italy: 35%; Poland: 18%	United Kingdom: 78.0; Italy: 79.0; Poland: 18.0	Cross-sectional	Self-report	Elderly	Care duration United Kingdom sample: 12% <2 years; 37% 2-5 years; 51% more than 5 years; Italian sample respectively 22%; 26% and 52%; Polish sample respectively 9%; 53% and 38%	Dependency level cared person United Kingdom sample: 56% severe, 32% moderate and 12% slights/independent; Italian sample respectively 75%, 21% and 5%; Polish sample respectively 42%, 43% and 15%	Carers of older people	United Kingdom, Italy and Poland	English, Italian and Polish	78.0	Means and SD's were reported for 7 of the ten sub dimensions	Not reported	Not reported

Scale	Reference	N	Mean age (SD; range)		Study design	Data collection	Diagnosis	Care duration	Dependency/cognitive status	Caregiver	Country	Language		Score (SD; range)	Missing data	Distribution
Modified Pearlin Burden Scale (1990) [149]	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
Oberst Caregiving Burden Scale (1991) [142]	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
Positive Aspects of Caregiving Questionnaire (2004) [211]	Tarlow, Wisniewski, Belle, Rubert, Ory, & Gallagher-Thompson (2004) Positive Aspects of Caregiving: Contributions of the REACH Project to the development of new measures for Alzheimer's caregiving [241].	1229	62.6 (0.4; range not reported)	81.4	Cross-sectional (part of a longitudinal study)	Self-report	Probable Alzheimer's disease or cognitive impairment	Not reported	Average score on the Mini-Mental State Exam is 12.6 (0.2) (scores between 10-19 considered to be moderately impaired)	Caregivers who lived with the care recipients, provided care for at least six months	United States of America	English and Spanish	Not reported	Not reported	More than 99% of the cases were complete on main scale	Not reported
	Las Hayas, Lopez, & Calvete (2014) Positive Aspects of Caregiving in Spanish caregivers of individuals with acquired brain injury [284].	141	58.3 (12.2; 27-85)	78.0	Cross-sectional	Self-report and interview based	Acquired brain injury	Mean years receiving care 11.1 (7.5; range not reported)	41.2% totally dependent; 37.0% severely dependent; 12.6% moderately dependent; 4.2% mildly dependent; 5.0% independent	Family members who are the primary caregivers of relatives with acquired brain injury	Spain	Spanish	27.8	Not reported	1 case excluded due to missing scores	Not reported
	Lou, Lau, & Cheung (2015) Positive Aspects of Caregiving (PAC): Scale validation among Chinese dementia caregivers (CG) [285].	374	62.9 (12.4; 23-89)	76.7	Cross-sectional (part of intervention study)	Interview based	Dementia	Not reported	Not reported	Caregivers of patients with dementia	Hong Kong	Chinese	Not reported	27.8 (9.5; range not reported); based on scale with 11 items	82 cases excluded due to missing scores: 36 had one missing, 14 had two missings, 10 had three missings and 22 >3 missings	Total score was negatively skewed
Relative Stress Scale (1982) [223]	Greene, Smith, Gardiner, & Timbury (1982) Measuring behavioural disturbance of elderly demented patients in the community and its effects on relatives: A factor analytic study [84].	38	Not reported	Not reported	Repeated administration (repeated measure after 3 weeks)	Self-report	Senile dementia	Not reported	Not reported	Relatives of elderly demented patients in the community	Scotland	English	Not reported	27.2 (10.3; range not reported)	Not reported	Not reported
	Ulstein, Bruun, & Engedal (2007) The Relative Stress Scale, a useful instrument to identify various aspects of carer burden in dementia? [286].	196	63.8 (13.0; range not reported)	65.0	Cross-sectional	Self-report	Dementia	Mean duration 3.8 years (SD/range not reported)	Mean Mini Mental State Examination score of 20.6 (5.3)	Carer of persons with dementia living at home	Norway	Norwegian	Not reported	23.8 (11.3; range not reported)	Not reported	Not reported
Revised 15-item Bakas Caregiving Outcomes Scale (2006) [171]	Bakas, Champion, Perkins, Farran, & Williams (2006) Psychometric testing of the Revised 15-item Bakas Caregiving Outcomes Scale [197].	147	51.6 (SD not reported; 21-78)	78.9	Repeated administration (repeated measure after 2 weeks)	Self-report and interview based	Stroke	4 months (SD/range not reported)	Not reported	Family caregivers of stroke survivors	United States of America	English	Not reported	58.4 (10.8; range not reported)	No more than 3 items were missing for any caregiver, 1 caregiver had 3 missings, 3 had 2 missings, 4 had 1 missing	The highest ceiling effect was 10.7% (relationship with survivor) and the highest floor effect was 14.1% (time for social activities)
	Govina, Kotronoulas, Mystakidou, Giannakopoulou, Galanos, & Patiraki (2013) Validation of the revised Bakas Caregiving Outcomes Scale in Greek caregivers of patients with advanced cancer receiving palliative radiotherapy [287].	100	52.9 (13.5; 26-88)	76.0	Repeated administration (repeated measure after 2 weeks)	Interview based	Advanced cancer	Mean duration 30.3 months (49.5; 1-276); mean care duration 14.4 months (19.6; 1-144)	Not reported	Greek caregivers of patients with advanced cancer receiving palliative radiotherapy	Greece	Greek	74.0	50.0 (10.5; 24-77)	Less than 5% of the data of the dependent variable for the total sample	Not reported

Instrument	Reference	N	Age mean	%	Design	Method	Condition	Duration	Severity	Population	Country	Language	%	Score mean (SD; range)	Missing values	Distribution notes
Sense of Competence Questionnaire (1993) [196]	Vernooij-dassen (1993) Dementie en thuiszorg: Een onderzoek naar determinanten van het competentiegevoel van centrale verzorgers en het effect van professionele interventie [Dementia and home care: Determinants of the sense of competence of primary caregivers ...] [288].	140	Not reported	Not reported	Cross-sectional	Interview based	Dementia	Not reported	Not reported	Home caregivers of dementia patients (>55 years)	The Netherlands	Dutch	Not reported	Not reported	Not reported	Not reported
	Scholte op Reimer, De Haan, Pijnenborg, Limburg, & Van den Bos (1998) Assessment of burden in partners of stroke patients with the Sense of Competence Questionnaire [289].	Sample A:166; sample B (used for test-retest reliability): 47	A: 66.0 (SD not reported; 29-92); B: 60.0 (SD not reported; 23-85)	A: 77.1; B: 61.7	A: cross-sectional (part of longitudinal study); B: repeated administration (repeated measure after 1 week)	Self-report and interview based	Stroke	Sample A and B: 6 months (SD/range not reported)	Not reported	Partners of noninstitutionalized patients who had been hospitalized because of stroke	The Netherlands	Dutch	74.0	A: 43.5 (10.6; 27-84); B: 42.0 (13.9; 27-77)	8 of the 174 partners (sample A) were excluded due to >10% missings, none in sample B; respectively 0-4.2% (A) and 0-1.3% (B) missing values per item; standardized index of missing values: 1.7 (A) and 0.1 (B)	Total scores were somewhat positively skewed (low estimates and fairly small standard deviations)
	Van Exel, Scholte Op Reimer, Brouwer, Van den Berg, Koopmanschap, & Van den Bos (2004) Instruments for assessing the burden of informal caregiving for stroke patients in clinical practice: A comparison of CSI, CRA, SCO and self-rated burden [271].	148	60.6 (13.7; range not reported)	69.6	Cross-sectional	Self-report	Stroke	Not reported	59.5% mild/severe disability; 40.5% were independent	Primary informal caregivers of stroke patients	The Netherlands	Dutch	35.9	57.4 (8.9; 39-79)	71.6% of the scores had no missing values; 85.1% <10% missing values; % missing values per item: 4.7-17.6%; standardized index of missing values: 8.6	Indication for a somewhat positively skewed distribution
	Jansen, Van Hout, Van Marwijk, Nijpels, Gundy, Vernooij-Dassen, De Vet, Schellevis, & Stalman (2007) Sense of Competence Questionnaire among informal caregivers of older adults with dementia symptoms: A psychometric evaluation [290].	93	62.9 (14.4; 32.5-91.2)	71.0	Cross-sectional (part of trial study)	Self-report	Dementia symptoms	Median months with symptoms (25th percentile, 75th percentile)=26.0 (19-48), SD not reported	Severity of neuropsychiatric symptoms based on the Neuropsychiatric Inventory-Questionnaire, mean: 6.3 (5.6; 0-27)	Caregivers of community dwelling older adults with dementia symptoms who received no assistance from outpatient geriatric services or outpatient diagnostic services	The Netherlands	Dutch	93.9	107.7 (13.7; 65.9-132.0)	% missing per item:0-3%; no missings on 18 items	Subscale 'satisfaction with the care recipient' demonstrated a ceiling effect (18% had a maximum score)
	Pendergrass, Beische, Becker, Hautzinger, & Pfeiffer (2015) An abbreviated German version of the Sense of Competence Questionnaire among informal caregivers of relatives who had a stroke: Development and validation [291].	198 (sum of two samples: A=122 and B=76)	64.93 (9.71; 27-87)	74.2	A: cross-sectional (part of randomized controlled trial); B: cross-sectional	interview based	Stroke	Not reported	Not reported	A: caregivers of stroke patients who participated in a randomized controlled trial of a telephone-based problem-solving intervention, and B: informal caregivers of stroke patients who were members of self help groups	Germany	German	Not reported	95.3 (16.4; range not reported)	Not reported	A ceiling effect occurred in the first subscale (satisfaction with the care recipient): 16,2% had the maximum score
Zarit Burden Interview (1980) [80]	Zarit, Reever, & Bach-Peterson (1980) Relatives of the impaired elderly: Correlates of feelings of burden [89].	29	65.0 (SD not reported; 42-82)	86.2	Cross-sectional	Self-report	Senile dementia	Mean duration 3.1 years (0.5-10 years; SD not reported)	Major deficits	Primary caregivers of elderly persons with senile dementia	United States of America	English	Not reported	30.8 (13.8; 1-66); possible scores ranged 0-84 instead of the regular 0-88	Not reported	Not reported
	Arai, Kudo, Hosokawa, Washio, Miura, & Hisamichi (1997) Reliability and validity of the Japanese version of the Zarit Caregiver Burden interview [292].	66	62.0 (12.4; range not reported)	Not reported	Repeated administration (repeated measure after 2 weeks)	Self-report	Elderly	Mean care duration 74.2 months (63.3; range not reported)	N.a.	Carers of elderly	Japan (northern)	Japanese	85.4	38.7 (18.4; range not reported)	80.5% completion rate	Not reported

Study	N	Age	% female	Design	Assessment	Patient condition	Care duration	Impairment	Caregiver description	Country	Language	%	Burden score	Response	Other
Arai, & Washio (1999) Burden felt by family caring for the elderly members needing care in southern Japan.[293]	45	60.4 (14.7; range not reported)	77.8	Cross-sectional	Self-report	Elderly in need of care	Mean care duration 65.8 months (52.6; range not reported)	More than two behavioural disturbances and/or partial limitations in ADL	Carers of elderly	Japan (southern)	Japanese	97.8	42.4 (14.9; range not reported)	Not reported	Not reported
Knight, Fox, & Chou (2000) Factor structure of the Burden Interview [294].	Two separate samples: 220 and 108	Sample 1: 60.1 (13.4); sample 2: 65.5 (11.92); ranges not reported	71.0 (both samples)	Cross-sectional	Interview based	Dementia	Not reported	Not reported	Primary caregivers of demented persons	United States of America	English	Not reported	Not reported	Not reported	Not reported
Taub, Andreoli, & Bertolucci (2004) Dementia caregiver burden: Reliability of the Brazilian version of the Zarit Caregiver Burden Interview [295].	55	56.0 (SD not reported; 23-81)	82.0	Repeated administration (repeated measure after 3-6 days)	Not described	Dementia	Not reported	98% memory deficits; 100% orientation deficits; 96% progressive aggravation of symptoms	Primary informal caregivers of demented patients coming from 3 different health care centres	Brazil	Brazilian	Not reported	Not reported	Not reported	Not reported
Ankri, Andrieu, Beaufils, Grand, & Henrard (2005) Beyond the global score of the Zarit Burden Interview: Useful dimensions for clinicians [296].	152	80.9 (7.0: range not reported)	69.3	Cross-sectional	Self-report	Dementia	Not reported	43.4% mildly impaired; 49.6% moderately	Informal caregivers of patients suffering from Alzheimer's Disease or related disorders, in community dwelling	France	French	Not reported	32.9 (17.9; 1-82)	Not reported	Not reported
Schreiner, Morimoto, Arai, & Zarit (2006) Assessing family caregiver's mental health using a statistically derived cut-off score for the Zarit Burden Interview [297].	Three separate samples: 80 (stroke); 48 (Chronic Obstructive Pulmonary Disease) and 70 (general disabilities)	Stroke: 60.2 (14.0; range not reported); Chronic Obstructive Pulmonary Disease: 65.4 (12.8; range not reported); general disabilities: 60.2 (12.7; range not reported)	Stroke: 71.3; Chronic Obstructive Pulmonary Disease: 91.1; general disabilities: 77.1	Cross-sectional	Self-report and interview based	Stroke, chronic obstructive pulmonary disease and general disabilities	Not reported	Not reported	Co-resident family members who assisted with most of the patient's daily care needs	Japan	Japanese	Not reported	Stroke: 28.3 (12.7; range not reported); chronic obstructive pulmonary disease: 20.4 (13.0); general disabilities: 30.3 (17.7)	Not reported	Not reported
Lai (2007) Validation of the Zarit Burden Interview for Chinese Canadian caregivers [298].	339	Age categories: 18-34 (19.2%); 35-54 (54.7%); 55 and above (26.1%)	65.5	Cross-sectional	Interview based	Not specified	Not reported	Not reported	Chinese caregivers in Canada	Canada	Chinese	85.6	Not reported	Not reported	Not reported
Ko, Yip, Liu, & Huang (2008) Chinese version of the Zarit caregiver Burden Interview: A validation study [299].	168	54.0 (13.0; range not reported)	78.6	Repeated administration (repeated measure after 2 weeks)	Self-report	Dementia	Not reported	57.1% mild dementia; 33.9% moderate; 8.9% severe	Caregivers of patients with dementia	Taiwan	Chinese	Not reported	35.5 (14.4; 4-78)	92.8% completion rate	Not reported
Lu, Wang, Yang, & Feng (2009) Zarit Caregiver Burden Interview: Development, reliability and validity of the Chinese version [300].	523	44.2 (12.4; 16-79)	59.3	Cross-sectional	Interview based	Cancer, cardiovascular disease, neurological disease, orthopaedic condition	Median duration 3 months (SD/range not reported)	Not specified	Caregivers of hospital patients	China	Chinese	98.3	Not reported	Not reported	Not reported
Braun, Scholz, Hornung, & Martin (2010) The burden of spousal caregiving: A preliminary psychometric evaluation of the German version of the Zarit Burden Interview [301].	28	68.0 (8.6; 53-92)	100	Cross-sectional	Self-report	Dementia	Mean duration 38.9 (21.8; 1-85); unclear time frame	40.7% mild; 37.1% moderate; 22.2% severe	Community dwelling older couples with husband suffering from dementia and the wife being the primary caregiver	Switzerland and Germany	German	Not reported	31.7 (15.5; range not reported)	Not reported	Not reported

Reference	N	Mean age		Design	Assessment	Patient condition	Care duration	Severity	Caregiver description	Country	Language	%	Burden score	Missing data	Notes
Martin-Carrasco, Otermin, Perez-Camo, Pujol, Aguera, Martin, Gobartt, Pons, & Balana (2010) EDUCA study: Psychometric properties of the Spanish version of the Zarit Caregiver Burden Scale [302].	104	62.8 (13.6; range not reported)	68.7	Repeated administration (time frame repeated measure not reported)	Self-report	Alzheimer's disease	Mean duration 3.1 years (2.0; range not reported)	On average moderate dementia	Caregivers of patients with Alzheimer's disease	Spain	Spanish	90.4	60.7 (15.8; range not reported)	Not reported	Not reported
Seng, Luo, Yee, Lim, Chionh, Goh, & Yap (2010) Validity and reliability of the Zarit Burden Interview in assessing caregiving burden [303].	238	50.1 (10.5; 22-84)	68.1	Repeated administration (repeated measure after 2 weeks)	Self-report	Dementia	Care duration: 12.7% ≤1 year; 33.8% >1-3 year; 53.6% >3 years)	Not reported	Caregivers of patients with dementia	Singapore	English	76.6	35.4 (15.2; 0-77)	8 of the 246 cases excluded due to missings	Not reported
Siegert, Jackson, Tennant, & Turner-Strokes (2010) Factor analysis and Rasch analysis of the Zarit Burden Interview for acquired brain injury carer research [304].	222	54.0 (10.9; range not reported)	80.6	Cross-sectional	Self-report	Acquired brain injury	Not reported	Not reported	Carers of adults with acquired brain injury	United Kingdom	English	Not reported	Not reported	Not reported	Not reported
Chattat, Cortesi, Izzicupo, Del Re, Sgarbi, Fabbo, & Bergonzini (2011) The Italian version of the Zarit Burden Interview: A validation study [305].	273	55.5 (11.6; range not reported)	72.0	Cross-sectional	Interview based	Dementia	Not reported	Not reported	Caregivers related to consecutively attending outpatients with a previously established primary diagnosis of dementia	Italy	Italian	Not reported	33.4 (15.9; range not reported)	Not reported	Not reported
Cheah, Han, Chong, Anthony, & Lim (2012) Multidimensionality of the Zarit Burden Interview across the severity spectrum of cognitive impairment: An Asian perspective [306].	130	53.9 (13.0; range not reported)	71.5	Cross-sectional	Interview based	Dementia	Not reported	90% had very mild to moderate dementia	Caregivers of persons with cognitive impairment in a predominantly Chinese multi-ethnic Asian population	Singapore	English and Chinese	62.5	18.9 (16.5; 0-71)	Not reported	Not reported
Özer, Yurttas, & Akyil (2012) Psychometric evaluation of the Turkish version of the Zarit Burden Interview in family caregivers of inpatients in medical and surgical clinics [307].	223	35.3 (SD not reported; 17-76)	54.7	Repeated administration (time frame repeated measure not reported)	Not described	Neurological disease, orthopaedic disease or traumatology	Mean care duration 22.1 days (7-540; range not reported)	Not reported	Family caregivers of inpatients in medical and surgical clinics	Turkey	Turkish	Not reported	Not reported	Not reported	Not reported
Torres, Hoff, Padovani, & Ramos-Cerqueira (2012) Dimensional analysis of burden in family caregivers of patients with obsessive-compulsive disorder [308].	47	46.0 (13.0; range not reported)	51.1	Cross-sectional	Self-report	Obsessive-compulsive disorder	Not reported	Mean of 20 on a severity scale (ranged 0-40)	Adult family caregivers of patients with obsessive-compulsive disorder in treatment, outpatient clinic	Brazil	Brazilian	Not reported	24.0 (SD not reported; 2-72)	3 of the 50 cases excluded based on missings	Not reported
Cheng, Kwok, & Lam (2014) Dimensionality of burden in Alzheimer caregivers: Confirmatory factor analysis and correlates of the Zarit Burden Interview [309].	395	56.6 (11.7; range not reported)	81.0	Cross-sectional	Interview based	Alzheimer's disease	Mean care duration 3.7 years (3.1; range not reported)	Not reported	Hong Kong Chinese Alzheimer caregivers	Hong Kong	Chinese	Not reported	Not reported	Not reported	Data positively skewed
Lim, Cheah, Ali, Han, Anthony, Chan, & Chong (2014) Worry about performance: A unique dimension of caregiver burden [310].	130	53.9 (13.0; range not reported)	71.5	Cross-sectional	Interview based	Dementia	Not reported	90% had very mild to moderate dementia	Caregivers of persons with cognitive impairment in a predominantly Chinese multi-ethnic Asian population	Singapore	English and Chinese	62.5	18.9 (16.5; 0-71)	Not reported	Not reported
Chan, Lam, & Chiu (2015) Validation of the Chinese version of the Zarit Burden Interview [311].	40	51.0 (SD and range not reported)	52.5	Cross-sectional	Interview based	Dementia	Not reported	Mild to moderate dementia	Carers of patients with mild-moderate dementia in psychiatric clinics in Hong Kong	Hong Kong	Chinese	Not reported	24.6 (SD and range not reported)	Not reported	Not reported
Galindo-Vazquez, Benjet, Cruz-Nieto, Rojas-Castillo, Riveros-Rosas, Meneses-Garcia, Aguilar-Ponce, Alvarez, Avitia, & AlvaradoAguilar (2015) Psychometric properties of the Zarit Burden Interview in Mexican caregivers of cancer patients [312].	359	43.4 (SD not reported; 17-79)	72.2	Cross-sectional	Self-report	Cancer	Not reported	Not reported	Mexican caregivers of cancer patients	Mexico	Spanish	Non-response <5%	Not reported	Not reported	Not reported

Study	N	Age mean (SD; range)	%	Study design	Method	Patient condition	Care duration	Disease severity	Caregivers	Country	Language	Response rate	ZBI score	Missing data	Floor/ceiling effects
Al-Rawashdeh, Lennie, & Chung (2016) Psychometrics of the Zarit Burden Interview in caregivers of patients with heart failure [313].	124	56.4 (14.4; range not reported)	76.6	Cross-sectional	Self-report	Chronic heart failure	Not reported	Not reported	Caregivers of patients with heart failure, outpatient clinics	United States of America	English	Not reported	Only stated that 41.1% of the cases had a burden score of ≥17	13.3% of the cases had missing data and were excluded	Not reported
Bianchi, Flesch, Alves, Batistoni, & Neri (2016) Zarit Burden Interview psychometric indicators applied in older people caregivers of other elderly [314].	121	70.5 (7.2; range not reported)	73.0	Cross-sectional	Self-report	Elderly	Care duration at least 6 months	Not reported	Caregivers aged ≥60, informally in charge of care to an elderly family member, home context, at least some degree of dependency	Brazil	Spanish	Not reported	26.1 (14.0; 3-80)	Not reported	Not reported
Goncalves-Pereira, Gonzalez-Fraile, Santos-Zorrozua, Martin-Carrasco, Fernandez-Catalina, Dominguez-Panchon, Munoz-Hermoso, & Ballesteros (2017) Assessment of the consequences of caregiving in psychosis: A psychometric comparison of the Zarit Burden Interview (ZBI) and the Involvement Evaluation Questionnaire (IEQ) [315].	223 (109 intervention, 114 control)	60.0 (11.0; range not reported)	76.0	Repeated administration (part of randomized controlled trail, 4 months between baseline and endpoint)	Not described	Schizophrenia and related disorders	Mean care duration 15.0 years (10.0; range not reported)	Not reported	Caregivers caring for relative with diagnosis of schizophrenia or schizoaffective disorder	Spain	Spanish and Portuguese	Not reported	32.3 (16.0; range not reported)	Not reported	Not reported
Hagell, Alvariza, Westergren, & Arestedt (2017) Assessment of burden among family caregivers of people with Parkinson's Disease using the Zarit Burden Interview [316].	66	69.6 (8.2; 44-86)	70.0	Cross-sectional	Self-report	Parkinson's disease	Median care duration 3 years (q1-q3: 2-5) and max 18 years	Not reported	Family caregivers of people with Parkinson's disease	Sweden	Swedish	61.0	28.3 (18.0; range not reported)	Missing item responses were ≤5%	Floor/ceiling effects were 2/0%
Imarhiagbe, Asemota, Oripelay, Akpekpe, Owolabi, Abidakun, Akemokwe, Ogundare, Azzez, & Osakue (2017) Burden of informal caregivers of stroke survivors: Validation of the Zarit burden interview in an African population [317].	64	40.7 (14.3; range not reported)	51.6	Cross-sectional	Self-report	Stroke	Median duration after discharge 330 days (Interquartile range=730-739)	89.1% cerebral infarct; 10.9% cerebral haemorrhage	Informal caregivers of outpatient stroke patients of a neurology clinic	Sub-Saharan Africa	French	Not reported	Mean and SD only reported for subgroups: caregivers who reported that caregiving was telling on their health versus not: respectively 30.1 (14.2) and 20.3 (13.0); ranges not reported	Not reported	Not reported
Landfeldt, Mayhew, Straub, Bushby, Lochmuller, & Lindgren (2017) Psychometric properties of the Zarit Caregiver Burden Interview administered to caregivers to patients with Duchenne muscular dystrophy: A Rasch analysis [318].	475	44.0 (SD not reported; 23-76)	81.0	Cross-sectional	Self-report	Duchenne muscular disease	Not reported	Not reported	Caregivers to patients with Duchenne muscular dystrophy	United Kingdom and United States of America	English	Not reported	Not reported	Not reported	Not reported
Tang, Yu, Liu, Lin, Chen, Zhao, Xiao (2017) Factor analyses of the Chinese Zarit Burden Interview among caregivers of patients with schizophrenia in a rural Chinese community [319].	324	57.8 (12.9; range not reported)	53.1	Cross-sectional	Interview based	Schizophrenia	Mean care duration 16.5 years (10.92; range not reported)	Not reported	Caregivers of patients with schizophrenia, rural Chinese community	China	Chinese	92.9	Not reported	3 cases excluded due to missing values	Not reported
Oh, & Kim (2018) Factor analysis of the Zarit Burden Interview in family caregivers of patients with amyotrophic lateral sclerosis [320].	202	51.0 (14.1; range not reported)	60.4	Cross-sectional	Self-report	Amyotrophic lateral sclerosis	Mean duration 45.9 months (98.7; range not reported)	19.8% stage II; 49.5% stage III; 30.7% stage IV	Family caregivers of patients with amyotrophic lateral sclerosis	South Korea	Korean	Not reported	43.1 (19.5; 5-83)	14 cases excluded due to missing in any item	Not reported
Smith, George, & Ferreira (2018) Factors emerging from the "Zarit Burden Interview" and predictive variables in a UK sample of caregivers for people with dementia [321].	110	70.0 (12.0; range not reported)	68.2	Cross-sectional	Self-report	Dementia	Not reported	Not reported	Unpaid community caregivers of patients with dementia	Scotland	English	17.2	41.6 (14.2; 14-69)	Not reported	Not reported

Measure	Study	N	Age mean (SD; range)	%	Design	Administration	Condition	Care duration	Cognitive status	Sample	Country	Language	Response rate	Score mean (SD; range)	Missing data	Floor/ceiling effect
	Nagata, Yada, & Inagaki (2018) Exploration of thefactor structure of the burden experienced by individuals providing end-of-life care at home [322].	247	64.8 (14.6; range not reported)	79.4	Cross-sectional	Self-report	Elderly	Not reported	Not reported	Caregivers who provide end-of-life care at home	Japan	Japanese	99.6 (effective)	Only reported per item	No missing values for any item	4 items had floor of ceiling effect in terms of +/- 1 SD
	Vatter, McDonald, Stanmore, Clare, & Leroi (2018) Multidimensional care burden in Parkinson-related dementia [323].	127	67.4 (7.6; range not reported)	85.3	Cross-sectional	Self-report	Parkinson-related dementia	Mean care duration 5.5 year (4.7; range not reported)	Not reported	Spouses and lofe partners	United Kingdom	English	Not reported	35.5 (15.4; range not reported)	Not reported	Not reported
	Yu, Liu, Zhou, Chen, Zhang, Hu, et al. (2018) Assessment of burden among family caregivers of schizophrenia: Psychometric testing for short-form Zarit Burden Interviews [324].	327	57.7 (12.5; 17-81)	53.8	Cross-sectional	Self-report	Schizophrenia	Median care duration 15 years (range 1-49)	Not reported	Family caregivers of patients with schizophrenia	China	Chinese	Not reported	Not reported	Not reported	Not reported
Zarit Burden Interview Short Form (2001) [112]	Bédard, Molloy, Squire, Dubois, Lever, & O'Donnell (2001) The Zarit Burden Interview: A new short version and screening version [126].	413	61.0 (13.9; range not reported)	Not reported	Repeated administration (6 months between baseline and second measurement)	Not described	Cognitively impaired elderly	Not reported	Not reported	Caregivers of community dwelling older adults with cognitively impaired older adults	Canada	English	Not reported	20.6 (15.6; range not reported); scores at baseline on 22-item scale	Not reported	Not reported
	O'Rourke, & Tuokko (2003) Psychometric properties of an abridged version of the Zarit Burden Interview within a representative Canadian caregiver sample [325].	770	58.4 (12.7; range not reported)	71.2	Repeated administration (5 years between measurements)	Interview based	Dementia	Not reported	Not reported	Informal caregivers of surviving institutionalized and community dwelling subjects	Canada	English	93.8	7.6 (7.8; range not reported)	Data were missing at both points of measurement in 5% of the cases	Not reported
	O'Rourke, & Tuokko (2003) The relative utility of four abridged versions of the Zarit Burden interview [326].	503	58.9 (13.4; range not reported)	67.4	Cross-sectional	Not described	Dementia	Not reported	Not reported	Caregivers of patients with dementia living in the community or residing in institutions	Canada	English	93.8	Not reported	Missing is estimated at roughly 5%	Not reported
	Bachner, & Ayalon (2010) Initial examination of the psychometric properties of the short Hebrew version of the Zarit Burden Interview [327].	148	58.8 (13.1; range not reported)	60.1	Cross-sectional	Interview based	Cognitive or physical impairment	Not reported	60.8% had cognitive impairment	Primary caregivers of individuals with cognitive and/or physical impairments	Israel	Hebrew	Not reported	Not reported	Not reported	Not reported
	Higginson, Gao, Jackson, Murray, & Harding (2010) Short-form Zarit Caregiver Burden Interviews were valid in advanced conditions [328].	452 (advanced cancer: 105; dementia: 131; acquired brain injury: 215)	Cancer: 66.0 (12.0); dementia: 62.0 (13.0); acquired brain injury: 54.0 (11.0); ranges not reported	Cancer: 72.0; dementia: 72.0; ABI: 81.0	Secondary analysis using data pooled from four different studies	Self-report	Advanced cancer, dementia and acquired brain injury	Not reported	Not reported	Caregivers of patients with advanced cancer, dementia of acquired brain injury	United Kingdom	English	Not reported	Cancer: 12.0 (8.5); dementia: 15.1 (10.0); ABI: 21.7 (10.1); ranges not reported	Not reported	Not reported
	Bachner, O'Rourke, Ayalon, & Bedard (2011) Comparison of caregiver responses to English and Hebrew language versions of an abridged Zarit Burden Interview [329].	341 (142 from Israel and 199 from Canada)	Not reported	Not reported	Cross-sectional (Canadian part is part of a longitudinal study)	Interview based	Cognitive deficits (Israel) and dementia (Canada)	Not reported	Not reported	Israel: caregivers of patients with cognitive deficits; Canada: caregivers of patients with dementia	Canada and Israel	English and Hebrew	Not reported	22.5% of the Israeli sample report significant burden; 23.6% of the Canadian caregivers	Not reported	Not reported
	Brink, Stones, & Smith (2012) Confirmatory factor analysis of the Burden Interview of the caregivers of terminally ill home care clients [330].	71	60.7 (1.4; range not reported)	70.0	Cross-sectional	Self-report	Terminally ill with an estimated prognosis of six months or less	Mean care duration 21.4 months (3.4; range not reported)	Terminally	Primary caregivers of terminally ill persons receiving palliative home care	Canada	English	65.0	19.9 (9.2; range not reported); possible scores ranged 0-60 instead of the regular 0-48	Not reported	Not reported
	Iecovich (2012) Psychometric properties of the Hebrew version of the Zarit Caregiver Burden Scale short version [331].	456 (data from two samples 121 and 335)	56.2 (11.9; range not reported)	60.6	Cross-sectional	Interview based	Physically frail elderly	Not reported	Not reported	Informal family caregivers of cognitively intact but physically frail older family members	Israel	Hebrew	60.5	27.4 (9.2; 12-58); possible range 12-60	Not reported	Not reported

Reference	N	Age M (SD; range)	%	Design	Method	Diagnosis	Onset	Severity	Sample	Country	Language	%	ZBI score M (SD; range)	Missing data	Floor/ceiling effects
Bachner (2013) Preliminary assessment of the psychometric properties of the abridged Arabic version of the Zarit Burden Interview among caregivers of cancer patients [332].	96	38.1 (15.5; range not reported)	61.5	Cross-sectional	Self-report	Cancer	Not reported	Not reported	Bedouin Arab primary caregivers of cancer patients	Israel	Arabic	Not reported	Not reported	Not reported	Not reported
Rajabi-Mashhadi, Mashhadinejad, Ebrahimzadeh, Golhasani-Keshtan, et al. (2015) The Zarit Caregiver Burden Interview Short Form (ZBI-12) in spouses of veterans with chronic spinal cord injury, validity and reliability of the Persian version [333].	72	44.7 (6.5; 31-66)	Not reported	Repeated administration (repeated measure after 3 days)	Self-report	Spinal cord injury	Onset in Iran-Iraq war (1980-1988)	Not reported	Spouses of veterans of the Iran-Iraq war (1980-1988) with chronic spinal cord injury	Iran	Persian	72.0	Not reported	Not reported	Floor/ceiling effects were considered as being present if more than 15% of the sample had a score between 0-4 and 44-48 respectively. Floor effect was found in item 2 and 11; ceiling effect in item 1, 3, 4, 5, 7, 8, 9 and 10; floor and ceiling effect in item 6
Stagg, & Larner (2015) Zarit Burden Interview: Pragmatic study in a dedicated cognitive function clinic [334].	45	Not reported	69.0	Cross-sectional	Self-report	Dementia, mild cognitive impairment and subjective memory complaint	Not reported	Not reported	Cohabiting spouses/partners of new patients	United Kingdom	English	Not reported	Dementia: 16.5 (10.6; range not reported); mild cognitive impairment: 17.5 (10.9); subjective memory complaint: 12.2 (8.3)	Not reported	Not reported
Branger, O'Connell, & Morgan (2016) Factor analysis of the 12-Item Zarit Burden Interview in caregivers of persons diagnosed with dementia [335].	194	61.9 (14.8; range not reported)	69.5	Cross-sectional	Self-report	Dementia	Not reported	Not reported	Informal rural and urban caregivers of persons diagnosed with dementia	Canada	English	Not reported	14.2 (8.7; 0-48)	3 cases excluded due to missing scores	Not reported
Tang, Ho, Luo, Wong, Lau, Lum, & Cheung (2016) Validating a Cantonese short version of the Zarit Burden Interview (CZBI-Short) for dementia caregivers [336].	447	63.6 (12.6; range not reported)	78.0	Cross-sectional (part of pretest-posttest study)	Interview based	Dementia	Not reported	Not reported	Chinese caregivers of patients with dementia	China	Cantonese	Not reported	19.1 (8.8; range not reported)	53 cases of the 500 excluded due to missing values	Not reported
Hagell, Alvariza, Westergren, & Arestedt (2017) Assessment of Burden Among Family Caregivers of People With Parkinson's Disease Using the Zarit Burden Interview [316].	66	69.6 (8.2; 44-86)	70.0	Cross-sectional	Self-report	Parkinson's disease	Median care duration 3 years (q1-q3: 2-5) and max 18 years	Not reported	Family caregivers of people with Parkinson's disease	Sweden	Swedish	61.0	15.3 (11.0; range not reported)	Missing item responses were ≤5%	Floor/ceiling effects were 5/0%
Yu, Yap, & Liew (2018) The optimal short version of the Zarit Burden Interview for dementia caregivers: Diagnostic utility and externally validated cut-offs [337].	394	53.0 (10.7; 22-83)	59.9	Cross-sectional	Self-report	Dementia	Mean care duration 6.8 years (6.7; 0.1-17.7)	15.7% mild; 41.4% moderate; 42.9% severe	Caregivers of patients with dementia recruited from dementia services	Singapore	Not clearly specified (86,6% had a Chinese ethnicity)	87.8	34.8 (16.8; 0-80); scores on the 22-item scale	Not reported	Not reported
Gratão, Brigola, Ottaviani, Luchesi, Souza, Rossetti, et al. (2019) Brief version of Zarit Burden Interview (ZBI) for burden assessment in older caregivers [338].	341	69.6 (7.1; range not reported)	76.8	Cross-sectional	Interview based	Elderly	Care duration in 75.2% <5 years	13.5% completely dependent	Older caregivers of community-dwelling older dependent individuals	Brazil	Not reported	Not reported	Not reported	Not reported	Not reported
Yu, Liu, Zhou, Chen, Zhang, Hu, et al. (2018) Assessment of burden among family caregivers of schizophrenia: Psychometric testing for short-form Zarit Burden Interviews [324].	327	57.7 (12.5; 17-81)	53.8	Cross-sectional	Self-report	Schizophrenia	Median care duration 15 years (range 1-49)	Not reported	Family caregivers of patients with schizophrenia	China	Chinese	Not reported	Not reported	Not reported	Not reported

M: mean; SD: Standard deviation; N.a.: Not applicable

Supplementary Table S7. Summary of measurement properties of the 18 selected measures per publication.

Measurement		Structural validity		Internal consistency		Cross-cultural validity/measurement invariance		Reliability		Measurement error		Criterion validity		Hypotheses testing		Responsiveness	
		Methodological quality	Result (rating)	Methodological quality	Result (rating)	Methodological quality	Result (rating)	Methodological quality	Result (rating)	Methodological quality	Result (rating)	Methodological quality	Result (rating)	Methodological quality	Result (rating)	Methodological quality	Result (rating)
Appraisal of Caregiving Scale (1991) [215]	Oberst, Thomas, Gass, & Ward (1989) Caregiving demands and appraisal of stress among family caregivers [256].	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
	Lambert, Yoon, Ellis, & Northouse (2015) Measuring appraisal during advanced cancer: Psychometric testing of the Appraisal of Caregiving Scale [257].	A	3-dimensional, no goodness of fit indicators reported (?)	V	α general stress=0.79; threat=0.89; benefit=0.72 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	V	Correlation with gold standard <0.7; hypotheses regarding predictive regression partly confirmed and partly rejected (-)	A	Results in line with 7 hypo's (7+), not in line with 5 hypo's (5-)	N.a.	N.a.
Burden Assessment Scale (1994) [153]	Reinhard, Gubman, Horwitz, & Minsky (1994) Burden Assessment Scale for families of the seriously mentally ill [174].	D	5-dimensional, not identical in two different samples (-)	I	α total scale in two different samples=0.89 and 0.91 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	A	Results in line with 1 hypo (1+)	A	Results in line with 1 hypo (1+)
	Ivarsson, Sidenvall, & Carlsson (2004) The factor structure of the Burden Assessment Scale and the perceived burden of caregivers for individuals with severe mental disorders [258].	A	3-dimensional, no goodness of fit indicators reported (?)	V	α total scale=0.90; sub dimensions=0.73; 0.75; 0.88 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
	Guada, Land, & Han (2011) An exploratory factor analysis of the Burden Assessment Scale with a sample of African-American families [259].	D	2-dimensional, no goodness of fit indicators reported (?)	I	α total scale=0.94 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
	Murdoch, Rahman, Barsky, Maunula, & Cawthorpe (2014) The use of the Burden Assessment Scale with families of a paediatric population [260].	A	4-dimensional, no goodness of fit indicators reported (?)	I	α total scale=0.89 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
	Hunger, Krause, Hilzinger, Ditzen, & Schweitzer (2016) When significant others suffer: German Validation of the Burden Assessment Scale (BAS) [261].	V	4-dimensional (c2(146)=287.08; p<0.001; RMSEA=0.07; CFI=0.92) (-)	V	α total scale=0.92; sub dimensions=0.64-0.90 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	V	Results in line with 3 hypo's (3+)	N.a.	N.a.
Burden Assessment Schedule Modified (2010) [156]	Das, Hazra, Ray, Ghosal, Banerjee, Roy, Chaudhuri, Raut, & Das (2010) Burden among stroke caregivers: Results of a community-based study from Kolkata, India [177].	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	I	ICC=0.80-0.84 (inter-rater and intra-rater. for every item) (+)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
Caregiver Burden Inventory (1989) [133]	Novak, & Guest (1989) Application of a multidimensional Caregiver Burden Inventory [150].	A	4-dimensional, no goodness of fit indicators reported (?)	V	α subscales ranged 0.73-0.86 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
	Marvardi, Mattioli, Spazzafumo, Martriforti, et al. (2005) The Caregiver Burden Inventory in evaluating the burden of caregivers of elderly demented patients: Results from a multicentre study [262].	A	4-dimensional, no goodness of fit indicators reported (?)	A	α total scale and sub dimensions >0.80 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.

Scale	Study																
	McCleery, Addington, & Addington (2007) Family assessment in early psychosis [263].	A	4-dimensional, no goodness of fit indicators reported (?)	A	α total scale ranged 0.90-0.94 (different time points); sub dimensions 0.72-0.88 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	V	Results in line with 2 hypo's (+2)	N.a.	N.a.
	Valer, Aires, Fengler, & Paskulin (2015) Adaptation and validation of the Caregiver Burden Inventory for use with caregivers of elderly individuals [264].	A	5-dimensional (RMSEA=0.065; CFI=0.935) (-)	V	α total scale=0.94; sub dimensions ranged 0.76-0.92 (?)	N.a.	N.a.	A	ICC total scale=0.94; sub dimensions ICC ranged 0.87-0.93 (+)	N.a.	N.a.	N.a.	N.a.	D	Results in line with 1 hypo (+1)	N.a.	N.a.
	Greco, Pancani, Sala, Annoni, Steca, Paturzo, D'Agostino, Alvaro, & Vellone (2017) Psychometric characteristics of the Caregiver Burden Inventory in caregivers of adults with heart failure [265].	V	5-dimensional (c2(242)=513.29; RMSEA=0.05; CFI=0.95; SRMR=0.07) (+)	V	α total scale=0.96; sub dimensions ranged 0.88-0.93 (+)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	V	Results in line with 2 hypo's (+2)	N.a.	N.a.
	Farmer, Thienemann, Leibold, Kamalani, Sauls, & Frankovich (2018) Psychometric evaluation of the Caregiver Burden Inventory in children and adolescents with PANS [266].	D	5-dimensional (c2(247)=341.05; p=0.001; TLI=0.981; RMSEA=0.061; CFI=0.983; SRMR=0.078) (+)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	A	Results in line with 3 hypo's (+3), not in line with 5 hypo's (-5)	N.a.	N.a.
	Vázquez, Otero, Simón, Bueno, & Blanco (2019) Psychometric properties of the Spanish version of the caregiver burden inventory [267].	V	5-dimensional (RMSEA=0.098; CFI=0.89) (-)	V	α total scale=0.89; sub dimensions ranged 0.74-0.83(+)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	V	AUC=0.81 for the total scale; sub dimensions ranged 0.67-0.78 (+)	D	Results in line with 2 hypo's (+2)	N.a.	N.a.
Caregiver Burden Scale (1996) [115]	Elmstahl, Malmberg, & Annerstedt (1996) Caregiver's burden of patients 3 years after stroke assessed by a novel Caregiver Burden Scale [130].	A	5-dimensional, no goodness of fit indicators reported (?)	V	α sub dimensions ranged 0.70-0.87; excepted for the environment sub dimension=0.53 (?)	N.a.	N.a.	A	Cohen's Kappa ranged 0.69-1.00 (+)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
	Cil Akinci, & Pinar (2014) Validity and reliability of Turkish Caregiver Burden Scale among family caregivers of haemodialysis patients [268].	V	5-dimensional (RMSEA=0.07; CFI=0.96; SRMR=0.07) (+)	V	α total scale=0.91; sub dimensions ranged 0.69-0.83; excepted for emotional involvement (0.61) and environment (0.63) (±)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	V	Results in line with 4 hypo's (4+), not in line with 1 hypo (1-)	N.a.	N.a.
	Farajzadeh, Akbarfahimi, Maroufizadeh, Rostami, & Kohan (2018) Psychometric properties of Persian version of the Caregiver Burden Scale in Iranian caregivers of patients with spinal cord injury [269].	A	5-dimensional (c2(188)=222.2; RMSEA=0.04; CFI=0.96; SRMR=0.05) (+)	V	α sub dimensions ranged 0.70-0.76; excepted for the environment sub dimension (0.56) (±)	N.a.	N.a.	A	ICC ranged 0.75-0.90 (+)	N.a.	N.a.	N.a.	N.a.	V	Results in line with 2 hypo's (+2)	N.a.	N.a.

Instrument	Reference																
Caregiver Reaction Assessment (1992) [182]	Given, Given, Stommel, Collins, King, & Franklin (1992) The Caregiver Reaction Assessment (CRA) for Caregivers to persons with chronic physical and mental impairments [210].	V	5-dimensional (CFI=0.98; CFI=1.00) (+)	V	α subscales=0.80-0.90 (+)	V	No important differences between group factors (disease and relationship) and over time (+)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	A	Results in line with 10 hypo's (10+)	N.a.	N.a.
	Nijboer, Triemstra, Tempelaar, Sanderman, Van den Bos (1999) Measuring both negative and positive reactions to giving care to cancer patients: Psychometric qualities of the Caregiver Reaction Assessment (CRA) [270].	A	5-dimensional, no goodness of fit indicators reported (?)	V	α subscales=0.62-0.83 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	A	Results in line with 4 hypo's (4+), not in line with 1 hypo (1-)	N.a.	N.a.
	Van Exel, Scholte Op Reimer, Brouwer, Van den Berg, Koopmanschap, & Van den Bos (2004) Instruments for assessing the burden of informal caregiving for stroke patients in clinical practice: A comparison of CSI, CRA, SCO and self-rated burden [271].	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	V	Results in line with 12 hypo's (12+), results not in line with 2 hypo's (2-)	N.a.	N.a.
	Post, Festen, Van de Port, & Visser-Meily (2007) Reproducibility of the Caregiver Strain Index and the Caregiver Reaction Assessment in partners of stroke patients living in the Dutch community [272].	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	A	ICC ranges 0.58-0.86 (two below 0.70) (3+/2-)	A	SDD (range: 0,8-1,0) for all subscales 1 SD or more, indicating that large score differences are needed to exceed chance. The Standard Error of Measurement of each scale were between 0.3 and 0.5, indicating that only two subscales showed reasonable sensitivity to change (-)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
	Persson, Wennman-Larsen, Sundin, & Gustavsson (2008) Assessing informal caregivers' experiences: A qualitative and psychometric evaluation of the Caregiver Reaction Assessment Scale [273].	A	5-dimensional, no goodness of fit indicators reported (?)	V	α subscales were=0.84; 0.83; 0.76; 0.86 and 0.80 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	A	Results in line with 2 hypo's (+2), not in line with 4 hypo's (-4)	N.a.	n.a.
	Misawa, Miyashita, Kawa, Abe, Abe, Nakayama, & Given (2009) Validity and reliability of the Japanese version of the Caregiver Reaction Assessment Scale (CRA-J) for community-dwelling cancer patients [274].	I	5-dimensional, no goodness of fit indicators reported (?)	V	α subscales were=0.80; 0.87; 0.88; 0.83; 0.73 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	A	Results in line with 5 hypo's (+5)	N.a.	N.a.

Scale	Study																
	Malhotra, Chan, Malhotra, & Ostbye (2012) Validity and reliability of the Caregiver Reaction Assessment scale among primary informal caregivers for older persons in Singapore [275].	V	5-dimensional (c2=1123.00; p<0.005; RMSEA=0.08; CFI=0.84) or 4-dimensional (c2=737.97; p<0.05; RMSEA=0.07; CFI=0.89) (-)	V	α subscales were=0.76; 0.66; 0.82; 0.67; 0.81. (?)	V	Partial invariance of the scale across language (Chinese, English and Malay) (-)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	V	Results in line with 6 hypo's (6+)	N.a.	N.a.
	Stephan, Mayer, Renom Guiteras, & Meyer (2013) Validity, reliability, and feasibility of the German version of the Caregiver Reaction Assessment scale (G-CRA): A validation study [276].	A	5-dimensional, no goodness of fit indicators reported (?)	V	α subscales were=0.76; 0.77; 0.67; 0.78; 0.78 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	A	Results in line with 3 hypo's (3+)	N.a.	N.a.
Caregiver Strain Index (1983) [19]	Robinson (1983) Validation of a Caregiver Strain Index [19].	N.a.	N.a.	D	α=0.86 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	D	Results in line with 5 hypo's (+5), not in line with 4 hypo's (-4)	N.a.	N.a.
	McGartland Rubio, Berg-Weger, & Tebb (1999) Assessing the validity and reliability of well-being and stress in family caregivers [277].	A	3-dimensional (c2(43)=52.91; p=0.14; RMSEA=0.04;) (+)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
	Chen, & Hu (2002) The generalizability of Caregiver Strain Index in family caregivers of cancer patients [278].	N.a.	N.a.	D	Average α total scale at four measurement occasions=0.66 (-)	N.a.	N.a.	I	Average ICC four measurement occasions=0.56 (-)	N.a.	N.a.	N.a.	N.a.	I	Results in line with 2 hypo's (2+)	N.a.	N.a.
	Van Exel, Scholte Op Reimer, Brouwer, Van den Berg, Koopmanschap, & Van den Bos (2004) Instruments for assessing the burden of informal caregiving for stroke patients in clinical practice: A comparison of CSI, CRA, SCO and self-rated burden [271].	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	V	Results in line with 6 hypo's (6+), results not in line with 1 hypo (1-)	N.a.	N.a.
	Post, Festen, Van de Port, & Visser-Meily (2007) Reproducibility of the Caregiver Strain Index and the Caregiver Reaction Assessment in partners of stroke patients living in the Dutch community [272].	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	A	ICC=0.93 (+)	A	SDD was 2.8 points (large effect), SEM of 1.0 point corresponds to 7.1% of the maximum scale range (is below the 10% threshold for reasonable responsiveness) (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
	Ugur, & Fadiloglu (2010) "Caregiver strain index" validity and reliability in Turkish society [279].	A	4-dimensional, no goodness of fit indicators reported (?)	D	α T1 and T2=0.77 and 0.73; α for dimensions all at least 0.77) (+)	N.a.	N.a.	D	Intra item Pearson Correlation Coefficient=0.75 (+)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
	Othman, & Siongteck (2014) Validation of Malay Caregiver Strain Index [280].	N.a.	N.a.	D	α=0.79 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.

Instrument	Reference																	
	García-Domíngue, Martínez-Ginés, Carmona, Caminero, Prefasi, Maurino, et al. (2019) Measuring burden in caregivers of people with multiple sclerosis: Psychometric properties of the CSI questionnaire [281].	D	Unidimensional, no goodness of fit indicators reported (?)	D	α=0.91 (+)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	V	AUC=0.77 (+)	N.a.	N.a.	N.a.	N.a.
Modified Caregiver Strain Index (2003) [69]	Thornton, & Travis (2003) Analysis of the reliability of the Modified Caregiver Strain Index [77].	N.a.	N.a.	D	α T1=0.90; T2=0.88 (?)	N.a.	N.a.	D	ICC=0.88 (+)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
	Chan, Chan, & Suen (2013) Validation of the Chinese version of the Modified Caregivers Strain Index among Hong Kong caregivers: An initiative of medical social workers [282].	A	Unidimensional, no goodness of fit indicators reported (?)	V	α=0.91 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	V	Results in line with 2 hypo's (+2)	N.a.	N.a.		
Carer Assessment Scale (1998) [205]	Mackenzie, Holroyd, & Lui (1998) Community nurses' assessment of the needs of Hong Kong family carers who are looking after stroke patients [234].	N.a.	N.a.		N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
Carer's Assessment of Managing Index (1998) [206]	McKee, Spazzafumo, Nolan, Wojszel, Lamura, & Bien (2009) Components of the difficulties, satisfactions and management strategies of carers of older people: A principal component analysis of CADI-CASI-CAMI [283].	A	10-dimensional, no goodness of fit indicators reported (?)	I	α only reported for the seven sub dimensions with the highest alpha's=0.64-0.80 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
Modified Pearlin Burden Scale (1990) [149]	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
Oberst Caregiving Burden Scale (1991) [142]	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
Positive Aspects of Caregiving Questionnaire (2004) [211]	Tarlow, Wisniewski, Belle, Rubert, Ory, & Gallagher-Thompson (2004) Positive Aspects of Caregiving: Contributions of the REACH Project to the development of new measures for Alzheimer's caregiving [241].	V	2-dimensional (c2(26)=54.47; RMSEA=0.059) (+)	V	α total scale=0.89; subscales=0.86 and 0.80 (+)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	D	Results in line with 6 hypo's (6+)	N.a.	N.a.
	Las Hayas, Lopez, & Calvete (2014) Positive Aspects of Caregiving in Spanish caregivers of individuals with acquired brain injury [284].	V	2-dimensional (c2(25)=39; RMSEA=0.067; CFI=0.98; SMSR=0.06) (+)	V	α total scale=0.82; subscales=0.74 and 0.81 (+)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	V	Results in line with 4 hypo's (4+), not in line with 1 hypo (1-)	N.a.	N.a.
	Lou, Lau, & Cheung (2015) Positive Aspects of Caregiving (PAC): Scale validation among Chinese dementia caregivers (CG) [285].	A	2-dimensional, no goodness of fit indicators reported (?)	V	α total scale=0.89; for the subscales=0.85 and 0.84 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	I	Results in line with 8 hypo's (8+), not in line with 4 hypo's (4-)	N.a.	N.a.
Relative Stress Scale (1982) [223]	Greene, Smith, Gardiner, & Timbury (1982) Measuring behavioural disturbance of elderly demented patients in the community and its effects on relatives: A factor analytic study [84].	I	3-dimensional, no goodness of fit indicators reported (?)	N.a.	N.a.	N.a.	N.a.	A	Correlation coefficient 0.72-0.88 (+)	N.a.	N.a.	N.a.	N.a.	I	Results in line with 3 hypo's (+3), not in line with 1 hypo (-1)	N.a.	N.a.

Instrument	Reference																
	Ulstein, Bruun, & Engedal (2007) The Relative Stress Scale, a useful instrument to identify various aspects of carer burden in dementia? [286].	A	3-dimensional, no goodness of fit indicators reported (?)	V	α for subscales=0.84; 0.86 and 0.70 (+)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
Revised 15-item Bakas Caregiving Outcomes Scale (2006) [171]	Bakas, Champion, Perkins, Farran, & Williams (2006) Psychometric testing of the Revised 15-item Bakas Caregiving Outcomes Scale [197].	V	Unidimensional (CFI=0.91; RMSEA=0.31) (-)	V	α total scale=0.90 (?)	N.a.	N.a.	A	ICC=0.66 (-)	N.a.	N.a.	N.a.	N.a.	V	Results in line with 2 hypo's (2+)	N.a.	N.a.
	Govina, Kotronoulas, Mystakidou, Giannakopoulou, Galanos, & Patiraki (2013) Validation of the revised Bakas Caregiving Outcomes Scale in Greek caregivers of patients with advanced cancer receiving palliative radiotherapy [287].	A	Unidimensional (TLI=0.91; RMSEA=0.08; CFI=0.90) (-)	V	α total scale=0.83 (+)	N.a.	N.a.	A	ICC=0.99 (+)	N.a.	N.a.	V	AUC=0.94; sensitivity=91%; specificity=86% (+)	A	Results in line with 4 hypo's (4+)	N.a.	N.a.
Sense of Competence Questionnaire (1993) [196]	Vernooij-dassen (1993) Dementie en thuiszorg: Een onderzoek naar determinenten van het competentiegevoel van centrale verzorgers en het effect van professionele interventie [Dementia and home care: Determinants of the sense of competence of primary caregivers ...] [288].	A	3-dimensional, no goodness of fit indicators reported (?)	V	α total scale=0.79; subscales=0.55; 0.63 and 0.50 (-)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
	Scholte op Reimer, De Haan, Pijnenborg, Limburg, & Van den Bos (1998) Assessment of burden in partners of stroke patients with the Sense of Competence Questionnaire [289].	A	3-dimensional, no goodness of fit indicators reported (?)	V	α total scale=0.83; subscales=0.68; 0.77 and 0.75 (?)	N.a.	N	A	ICC total scale: 0.93; subscales: 0.84; 0.89 and 0.92 (+)	N.a.	N.a.	N	N	A	Results in line with 4 hypo's (4+)	N.a.	N.a.
	Van Exel, Scholte Op Reimer, Brouwer, Van den Berg, Koopmanschap, & Van den Bos (2004) Instruments for assessing the burden of informal caregiving for stroke patients in clinical practice: A comparison of CSI, CRA, SCO and self-rated burden [271].	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	V	Results in line with 6 hypo's (6+), results not in line with 1 hypo (1-)	N.a.	N.a.
	Jansen, Van Hout, Van Marwijk, Nijpels, Gundy, Vernooij-Dassen, De Vet, Schellevis, & Stalman (2007) Sense of Competence Questionnaire among informal caregivers of older adults with dementia symptoms: A psychometric evaluation [290].	I	3-dimensional, no goodness of fit indicators reported (?)	V	α sub dimensions were=0.83; 0.83 and 0.85 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	I	Results in line with 2 hypo's (2+), not in line with 10 hypo's (-10)	N.a.	N.a.
	Pendergrass, Beische, Becker, Hautzinger, & Pfeiffer (2015) An abbreviated German version of the Sense of Competence Questionnaire among informal caregivers of relatives who had a stroke: Development and validation [291].	A	3-dimensional, no goodness of fit indicators reported (?)	V	α total scale=0.89; subscales=0.82-0.85 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	A	Results in line with 4 hypo's (4+)	N.a.	N.a.
Zarit Burden Interview (1980) [80]	Zarit, Reever, & Bach-Peterson (1980) Relatives of the impaired elderly: Correlates of feelings of burden [89].	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	I	Results in line with 1 hypo (+1), not in line with 6 hypo's (-6)	N.a.	N.a.

Study																
Arai, Kudo, Hosokawa, Washio, Miura, & Hisamichi (1997) Reliability and validity of the Japanese version of the Zarit Caregiver Burden interview [292].	N.a.	N.a.	D	α=0.93 (?)	D	Correlations between Zarit Burden Interview and CES-D and a single global rating comparable to the original study, demographic distribution of the scores had a similar trend to those of the original version (?)	A	Pearson's product moment correlation coefficient=0.76 (+)	N.a.	N.a.	N.a.	N.a.	A	Results in line with 2 hypo's (+2)	N.a.	N.a.
Arai, & Washio (1999) Burden felt by family caring for the elderly members needing care in southern Japan.[293]	N.a.	N.a.	D	α=0.93 (?)	D	Correlations between Zarit Burden Interview and CES-D and a single global rating comparable to former study in Japan and original study (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	A	Results in line with 2 hypo's (+2)	N.a.	N.a.
Knight, Fox, & Chou (2000) Factor structure of the Burden Interview [294].	V	3-dimensional (sample 1: c2(75)=207.30, p<0.001, CFI=0.91; sample 2: c2(75)=102.71, p=0.01, CFI=0.94) (-)	I	α=0.92 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
Taub, Andreoli, & Bertolucci (2004) Dementia caregiver burden: Reliability of the Brazilian version of the Zarit Caregiver Burden Interview [295].	N.a.	N.a.	D	α=0.77 and 0.80 (?)	N.a.	N.a.	A	ICC=0.88 (+)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
Ankri, Andrieu, Beaufils, Grand, & Henrard (2005) Beyond the global score of the Zarit Burden Interview: Useful dimensions for clinicians [296].	A	5-dimensional, no goodness of fit indicators reported (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
Schreiner, Morimoto, Arai, & Zarit (2006) Assessing family caregiver's mental health using a statistically derived cut-off score for the Zarit Burden Interview [297].	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	V	Sensitivity 47-77%; specificity 59-79% (±)	N.a.	N.a.	N.a.	N.a.
Lai (2007) Validation of the Zarit Burden Interview for Chinese Canadian caregivers [298].	V	5-dimensional (RMSEA=0.09; CFI=0.093) (-)	V	α subscales=0.88; 0.88; 0.72; 0.67; 0.70 (±)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
Ko, Yip, Liu, & Huang (2008) Chinese version of the Zarit caregiver Burden Interview: A validation study [299].	A	5-dimensional, no goodness of fit indicators reported (?)	I	α total scale=0.88 (?)	N.a.	N.a.	A	ICC=0.88 (+)	N.a.	N.a.	N.a.	N.a.	A	Results in line with 2 hypo's (+2)	N.a.	N.a.

Study																
Lu, Wang, Yang, & Feng (2009) Zarit Caregiver Burden Interview: Development, reliability and validity of the Chinese version [300].	V	5-dimensional (c2(160)=660.9 p=0.001; RMSEA=0.077; CFI=0.841) (-)	I	α total scale=0.88 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
Braun, Scholz, Hornung, & Martin (2010) The burden of spousal caregiving: A preliminary psychometric evaluation of the German version of the Zarit Burden Interview [301].	N.a.	N.a.	D	α total scale=0.92 (?)	I	Correlation between Zarit Burden Interview (21 items) and the global rating item of the Zarit Burden Interview (1 item) comparable to the corresponding correlation in the original version (?)	N.a.	N.a.	N.a.	N.a.	N.a.	D	Results in line with 2 hypo's (+2)	N.a.	N.a.
Martin-Carrasco, Otermin, Perez-Camo, Pujol, Aguera, Martin, Gobartt, Pons, & Balana (2010) EDUCA study: Psychometric properties of the Spanish version of the Zarit Caregiver Burden Scale [302].	D	3-dimensional, no goodness of fit indicators reported(?)	I	α total scale=0.92 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	V	Results in line with 1 hypo (+1)		
Seng, Luo, Yee, Lim, Chionh, Goh, & Yap (2010) Validity and reliability of the Zarit Burden Interview in assessing caregiving burden [303].	N.a.	N.a.	D	α total scale=0.93 (?)	N.a.	N.a.	A	ICC=0.89 (+)	N.a.	N.a.	N.a.	N.a.	A	Convergent: results in line with 4 hypo's (+4), known-groups: results in line with 3 hypo's, one in line, but not significant (3+, 1+/-)	N.a.	N.a.
Siegert, Jackson, Tennant, & Turner-Strokes (2010) Factor analysis and Rasch analysis of the Zarit Burden Interview for acquired brain injury carer research [304].	V	2-dimensional (c2(43)=175.74; RMSEA=0.119; CFI=0.90) (-)	V	α sub dimensions=0.82 and 0.80 (+)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
Chattat, Cortesi, Izzicupo, Del Re, Sgarbi, Fabbo, & Bergonzini (2011) The Italian version of the Zarit Burden Interview: A validation study [305].	A	5-dimensional (c2(231)=2301, p=0,001) (-)	I	α total scale=0.90 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	A	Results in line with 3 hypo's (+3)	N.a.	N.a.
Cheah, Han, Chong, Anthony, & Lim (2012) Multidimensionality of the Zarit Burden Interview across the severity spectrum of cognitive impairment: An Asian perspective [306].	A	4-dimensional, no goodness of fit indicators reported (?)	I	α total scale=0.92 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
Özer, Yurttas, & Akyil (2012) Psychometric evaluation of the Turkish version of the Zarit Burden Interview in family caregivers of inpatients in medical and surgical clinics [307].	A	Unidimensional, no goodness of fit indicators reported (?)	V	α test and retest were=0.82 and 0.87 (+)	N.a.	N.a.	D	Test-retest correlation=0.72 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
Torres, Hoff, Padovani, & Ramos-Cerqueira (2012) Dimensional analysis of burden in family caregivers of patients with obsessive-compulsive disorder [308].	I	6-dimensional, no goodness of fit indicators reported (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.

Study															
Cheng, Kwok, & Lam (2014) Dimensionality of burden in Alzheimer caregivers: Confirmatory factor analysis and correlates of the Zarit Burden Interview [309].	V	4-dimensional (c2(130)=217.80, p<0.001; CFI=0.95; RMSEA=0.06; NNFI=0.94; SRMS=0.06)(+)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
Lim, Cheah, Ali, Han, Anthony, Chan, & Chong (2014) Worry about performance: A unique dimension of caregiver burden [310].	N.a.	N.a.	V	α total scale=0.92; four sub dimensions had α's ranging 0.83-0.89 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
Chan, Lam, & Chiu (2015) Validation of the Chinese version of the Zarit Burden Interview [311].	N.a.	N.a.	D	Split half correlation coefficient=0.81 (?)	N.a.	N.a.	A	ICC=0.99 (+) (inter-rater instead of intra-rater)	N.a.	N.a.	N.a.	N.a.	D	Results in line with 2 hypo's (+2)	
Galindo-Vazquez, Benjet, Cruz-Nieto, Rojas-Castillo, Riveros-Rosas, Meneses-Garcia, Aguilar-Ponce, Alvarez, Avitia, & AlvaradoAguilar (2015) Psychometric properties of the Zarit Burden Interview in Mexican caregivers of cancer patients [312].	A	4-dimensional, no goodness of fit indicators reported (?)	V	α total scale=0.90; decided to chose for 3 subscales with the following α's=0.90; 0.74 and 0.67 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	V	Results in line with 3 hypo's (+3)	N.a.	N.a.
Al-Rawashdeh, Lennie, & Chung (2016) Psychometrics of the Zarit Burden Interview in caregivers of patients with heart failure [313].	A	4-dimensional, no goodness of fit indicators reported (?)	I	α total scale=0.92 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	V	Results in line with 3 hypo's (+3)	N.a.	N.a.
Bianchi, Flesch, Alves, Batistoni, & Neri (2016) Zarit Burden Interview psychometric indicators applied in older people caregivers of other elderly [314].	A	3-dimensional, no goodness of fit indicators reported (?)	V	α total scale=0.86; subscales=0.83; 0.72; 0.72 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	A	Results in line with 3 hypo's (+3)	N.a.	N.a.
Goncalves-Pereira, Gonzalez-Fraile, Santos-Zorrozua, Martin-Carrasco, Fernandez-Catalina, Dominguez-Panchon, Munoz-Hermoso, & Ballesteros (2017) Assessment of the consequences of caregiving in psychosis: A psychometric comparison of the Zarit Burden Interview (ZBI) and the Involvement Evaluation Questionnaire (IEQ) [315].	N.a.	N.a.	V	α total scale=0.91; subscales=0.84 and 0.86 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	A	Result in line with 3 hypo's (3+)	V	Results in line with 1 hypo (1+)
Hagell, Alvariza, Westergren, & Arestedt (2017) Assessment of burden among family caregivers of people with Parkinson's Disease using the Zarit Burden Interview [316].	N.a.	N.a.	D	Ordinal α for the total scale=0.95 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	D	Result in line with 5 hypo's (5+)	N.a.	N.a.
Imarhiagbe, Asemota, Oripelay, Akpekpe, Owolabi, Abidakun, Akemokwe, Ogundare, Azzez, & Osakue (2017) Burden of informal caregivers of stroke survivors: Validation of the Zarit burden interview in an African population [317].	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	V	AUC=0.69 (-)	A	Result in line with 1 hypo (1+)	N.a.	N.a.

Instrument	Study	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Landfeldt, Mayhew, Straub, Bushby, Lochmuller, & Lindgren (2017) Psychometric properties of the Zarit Caregiver Burden Interview administered to caregivers to patients with Duchenne muscular dystrophy: A Rasch analysis [318].	V	Rasch-analyses ($c_2(198)=499$ (-))	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
	Tang, Yu, Liu, Lin, Chen, Zhao, Xiao (2017) Factor analyses of the Chinese Zarit Burden Interview among caregivers of patients with schizophrenia in a rural Chinese community [319].	V	Confirmatory factor analysis: none of the existing models fit the data well. Exploratory factor analysis: 5 dimensions ($c_2(134)=381.13$; p=0.001; RMSEA=0.059; CFI=0.97) (+)	V	α total scale=0.88; subscales=0.68-0.84 (±)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
	Oh, & Kim (2018) Factor analysis of the Zarit Burden Interview in family caregivers of patients with amyotrophic lateral sclerosis [320].	A	3-dimensional; no goodness of fit indicators reported (?)	V	α ranged 0.68-0.92 (+)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
	Smith, George, & Ferreira (2018) Factors emerging from the "Zarit Burden Interview" and predictive variables in a UK sample of caregivers for people with dementia [321].	A	3-dimensional; no goodness of fit indicators reported (?)	I	α total scale=0.91 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
	Nagata, Yada, & Inagaki (2018) Exploration of the factor structure of the burden experienced by individuals providing end-of-life care at home [322].	V	4-dimensional (CFI=0.78; RMSEA=0.11) (-)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
	Vatter, McDonald, Stanmore, Clare, & Leroi (2018) Multidimensional care burden in Parkinson-related dementia [323].	A	5-dimensional; no goodness of fit indicators reported (?)	V	α total scale=0.92; subscales=0.77-0.85 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
	Yu, Liu, Zhou, Chen, Zhang, Hu, et al. (2018) Assessment of burden among family caregivers of schizophrenia: Psychometric testing for short-form Zarit Burden Interviews [324].	N.a.	N.a.	D	ω=0.89 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a	N.a.	A	Results in line with 5 hypo's (5+)	N.a.	N.a.
Zarit Burden Interview Short Form (2001) [112]	Bédard, Molloy, Squire, Dubois, Lever, & O'Donnell (2001) The Zarit Burden Interview: A new short version and screening version [126].	A	2-dimensional; no goodness of fit indicators reported (?)	V	α total score=0.88; sub dimensions: 0.89 and 0.77 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	V	Correlation original scale=0.92-0.97 (+)	V	Results in line with 8 hypo's (8+)	N.a.	N.a.
	O'Rourke, & Tuokko (2003) Psychometric properties of an abridged version of the Zarit Burden Interview within a representative Canadian caregiver sample [325].	V	2-dimensional ($c_2(36)=79.7$ p<.01; RMSEA=0.037; CFI=0.99) (+)	I	α total scale=0.85 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	V	No loss of predictive validity when brief version is used; specificity=90%; sensitivity=49% (±)	N.a.	N.a.	N.a.	N.a.

Study																
O'Rourke, & Tuokko (2003) The relative utility of four abridged versions of the Zarit Burden interview [326].	V	2-dimensional (c2(39)=82.93; p<0,01; RMSEA=0.047; CFI=0.98) (+)	V	α total scale=0.85; subscales= 0.89 and 0.77 (+)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.		V	Results in line with 1 hypo (1+)	N.a.	N.a.
Bachner, & Ayalon (2010) Initial examination of the psychometric properties of the short Hebrew version of the Zarit Burden Interview [327].	A	2-dimensional, no goodness of fit indicators reported (?)	V	α total scale=0.83; subscales: 0.86 and 0.76 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.		A	Results in line with 2 hypo's (2+)	N.a.	N.a.
Higginson, Gao, Jackson, Murray, & Harding (2010) Short-form Zarit Caregiver Burden Interviews were valid in advanced conditions [328].	N.a.	N.a.	I	α respectively for caregivers of patients with advanced cancer, dementia and acquired brain injury= 0.85; 0.87 and 0.89 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	V	Correlation with full-version for caregivers of patients with advanced cancer, dementia and acquired brain injury=0.95; 0.96; 0.97; AUC=0.99; sensitivity=92%; specificity=94% (+)	A	Results in line with 4 hypo's (4+)	N.a.	N.a.
Bachner, O'Rourke, Ayalon, & Bedard (2011) Comparison of caregiver responses to English and Hebrew language versions of an abridged Zarit Burden Interview [329].	V	2-dimensional (Isreal: c2(41)=68.40; RMSEA=0.069; CFI=0.96; Canada: c2(48)=67.58; RMSEA=0.045; CFI=0.98,) (+)	I	α Hebrew and Canadian version respectively= 0.83 and 0.90 (+)	A	Findings suggest generally reliable translation from English to Hebrew, since it appears that both Israeli and Canadian caregivers interpret and respond to the majority of the brief items in a consistent manner (based on group factor analysis) (+)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
Brink, Stones, & Smith (2012) Confirmatory factor analysis of the Burden Interview of the caregivers of terminally ill home care clients [330].	A	2-dimensional (c2(53)=131.72) (+)	V	α subscales=0.90 and 0.76 (+)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.
Iecovich (2012) Psychometric properties of the Hebrew version of the Zarit Caregiver Burden Scale short version [331].	A	2-dimensional (c2(66)=2772.38) (+)	V	α total scale=0.85; subscales=0.88 and 0.91 (+)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	V	Results in line with 7 hypo's (7+)	N.a.	N.a.
Bachner (2013) Preliminary assessment of the psychometric properties of the abridged Arabic version of the Zarit Burden Interview among caregivers of cancer patients [332].	A	2-dimensional, no goodness of fit indicators reported (?)	V	α total scale=0.77; subscales=0.81 and 0.77 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	A	Results in line with 3 hypo's (+3)	N.a.	N.a.

Study																
Rajabi-Mashhadi, Mashhadinejad, Ebrahimzadeh, Golhasani-Keshtan, et al. (2015) The Zarit Caregiver Burden Interview Short Form (ZBI-12) in spouses of veterans with chronic spinal cord injury, validity and reliability of the Persian version [333].	A	2-dimensional, no goodness of fit indicators reported (?)	I	α total scale=0.78 (?)	N.a.	N.a.	A	ICC=0.78 (+)	N.a.	N.a.	N.a.	N.a.	A	No clear hypo's stated (?)	N.a.	N.a.
Stagg, & Larner (2015) Zarit Burden Interview: Pragmatic study in a dedicated cognitive function clinic [334].	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	V	Correlation (Kappa)=0.86 (+)	N.a.	N.a.	N.a.	N.a.
Branger, O'Connell, & Morgan (2016) Factor analysis of the 12-Item Zarit Burden Interview in caregivers of persons diagnosed with dementia [335].	A	2-dimensional, no goodness of fit indicators reported (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a	N.a.	A	Results in line with 1 hypo (1+), not in line with 1 hypo (-)	N.a.	N.a.
Tang, Ho, Luo, Wong, Lau, Lum, & Cheung (2016) Validating a Cantonese short version of the Zarit Burden Interview (CZBI-Short) for dementia caregivers [336].	V	3-dimensional (c2(41)=87.0 p<0.001, CFI=0.94, RMSEA=0.07, SRMR=0.05) (-)	V	α total scale=0.84; for subscales=0.82; 0.87 and 0.59 (±)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	A	Results in line with 5 hypo's (5+), not in line with 2 hypo's (2-)	N.a.	N.a.
Hagell, Alvariza, Westergren, & Arestedt (2017) Assessment of Burden Among Family Caregivers of People With Parkinson's Disease Using the Zarit Burden Interview [316].	N.a.	N.a.	D	Ordinal alpha=0.91 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	V	Correlation=0.99; AUC=0.98 (+)	D	Results in line with 5 hypo's (5+)	N.a.	N.a.
Yu, Yap, & Liew (2018) The optimal short version of the Zarit Burden Interview for dementia caregivers: Diagnostic utility and externally validated cut-offs [337].	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	V	AUC=0.86; sensitivity=79.6%; specificity=76.2% (+)	V	Results in line with 1 hypo (1+)	N.a.	N.a.
Gratão, Brigola, Ottaviani, Luchesi, Souza, Rossetti, et al. (2019) Brief version of Zarit Burden Interview (ZBI) for burden assessment in older caregivers [338].	N.a.	N.a.	V	α total scale=0.81 (+)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	D	Results in line with 3 hypo's (3+)	N.a.	N.a.
Yu, Liu, Zhou, Chen, Zhang, Hu, et al. (2018) Assessment of burden among family caregivers of schizophrenia: Psychometric testing for short-form Zarit Burden Interviews [324].	N.a.	N.a.	D	ω=0.84 (?)	N.a.	N.a.	N.a.	N.a.	N.a.	N.a.	V	AUC=0.98; correlation=0.70 (+)	A	Reults in line with 4 hypo's; not in line with 1 hypo (1-)	N.a.	N.a.

N.a.: not applicable; -: insufficient; +: sufficient; ±: inconsistent; ?: inteterminate; ?/-: inconsistent based on indeterminate and insufficient scores; +/?: inconsistent based on sufficient and indeterminate scores; ?/+/-: inconsistent scores based on indeterminate, sufficient and insufficient scores; RMSEA: root mean

[a] Internal consistency is rated '+' when there is at least low evidence for sufficient structural validity and α ≥ 0.70 for each unidimensional scale or subscale. Internal consistency is rated '?' if α ≥ 0.70, but when the criteria for at least low evidence for sufficient structural validity have not met. Often rated