

Additional Files:

Preference-based outcome measures in people with dementia: A systematic review of psychometric properties

Anouk Lamé, MSc^{1,2,3}, Jan Heijdra Suasnabar, MSc¹, Esmee Adam, MSc^{2,3}, Ron Handels, PhD^{5,6}, Hanneke J.A. Smaling, PhD^{2,3}, Miriam L. Haaksma, PhD^{2,3,4}, M. Elske van den Akker-van Marle, PhD¹

¹Medical Decision Making, Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands

²University Network for the Care sector South-Holland, Leiden University Medical Center, The Netherlands

³Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, The Netherlands

⁴Department of Internal Medicine, Division of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands

⁵Department of Neurobiology, Care Sciences and Society, Division of Neurogeriatrics, BioClinicum J9:20, Karolinska Institutet, Solna, Sweden

⁶Maastricht University, Alzheimer Centre Limburg, Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Faculty of Health Medicine and Life Sciences, Maastricht, Netherlands

Corresponding author: Anouk Lamé, a.lame@lumc.nl

Additional File 1: Inclusion/exclusion criteria and search strategy

<i>Inclusion criteria</i>	1. Older adults living with dementia, either community-dwelling or living in residential care. 2. Preference-based measures (self- and/or proxy-reported). 3. One or more psychometric properties are described.
<i>Exclusion criteria</i>	• Early onset dementia or (mild) cognitive impairment with other clearly described etiologies (e.g., Parkinson's dementia). • Not published in English. • Reviews, commentaries, protocols, editorials, dissertations, abstracts and conference proceedings.
<i>Search strategy</i>	• Sources: PubMed, EMBASE, PsycINFO, Web of Science. Search was done in July 2025.
<i>Search terms</i>	• Keywords (and MeSH terms) focused on: dementia, (health-related) quality of life, preference-based measures. • This was combined with the COSMIN methodological search filter for assessing psychometric properties. • Full search strategy can be found in <i>Table S2</i>.

Additional File 2: Full search strategy and number of results in PubMed, Embase, PsycINFO and Web of Science, executed in July 2025.

Database	Search	Results
PubMed	("Alzheimer Disease"[Mesh] OR "Alzheimer*" [tiab] OR "Dement*" [tiab] OR "cognitive declin*" [tiab] OR "cognitive dysfunct*" [tiab] OR "cognitive impair*" [tiab] OR "cognitive deterior*" [tiab] OR "Homes for the Aged"[Mesh] OR "Homes for the Aged" [tiab] OR "Nursing home resident*" [tiab] OR "community-dwelling" [tiab]) AND ("Surveys and Questionnaires"[Mesh] OR "Survey*" [tiab] OR "Questionnaire*" [tiab] OR "self-report*" [tiab] OR "interview*" [tiab] OR "care-giver*" [tiab] OR "caregiver*" [tiab] OR "carer*" [tiab] OR "patient-outcome assessment" [tiab] OR "proxy" [tiab] OR "Patient Outcome Assessment"[Mesh] OR "outcome measure*" [tiab] OR "instrument*" [tiab] OR "eq5d*" [tiab] OR "eq-5d*" [tiab] OR "demqol*" [tiab] OR "ascot*" [tiab] OR "icecap*" [tiab] OR "qol-acc*" [tiab] OR "qolacc*" [tiab] OR "WOOP*" [tiab] OR "qol-ad*" [tiab] OR "qolad*" [tiab] OR "dqi" [tiab]) AND (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 of variation" [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 [tw] OR precision [tw] OR imprecision [tw] OR "precise values" [tw] 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* [tw] OR ((replicab* [tw] OR repeated [tw]) AND (measure [tw] OR measures [tw] OR findings [tw] OR result [tw] OR results [tw] OR test [tw] OR tests [tw])) 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 "factor structure" [tiab] OR "factor structures" [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 "interval variability" [tiab] OR "rate 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 (limit [tiab] AND detection [tiab]) OR "minimal detectable concentration" [tiab] OR interpretab* [tiab] OR ((minimal [tiab] OR minimally [tiab] OR clinically [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]) NOT ('delphi-technique' [ti] OR cross-sectional [ti] OR "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]) AND ("Quality of Life" [Mesh] OR "Quality of Life" [tiab] OR "QoL" [tiab] OR "Life Quality" [tiab] OR "utilit*" [tiab] OR "preference-based" [tiab] OR "preference based" OR "wellbeing" [tiab] OR "well-being" [tiab] OR "HRQoL" [tiab] OR "quality-adjusted life-years" [tiab])	9,610
Embase	1 exp "Alzheimer Disease"/ or "Alzheimer*" .ti,ab. or "Dement*" .ti,ab. or "cognitive declin*" .ti,ab. or "cognitive dysfunct*" .ti,ab. or "cognitive impair*" .ti,ab. or "cognitive deterior*" .ti,ab. or exp "Home for the Aged"/ or "Homes for the Aged".ti,ab. or "Nursing home resident*" .ti,ab. or "community-dwelling".ti,ab. or "community dwelling".ti,ab. 2 exp "Health Survey"/ or exp "Questionnaire"/ or "Survey*" .ti,ab. or "Questionnaire*" .ti,ab. or "self-report*" .ti,ab. or "interview*" .ti,ab. or "care-giver*" .ti,ab. or "caregiver*" .ti,ab. or "carer*" .ti,ab. or "patient-outcome assessment".ti,ab. or "proxy".ti,ab. or exp "Outcome Assessment"/ or "outcome measure*" .ti,ab. or "instrument*" .ti,ab. or "eq5d*" .ti,ab. or "eq-5d*" .ti,ab. or "demqol*" .ti,ab. or "ascot*" .ti,ab. or "icecap*" .ti,ab. or "qol-acc*" .ti,ab. or "qolacc*" .ti,ab. or "WOOP*" .ti,ab. or "qol-ad*" .ti,ab. or "qolad*" .ti,ab. or "dqi".ti,ab. 3 ("instrumentation" or "methods").mp. or ("validation study" OR "comparative study").mp. or exp "Psychometrics"/ or "psychometr*".tw. or ("clinimetr*" OR "clinometr*").mp. or exp "Outcome Assessment, Health Care"/ or "outcome assessment".tw. or "outcome measure*".mp. or exp "Observer Variation"/ or "observer variation".tw. or exp "Health Status Indicators"/ or exp "Reproducibility of Results"/ or "reproducib*".tw. or exp "Discriminant Analysis"/ or ("reliab*" OR "unreliab*" OR "valid*" OR	8,435

	"coefficient of variation" OR "coefficient" OR "homogeneity" OR "homogeneous" OR "internal consistency").tw. or ("cronbach*" AND ("alpha" OR "alphas")).tw. or ("item" AND ("correlation*" OR "selection*" OR "reduction*")).tw. or "agreement".mp. or "precision".mp. or "imprecision".mp. or "precise values".mp. or "test-retest".tw. or ("test" AND "retest").tw. or ("reliab*" AND ("test" OR "retest")).tw. or "stability".tw. or ("interrater" OR "inter-rater" OR "intrarater" OR "intra-rater").tw. or ("intertester" OR "inter-tester" OR "intratester" OR "intra-tester").tw. or ("interobserver" OR "inter-observer" OR "intraobserver" OR "intra-observer").tw. or ("intertechnician" OR "inter-technician" OR "intratechnician" OR "intra-technician").tw. or ("interexaminer" OR "inter-examiner" OR "intraexaminer" OR "intra-examiner").tw. or ("interassay" OR "inter-assay" OR "intraassay" OR "intra-assay").tw. or ("interindividual" OR "inter-individual" OR "intraindividual" OR "intra-individual").tw. or ("interparticipant" OR "inter-participant" OR "intraparticipant" OR "intra-participant").tw. or "kappa".tw. or "kappas".tw. or "repeatab*".mp. or (("replicab*" OR "repeated") AND ("measure" OR "measures" OR "findings" OR "result" OR "results" OR "test" OR "tests")).mp. or ("generaliza*" OR "generalisa*").tw. or "concordance".tw. or (intraclass AND correlation*).tw. or discriminative.tw. or "known group".tw. or ("factor analysis" OR "factor analyses" OR "factor structure" OR "factor structures").tw. or dimension*.tw. or subscale*.tw. or (multitrait AND scaling AND (analysis OR analyses)).tw. or "item discriminant".tw. or "interscale correlation*".tw. or (error OR errors).tw. or "individual variability".tw. or "interval variability".tw. or "rate variability".tw. or (variability AND (analysis OR values)).tw. or (uncertainty AND (measurement OR measuring)).tw. or "standard error of measurement".tw. or sensitiv*.tw. or responsive*.tw. or (limit AND detection).tw. or "minimal detectable concentration".tw. or interpretab*.tw. or ((minimal OR minimally OR clinical OR clinically) AND (important OR significant OR detectable) AND (change OR difference)).tw. or (small* AND (real OR detectable) AND (change OR difference)).tw. or "meaningful change".tw. or "ceiling effect".tw. or "floor effect".tw. or "item response model".tw. or IRT.tw. or rasch.tw. or "differential item functioning".tw. or DIF.tw. or "computer adaptive testing".tw. or "item bank".tw. or "cross-cultural equivalence".tw. 4 exp "Quality of Life"/ or "Quality of Life".ti,ab. or "QoL".ti,ab. or "Life Quality".ti,ab. or "utilit*".ti,ab. or "preference-based*".ti,ab. or "preference based*".ti,ab. or "wellbeing".ti,ab. or "well-being".ti,ab. or "HRQoL".ti,ab. or "quality-adjusted life-years".ti,ab. 5 1 and 2 and 3 and 4 9572 6 (address OR biography OR "case reports" OR comment OR directory OR editorial OR festschrift OR interview OR lecture OR "legal case" OR legislation OR letter OR news OR "newspaper article" OR "patient education handout" OR "popular work" OR congress OR "consensus development conference" OR "consensus development conference, NIH" OR "practice guideline").mp 7 1 and 2 and 3 and 4 not 6 7392	
PsycInfo	(DE "Alzheimer Disease" OR TI "Alzheimer*" OR AB "Alzheimer*" OR TI "Dement*" OR AB "Dement*" OR TI "cognitive declin*" OR TI "cognitive dysfunct*" OR TI "cognitive impair*" OR TI "cognitive deterior*" OR DE "Home for the Aged" OR TI "Homes for the Aged" OR TI "Nursing home resident*" OR TI "community-dwelling" OR TI "community dwelling" OR AB "cognitive declin*" OR AB "cognitive dysfunct*" OR AB "cognitive impair*" OR AB "cognitive deterior*" OR AB "Homes for the Aged" OR AB "Nursing home resident*" OR AB "community-dwelling" OR AB "community dwelling") (DE "Health Survey" OR DE "Questionnaire" OR TI "Survey*" OR TI "Questionnaire*" OR TI "self-report*" OR TI "interview*" OR TI "care-giver*" OR TI "caregiver*" OR TI "carer*" OR TI "patient-outcome assessment" OR TI "proxy" OR DE "Outcome Assessment" OR TI "outcome measure*" OR TI "instrument*" OR TI "eq5d*" OR TI "eq-5d*" OR TI "demqol*" OR TI "ascot*" OR TI "icecap*" OR TI "qol-acc*" OR TI "qolacc*" OR TI "WOOP*" OR TI "qol-ad*" OR TI "qolad*" OR TI "dqi" OR AB "Survey*" OR AB "Questionnaire*" OR AB "self-report*" OR AB "interview*" OR AB "care-giver*" OR AB "caregiver*" OR AB "carer*" OR AB "patient-outcome assessment" OR AB "proxy" OR AB "outcome measure*" OR AB "instrument*" OR AB "eq5d*" OR AB "eq-5d*" OR AB "demqol*" OR AB "ascot*" OR AB "icecap*" OR AB "qol-acc*" OR AB "qolacc*" OR AB "WOOP*" OR AB "qol-ad*" OR AB "qolad*" OR AB "dqi") cl("Psychometrics & Statistics & Methodology" OR "Research Methods & Experimental Design") OR TI (psychometr* OR clinimetr* OR clinometr* OR "outcome assessment" OR "outcome measure*" OR "observer variation" OR reproducib* OR reliab* OR unreliab* OR valid* OR coefficient OR homogeneity OR homogeneous OR "internal consistency" OR agreement OR precision OR imprecision OR "precise values" OR test-retest OR reliab* OR stability OR interrater OR inter-rater OR intrarater OR intra-rater OR intertester OR inter-tester OR intratester OR intra-tester OR interobserver OR inter-observer OR intraobserver OR intra-observer OR intertechnician OR inter-technician OR intratechnician OR intra-technician OR interexaminer OR inter-examiner OR intraexaminer OR intra-examiner OR interassay OR inter-assay OR intraassay OR intra-assay OR interindividual OR inter-individual OR intraindividual OR intra-individual OR interparticipant OR inter-participant OR intraparticipant OR intra-participant OR kappa OR kappa's OR kappas OR repeatab* OR generaliza* OR generalisa* OR concordance OR discriminative OR "known group" OR "factor analys*" OR dimension* OR subscale* OR "item discriminant" OR "interscale correlation*" OR error* OR "individual variability" OR "standard error of measurement" OR sensitiv* OR responsive* OR "meaningful change" OR "ceiling effect" OR "floor effect" OR "Item response model" OR IRT OR Rasch OR "Differential item functioning" OR DIF OR "computer adaptive testing" OR "item bank" OR "cross-cultural equivalence")	2,392

	OR AB (psychometr* OR clinimetr* OR clinometr* OR "outcome assessment" OR "outcome measure*" OR "observer variation" OR reproducib* OR reliab* OR unreliab* OR valid* OR coefficient OR homogeneity OR homogeneous OR "internal consistency" OR agreement OR precision OR imprecision OR "precise values" OR test-retest OR reliab* OR stability OR interrater OR inter-rater OR intrarater OR intra-rater OR intertester OR inter-tester OR intratester OR intra-tester OR interobserver OR inter-observer OR intraobserver OR intra-observer OR intertechnician OR inter-technician OR intratechnician OR intra-technician OR interexaminer OR inter-examiner OR intraexaminer OR intra-examiner OR interassay OR inter-assay OR intraassay OR intra-assay OR interindividual OR inter-individual OR intraindividual OR intra-individual OR interparticipant OR inter-participant OR intraparticipant OR intra-participant OR kappa OR kappa's OR kappas OR repeatab* OR generaliza* OR generalisa* OR concordance OR discriminative OR "known group" OR "factor analys*" OR dimension* OR subscale* OR "item discriminant" OR "interscale correlation*" OR error* OR "individual variability" OR "standard error of measurement" OR sensitiv* OR responsive* OR "meaningful change" OR "ceiling effect" OR "floor effect" OR "Item response model" OR IRT OR Rasch OR "Differential item functioning" OR DIF OR "computer adaptive testing" OR "item bank" OR "cross-cultural equivalence") OR TI ("cronbach* alpha*" OR "replicab* measure*" OR "replicab* finding*" OR "replicab* result*" OR "replicab* test*" OR "repeated measure*" OR "repeated finding*" OR "repeated result*" OR "repeated test*" OR "item correlation*" OR "item selection*" OR "item reduction*" OR "Test retest" OR "intraclass correlation*" OR "multitrait scaling analys*" OR "uncertainty measur*" OR "variability analys*" OR "variability value*" OR "minimal* important change" OR "minimal* important difference" OR "minimal* significant change" OR "minimal* significant difference" OR "minimal* detectable change" OR "minimal* detectable difference" OR "clinical* important change" OR "clinical* important difference" OR "clinical* significant change" OR "clinical* significant difference" OR "clinical* detectable change" OR "clinical* detectable difference" OR "small* real change" OR "small* real difference" OR "small* detectable change" OR "small* detectable difference") OR AB ("cronbach* alpha*" OR "replicab* measure*" OR "replicab* finding*" OR "replicab* result*" OR "replicab* test*" OR "repeated measure*" OR "repeated finding*" OR "repeated result*" OR "repeated test*" OR "item correlation*" OR "item selection*" OR "item reduction*" OR "Test retest" OR "intraclass correlation*" OR "multitrait scaling analys*" OR "uncertainty measur*" OR "variability analys*" OR "variability value*" OR "minimal* important change" OR "minimal* important difference" OR "minimal* significant change" OR "minimal* significant difference" OR "minimal* detectable change" OR "minimal* detectable difference" OR "clinical* important change" OR "clinical* important difference" OR "clinical* significant change" OR "clinical* significant difference" OR "clinical* detectable change" OR "clinical* detectable difference" OR "small* real change" OR "small* real difference" OR "small* detectable change" OR "small* detectable difference") OR (SU.EXACT.EXPLODE("Measurement") OR SU.EXACT.EXPLODE("Error Analysis") OR SU.EXACT.EXPLODE("Test Construction") OR SU.EXACT.EXPLODE("Interrater Reliability") OR SU.EXACT.EXPLODE("Content Analysis") OR SU.EXACT.EXPLODE("Error of Measurement") OR SU.EXACT.EXPLODE("Factor Structure") OR SU.EXACT.EXPLODE("Testing Methods") OR SU.EXACT.EXPLODE("Statistical Reliability") OR SU.EXACT.EXPLODE("Consistency (Measurement)") OR SU.EXACT.EXPLODE("Computer Assisted Testing") OR SU.EXACT.EXPLODE("Factor Analysis") OR SU.EXACT.EXPLODE("Prediction") OR SU.EXACT.EXPLODE("Statistical Validity") OR SU.EXACT.EXPLODE("Prediction Errors")) (DE "Quality of Life" OR TI "Quality of Life" OR TI "QoL" OR TI "Life Quality" OR TI "utilit*" OR TI "preference-based*" OR TI "preference based" OR TI "wellbeing" OR TI "well-being" OR TI "HRQoL" OR TI "quality-adjusted life-years" OR AB "Quality of Life" OR AB "QoL" OR AB "Life Quality" OR AB "utilit*" OR AB "preference-based*" OR AB "preference based*" OR AB "wellbeing" OR AB "well-being" OR AB "HRQoL" OR AB "quality-adjusted life-years")	
Web of Science	((TS=("Alzheimer Disease" OR Alzheimer* OR Dement* OR "cognitive declin*" OR "cognitive dysfunct*" OR "cognitive impair*" OR "cognitive deterior*" OR "Home for the Aged" OR "Homes for the Aged" OR "Nursing home resident" OR "Nursing home residents" OR "community-dwelling" OR "community dwelling")) AND TS=("Health Survey" OR Questionnaire OR Survey* OR Questionnaires OR self-report* OR interview* OR care-giver* OR caregiver* OR carer* OR "patient-outcome assessment" OR proxy OR "Outcome Assessment" OR "outcome measure*" OR instrument OR "eq5d*" OR "eq-5d*" OR "demqol*" OR "ascot*" OR "icecap*" OR "qol-acc*" OR "qolacc*" OR "WOOP*" OR "qol-ad*" OR "qolad*" OR "dqi")) AND TS=(instrumentation OR methods OR "validation study" OR "comparative study" OR Psychometrics OR psychometr* OR clinimetr* OR clinometr* OR "Outcome Assessment, Health Care" OR "outcome assessment" OR "outcome measure*" OR "Observer Variation" OR "Health Status Indicators" OR "Reproducibility of Results" OR reproducib* OR "Discriminant Analysis" OR reliab* OR unreliab* OR valid* OR "coefficient of variation" OR coefficient OR homogeneity OR homogeneous OR "internal consistency" OR (cronbach* AND (alpha OR alphas)) OR (item AND (correlation* OR selection* OR reduction*)) OR agreement OR precision OR imprecision OR "precise values" OR test-retest OR (test AND retest) OR (reliab* AND (test OR retest)) OR stability OR interrater OR inter-rater OR intrarater OR intra-rater OR intertester OR inter-tester OR intratester OR intra-tester OR interobserver OR inter-observer OR intraobserver OR intra-observer OR intertechnician OR inter-technician OR intratechnician OR intra-technician OR interexaminer OR inter-examiner OR intraexaminer OR intra-examiner OR interassay OR inter-assay OR intraassay OR intra-assay OR interindividual OR inter-individual OR intraindividual OR intra-individual OR interparticipant OR inter-participant OR intraparticipant OR intra-participant OR kappa OR kappas OR repeatab* OR ((replicab* OR repeated) AND (measure OR measures OR findings OR result OR results OR test OR tests)) OR generaliza* OR generalisa* OR concordance OR (intraclass AND correlation*) OR discriminative OR "known group" OR "factor analysis" OR "factor analyses" OR "factor structure" OR "factor structures" OR dimension* OR subscale* OR (multitrait AND scaling AND (analysis OR analyses)) OR "item discriminant" OR "interscale correlation*" OR (error OR errors) OR "individual variability" OR "interval variability" OR "rate variability" OR (variability AND (analysis OR values)) OR (uncertainty AND (measurement OR measuring)) OR "standard error of measurement" OR sensitiv* OR responsive* OR (limit AND detection) OR "minimal detectable concentration" OR interpretab* OR ((minimal OR minimally OR clinical OR clinically) AND (important OR significant OR detectable) AND (change OR	11,959

	difference)) OR (small* AND (real OR detectable) AND (change OR difference)) OR “meaningful change” OR “ceiling effect” OR “floor effect” OR “item response model” OR IRT OR rasch OR “differential item functioning” OR DIF OR “computer adaptive testing” OR “item bank” OR “cross-cultural equivalence”) AND TS=("Quality of Life" OR QoL OR "Life Quality" OR "utility*" OR "preference-based" OR "preference based" OR wellbeing OR well-being OR HRQoL OR "quality-adjusted life-years") NOT TS=((address OR biography OR “case reports” OR comment OR directory OR editorial OR festschrift OR interview OR lecture OR “legal case” OR legislation OR letter OR news OR “newspaper article” OR “patient education handout” OR “popular work” OR congress OR “consensus development conference” OR “consensus development conference, NIH” OR “practice guideline”))	
--	--	--

Additional File 3: Definitions and questions for the nine psychometric properties according to the COSMIN manual, rephrased for this systematic review about preference-based measures in people with dementia (Mokkink et al., 2024).

<i>Psychometric property</i>	Definition and question asked to assess this property in this systematic review
<i>Content validity</i>	The degree to which the content of a preference-based measure is an adequate reflection of concepts relevant to people with dementia • <i>Does the preference-based measure adequately assess health and well-being?</i>
<i>Structural validity</i>	The degree to which the scores of a preference-based measure are an adequate reflection of the dimensionality of health and well-being. • <i>Are the scores of the preference-based measure an adequate reflection of the dimensions of health and well-being?</i>
<i>Internal consistency</i>	The degree of interrelatedness among the items of a preference-based measure. • <i>Do the items in the preference-based measure assess various aspects health and well-being?</i>
<i>Cross-cultural validity</i>	The degree to which the performance of the items on a translated or culturally adapted preference-based measure are an adequate reflection of the performance of the items of the original version of the measure. • <i>Is the performance of the items on a translated or culturally adapted version of the preference-based measure an adequate reflection of the performance of the items in the original version?</i>

<i>Reliability</i>	The degree to which the measurement is free from measurement error, the total variance in measurements is due to differences between people. • <i>Is there no measurement error with repeated measurement?</i>
<i>Measurement error</i>	The systematic and random error of someone's score that is not attributed to true changes in health and well-being. • <i>Are the systematic and random errors of someone's score not related to changes in health and well-being?</i>
<i>Criterion validity</i>	There is no 'gold-standard' for preference-based measures, therefore, criterion validity can only be established when a short version of a measure is compared to its original long version. • <i>Are the scores of the shorter preference-based measure an adequate reflection of its original longer version?</i>
<i>Hypothesis testing for construct validity</i>	The degree to which the scores of a preference-based measure are consistent with hypothesis (e.g. relationships to score of other measures or clinical outcome measures). This is based on the assumption that the measure validly measures health and well-being. • <i>Are the preference-based scores consistent with a priori hypotheses regarding expected mean differences between groups, expected correlations with scores on other measures or with demographical/clinical variables?</i>
<i>Responsiveness</i>	The ability of a preference-based measure to detect change over time in health and well-being. • <i>What is the ability of the preference-based measure to detect change over time in health and well-being?</i>

Additional File 4: Predefined hypotheses for assessing construct validity and responsiveness based on the COSMIN guideline (Mokkink et al., 2024).

Hypotheses for construct validity and responsiveness	COSMIN guideline and rationale	Hypotheses used in this systematic review
<i>Hypothesis 1</i>	<i>Target and comparator instrument measure the same construct. These instruments should correlate highly ($r \geq 0.7$).</i>	When comparing measures assessing the same construct, with similar measurement strategy (e.g. both generic measures of HR-QoL), there should be a high correlation between these measures ($r \geq 0.7$).
<i>Hypothesis 2</i>	<i>Target and comparator instrument measure the same construct, but with a slightly different measurement strategy ($r \geq 0.6$).</i>	When comparing measures assessing the same construct, but with a slightly different measurement strategy (e.g. a generic with a disease-specific instrument), there should be a high correlation between these measures ($r \geq 0.6$).
<i>Hypothesis 3</i>	<i>Target and comparator instrument measure a largely related construct, and should correlate quite highly ($r = 0.3-0.7$).</i>	When comparing measures assessing a largely related construct (e.g. different QoL constructs, or with clinical outcome measures), there should be a moderate to high correlation ($r = 0.3-0.7$).
<i>Hypothesis 4 (known-groups validity)</i>	<i>There should be a marked difference in score between relevant subgroups. Relevant subgroups should differ at least 10% in scores, or correlation ≥ 0.30 between relevant subgroups and score.</i>	There should be a marked difference in score between relevant subgroups (e.g. comparing self- and proxy-rated, comparing between disease severities). Relevant subgroups should differ at least 10% in scores, or correlation ≥ 0.30 between relevant subgroups and score.
<i>Hypothesis 5 (known-groups validity)</i>	<i>There should be a marked difference in score between relevant subgroups. Differences between relevant subgroups should have an effect size (e.g. Cohen's D, standardized response mean) of at least 0.20.</i>	There should be a marked difference in score between relevant subgroups (e.g. comparing self- and proxy-rated, comparing between disease severities). Differences between relevant subgroups should have an effect size (e.g. Cohen's D, standardized response mean) of at least 0.20.

Additional File 5: COSMIN worksheet for assessing individual studies: Quality assessment based on the Risk of Bias checklist and rating of evidence based on the Criteria for good measurement properties (Mokkink et al., 2024).

Reference	Preference-based measure	Psychometric property	Risk of Bias?	Quality assessment	Data extracted	Rating of evidence
Aguirre et al. (2016)	EQ-5D-5L	Internal consistency	No	High	Cronbach's alpha = [0.60 – 0.64], but assumption of unidimensionality is not reported	Indeterminate (?)
		Reliability	Yes, time interval between repeated measures was not stated	Low	Inter-rater reliability: [0.63 – 0.67], but no ICC/kappa calculated	Indeterminate (?)
		Construct validity	No	High	4 hypotheses tested: 4 rejected	Insufficient (-)
Ankri et al. (2003)	EQ-5D-3L	Reliability	Yes, time interval between repeated measures < 2 weeks	Low	Test-retest reliability: ICC = 0.74; Kappa = [0.34 – 0.59]; Inter-rater reliability: ICC = [0.41 – 0.54]; Kappa = [0.00 – 0.61]	Inconsistent (±)
		Construct validity	No	High	4 hypotheses tested: 4 rejected	Insufficient (-)
Bibi et al. (2022)	ICECAP-O	Reliability	Yes, inappropriate ICC calculation	Low	Test-retest reliability: ICC = [0.56 – 0.68]	Insufficient (-)
Coucill et al. (2001)	EQ-5D-3L	Reliability	Yes, unclear if patients were stable between repeated measures, and weighting scheme not described	Moderate	Inter-rater agreement: Kappa = [0.03 – 0.63]	Insufficient (-)
Diaz-Redondo et al. (2014)	EQ-5D-3L	Structural validity	Yes, exploratory factor analysis was done	Moderate	2 factors: functional area & subjective area	Indeterminate (?)
		Internal consistency	No	High	Cronbach's alpha = 0.64, but assumption of unidimensionality is not reported	Indeterminate (?)

		Reliability	Yes, time interval between repeated measures was not stated and model for ICC calculation is not described	Low	Inter-rater reliability: ICC = 0.72	Sufficient (+)
		Construct validity	No	High	6 hypotheses tested: 4 rejected, 2 confirmed	Insufficient (-)
Easton et al. (2018)	EQ-5D-5L	Construct validity	No	High	11 hypotheses tested: 6 rejected, 5 confirmed.	Insufficient (-)
	DEMQOL-U				6 hypotheses tested: 6 rejected.	Insufficient (-)
	DEMQOL-Proxy-U				6 hypotheses tested: 6 rejected.	Insufficient (-)
Engel et al. (2020)	EQ-5D-5L	Content validity	Yes, interviewers had limited experience and no clear description of questions in the topic guide	Moderate	Insufficient (-) relevance; Sufficient (+) comprehensiveness; Insufficient (-) comprehensibility	Inconsistent (±)
	AQoL-8D				Inconsistent (±) relevance; Inconsistent (±) comprehensiveness; Inconsistent (±) comprehensibility	Inconsistent (±)
	ICECAP-O				Inconsistent (±) relevance; Insufficient (-) comprehensiveness; Insufficient (-) comprehensibility	Inconsistent (±)
	ASCOT-SCT4				Insufficient (-) relevance; Sufficient (+) comprehensiveness; Insufficient (-) comprehensibility	Inconsistent (±)
	DEMQOL-U				Inconsistent (±) relevance; Insufficient (-) comprehensiveness; Insufficient (-) comprehensibility	Inconsistent (±)
	AD-5D				Inconsistent (±) relevance; Sufficient (+) comprehensiveness; Sufficient (+) comprehensibility	Inconsistent (±)

Griffiths et al. (2020)	EQ-5D-5L	Internal consistency	No	High	Cronbach's alpha = [0.62 – 0.69], but assumption of unidimensionality is not reported	Indeterminate (?)
		Reliability	Yes, unclear if patients were stable between repeated measures, and if measurement conditions were similar	Moderate	Inter-rater reliability: ICC = [0.00 – 0.32]	Insufficient (-)
		Construct validity	No	High	18 hypotheses tested: 14 rejected, 4 confirmed	Insufficient (-)
Hussain et al. (2025)	EQ-5D-5L	Reliability	Yes, time interval between repeated measures < 2 weeks	Low	Inter-rater reliability: ICC = [0.00 – 0.36]	Insufficient (-)
		Responsiveness	No	High	80 hypotheses tested: 58 rejected, 22 confirmed	Insufficient (-)
Hutchinson et al. (2024)	QOL-ACC	Reliability	Yes, inappropriate ICC calculation, with evidence of systematic changes	Very Low	Inter-rater reliability: ICC = [0.52 – 0.56]; Kappa = [-0.00 – 0.60]	Insufficient (-)
Karlawish et al. (2008) (2 studies)	EQ-5D-3L	Reliability	Yes, inappropriate ICC calculation	Low	Test-retest reliability: ICC = [0.58 – 0.63]	Insufficient (-)
		Construct validity	No	High	11 hypotheses tested: 3 rejected, 8 confirmed	Insufficient (-)
	HUI2	Reliability	Yes, inappropriate ICC calculation	Low	Test-retest reliability: ICC = [0.62 – 0.66]	Insufficient (-)
		Construct validity	No	High	11 hypotheses tested: 2 rejected, 9 confirmed	Sufficient (+)
Kavirajan et al. (2009)	HUI2	Construct validity	No	High	4 hypotheses tested: 1 rejected, 3 confirmed	Sufficient (+)
		Responsiveness	No	High	9 hypotheses tested: 2 rejected, 7 confirmed	Sufficient (+)
	HUI3	Construct validity	No	High	4 hypotheses tested: 1 rejected, 3 confirmed	Sufficient (+)

		Responsiveness	No	High	9 hypotheses tested: 3 rejected, 6 confirmed	Insufficient (-)
Kerner et al. (1998)	QWB	Construct validity	No	High	3 hypotheses tested: 1 rejected, 2 confirmed	Insufficient (-)
Kunz et al. (2010)	EQ-5D-3L	Reliability	Yes, time interval between repeated measures was not stated	Low	Inter-rater reliability: ICC = 0.48, Kappa = [0.25 – 0.45]	Insufficient (-)
		Construct validity	No	High	7 hypotheses tested: 2 rejected, 5 confirmed	Insufficient (-)
		Responsiveness	No	High	10 hypotheses tested: 7 rejected, 3 confirmed	Insufficient (-)
Lacey et al. (2015)	HUI3	Responsiveness	Yes, unclear description of statistical models	Moderate	33 hypotheses tested: 11 rejected, 22 confirmed	Insufficient (-)
Lay et al. (2023)	EQ-5D-5L	Content validity	No	High	Insufficient (-) comprehensibility	Insufficient (-)
Lay et al. (2024)	QOL-ACC	Content validity	Yes, unclear if interviewers were trained/experienced	Low	Insufficient (-) comprehensibility	Insufficient (-)
Makai et al. (2014)	ICECAP-O	Construct validity	No	High	6 hypotheses tested: 6 confirmed	Sufficient (+)
Martin et al. (2019)	EQ-5D-5L	Criterion validity	No	High	Spearman's rank correlation coefficient = [-0.49 – 0.28]	Insufficient (-)
		Construct validity	No	High	5 hypotheses tested: 5 rejected	Insufficient (-)
	DEMQOL-Proxy-U	Criterion validity	No	High	Spearman's rank correlation coefficient = [-0.46 – 0.13]	Insufficient (-)
		Construct validity	No	High	3 hypotheses tested: 3 rejected	Insufficient (-)
Michalowsky et al. (2020)	EQ-5D-3L	Internal consistency	Yes, implausible findings, suggesting methodological flaws	Low	Cronbach's alpha = 0.98, but assumption of unidimensionality is not reported	Indeterminate (?)
		Construct validity	No	High	9 hypotheses tested: 3 rejected, 6 confirmed	Insufficient (-)

Michalowsky et al. (2021)	EQ-5D-3L	Construct validity	No	High	14 hypotheses tested: 8 rejected, 6 confirmed	Insufficient (-)
	EQ-5D-5L				14 hypotheses tested: 9 rejected, 5 confirmed	Insufficient (-)
Michalowsky et al. (2022)	EQ-5D-3L	Construct validity	Yes, important methodological flaw	Low	24 hypotheses tested: 12 rejected, 12 confirmed	Insufficient (-)
	EQ-5D-5L				24 hypotheses tested: 14 rejected, 10 confirmed	Insufficient (-)
Naglie et al. (2006)	EQ-5D-3L	Reliability	Yes, unclear if patients were stable between repeated measures, and if measurement conditions were similar, and model for ICC calculation is not described	Moderate	Test-retest reliability: ICC = [0.71 – 0.79]	Sufficient (+)
		Construct validity	No	High	28 hypotheses tested: 14 rejected, 14 confirmed	Insufficient (-)
	HUI3	Reliability	Yes, unclear if patients were stable between repeated measures, and if measurement conditions were similar, and model for ICC calculation is not described	Moderate	Test-retest reliability: ICC = [0.25 – 0.81]	Inconsistent (±)
		Construct validity	No	High	32 hypotheses tested: 26 rejected, 6 confirmed	Insufficient (-)
	QWB	Reliability	Yes, unclear if patients were stable between repeated measures, and if measurement conditions were similar, and model for ICC calculation is not described	Moderate	Test-retest reliability: ICC = 0.70	Sufficient (+)

		Construct validity	No	High	32 hypotheses tested: 26 rejected, 6 confirmed	Insufficient (-)
Nyman et al. (2021)	ICECAP-O	Structural validity	Yes, inadequate sample size	Moderate	1 factor, but not enough information reported	Indeterminate (?)
		Construct validity	No	High	1 hypotheses tested: 1 confirmed.	Sufficient (+)
		Responsiveness	No	High	1 hypotheses tested: 1 confirmed	Sufficient (+)
Orgeta et al. (2015)	EQ-5D-3L	Reliability	Yes, time interval between repeated measures was not stated	Low	Inter-rater reliability: ICC = 0.34, Kappa = [0.09 – 0.49]	Insufficient (-)
Sanchez-Arenas et al. (2014)	EQ-5D-3L	Internal consistency	Yes, important methodological flaw	Low	Cronbach's alpha = [0.76 – 0.80], but this is implausible	Indeterminate (?)
		Construct validity	No	High	12 hypotheses tested: 7 rejected, 5 confirmed	Insufficient (-)
Sarabia-Cobo et al. (2017)	ICECAP-O	Structural validity	Yes, important methodological flaw	Very low	5 factors, but not enough information reported and this is implausible	Indeterminate (?)
		Internal consistency	No	High	Cronbach's alpha = 0.82, but assumption of unidimensionality is not met	Indeterminate (?)
		Construct validity	No	High	5 hypotheses tested: 5 confirmed	Sufficient (+)
Schölzel-Dorenbos et al. (2012)	DQI	Reliability	Yes, unclear if patients were stable between repeated measures	Moderate	Inter-rater reliability: Kappa < 0.70.	Insufficient (-)
		Construct validity	No	High	2 hypotheses tested: 2 rejected	Insufficient (-)
Silarova et al. (2023)	ASCOT-Proxy	Structural validity	Yes, exploratory factor analysis was done	Moderate	2 or 1 factor(s), but not enough information reported	Indeterminate (?)
		Internal consistency	No	High	Cronbach's alpha = 0.69	Insufficient (-)
		Construct validity	No	High	7 hypotheses tested: 7 confirmed	Sufficient (+)

Silberfeld et al. (2002)	EQ-5D-3L	Content validity	Yes, inappropriate method used for assessing comprehensiveness	Low	Insufficient (-) comprehensiveness	Insufficient (-)
	HUI2				Insufficient (-) comprehensiveness	Insufficient (-)
	HUI3				Insufficient (-) comprehensiveness	Insufficient (-)
	QWB				Insufficient (-) comprehensiveness	Insufficient (-)
Towers et al. (2025)	EQ-5D-5L	Internal consistency	Yes, important methodological flaw	Low	Cronbach's alpha = 0.74, but not enough information reported	Indeterminate (?)
		Construct validity	No	High	5 hypotheses tested: 1 rejected, 4 confirmed	Sufficient (+)
	ASCOT-Proxy	Internal consistency	Yes, important methodological flaw	Low	Cronbach's alpha = 0.83, but not enough information reported	Indeterminate (?)
		Construct validity	No	High	11 hypotheses tested: 11 confirmed	Sufficient (+)
	ICECAP-O	Internal consistency	Yes, important methodological flaw	Low	Cronbach's alpha = 0.81, but not enough information reported	Indeterminate (?)
		Construct validity	No	High	6 hypotheses tested: 6 confirmed	Sufficient (+)

Abbreviations: EQ-5D-5L: 5-level EQ-5D version; EQ-5D-3L: 3-level EQ-5D version; HUI2: Health Utilities Index Mark 2; HUI3: Health Utilities Index Mark 3; QWB: Quality of Well-Being Scale; AQoL-8D: Assessment of Quality of Life 8 Dimensions; QoL-ACC: Quality of Life Aged Care Consumers; ICECAP-O: ICEpop CAPability measure for Older people; ASCOT-SCT4: Adult Social Care Outcomes Toolkit Self-Completion; ASCOT-Proxy: Adult Social Care Outcomes Toolkit Proxy-Completion; DEMQOL-U: Dementia Quality of Life Measure Utility version; DEMQOL-Proxy-U: Dementia Quality of Life Measure Proxy Utility version; AD-5D: Alzheimer's Disease Five Dimensions; DQI: Dementia Quality of Life Instrument.

Additional File 6 (6a, 6b, 6c): COSMIN worksheet for assessing final ratings: Summarizing quality assessments and ratings of evidence per measure and psychometric property.

AF 6a:

		EQ-5D-3L (12 studies)							EQ-5D-5L (10 studies)		
Content validity	Relevance	NA							- Moderate		
	Comprehensiveness	- Low							+ Moderate		
	Comprehensibility	NA							- Moderate	- High	
	Summarized	- Low							± Moderate		
Structural validity	Individual	- Moderate							NA		
	Summarized	- Moderate									
Internal consistency	Individual	? High	? Low	? Low					? High	? High	? Low
	Summarized	? Moderate							? Moderate		
Cross-cultural validity		NA							NA		
Measurement invariance		NA							NA		
Reliability	Individual	+ L	± L	- M	- L	+ M	- L	- L	? Low	- Mod	- Low
	Summarized	± Low							- Low		
Measurement error		NA							NA		
Criterion validity	Individual	NA							- High		

	Summarized		- High														
Hypothesis testing for construct validity	Individual	- H	- H	- H	- H	- H	- H	- H	- L	- H	- H	- H	- H	- H	- L	- H	+ H
	Summarized	- High									- High						
Responsiveness	Individual	- High									- High						
	Summarized	- High									- High						

Summarized results for psychometric properties are shown, as well as rating of summarized result (+ = sufficient, ± = inconsistent, - = insufficient, ? = indeterminate) and quality of evidence (H = High, M = Moderate, L = Low, VL = Very Low) based on COSMIN criteria. NA = not assessed. Colors are based on summarized result (+ = green, ± = yellow, - = orange, ? = grey) and shadings are based on quality of evidence (low quality evidence has lighter shading, high quality evidence has darker shading). Abbreviations: EQ-5D-5L: 5-level EQ-5D version; EQ-5D-3L: 3-level EQ-5D version.

AF 6b:

		HUI2 (5 studies)	HUI3 (4 studies)	QWB (3 studies)	AQoL-8D (1 study)	QoL-ACC (2 studies)	
Content validity	Relevance	NA	NA	NA	± Moderate	NA	
	Comprehensiveness	- Low	- Low	- Low	± Moderate	NA	
	Comprehensibility	NA	NA	NA	± Moderate	- Low	
	Summarized	- Low	- Low	- Low	± Moderate	- Low	
Structural validity		NA	NA	NA	NA	NA	
Internal consistency		NA	NA	NA	NA	NA	
Cross-cultural validity		NA	NA	NA	NA	NA	
Measurement invariance		NA	NA	NA	NA	NA	
Reliability	Individual	- Low	± Moderate	+ Moderate	NA	- Very Low	
	Summarized	- Low	± Moderate	+ Moderate		- Very Low	
Measurement error		NA	NA	NA	NA	NA	
Criterion validity		NA	NA	NA	NA	NA	
Hypothesis testing for construct validity	Individual	+ High	+ High	- High	+ High	NA	NA
	Summarized	+ High	- High	- High	- High		
Responsiveness	Individual	+ High	- High	- High	NA	NA	NA
	Summarized	+ High	- High				

Summarized results for psychometric properties are shown, as well as rating of summarized result (+ = sufficient, ± = inconsistent, - = insufficient, ? = indeterminate) and quality of evidence (H = High, M = Moderate, L = Low, VL = Very Low) based on COSMIN criteria. NA = not assessed. Colors are based on summarized result (+ = green, ± = yellow, - = orange, ? = grey) and shadings are based on quality of evidence (low quality evidence has lighter shading, high quality evidence has darker shading). Abbreviations: HUI2: Health Utilities Index Mark 2; HUI3: Health Utilities Index Mark 3; QWB: Quality of Well-Being Scale; AQoL-8D: Assessment of Quality of Life 8 Dimensions; QoL-ACC: Quality of Life Aged Care Consumers.

AF 6c:

		ICECAP-O (6 studies)	ASCOT-SCT4 (1 study)	ASCOT-Proxy (2 studies)	DEMQOL-U (2 studies)	DEMQOL-Proxy-U (1 study)	AD-5D (1 study)	DQI (1 study)
Content validity	<i>Relevance</i>	± Moderate	- Moderate	NA	± Moderate	NA	± Moderate	NA
	<i>Comprehensiveness</i>	- Moderate	+ Moderate	NA	- Moderate	NA	+ Moderate	NA
	<i>Comprehensibility</i>	- Moderate	- Moderate	NA	- Moderate	NA	+ Moderate	NA
	Summarized	± Moderate	± Moderate	NA	± Moderate	NA	± Moderate	NA
Structural validity	Individual	? Mod	? VL	NA	? Moderate	NA	NA	NA
	Summarized	? Moderate			? Moderate			
Internal consistency	Individual	? High	? Low	NA	- High	? Low	NA	NA
	Summarized	? Moderate			? Moderate			
<i>Cross-cultural validity</i>		NA	NA	NA	NA	NA	NA	NA
<i>Measurement invariance</i>		NA	NA	NA	NA	NA	NA	NA
Reliability	Individual	- Low	NA	NA	NA	NA	NA	- Moderate
	Summarized	- Low						- Moderate
<i>Measurement error</i>		NA	NA	NA	NA	NA	NA	NA
Criterion validity	Individual	NA	NA	NA	NA	- High	NA	NA
	Summarized					- High		

Hypothesis testing for construct validity	Individual	+	+	+	+	NA	+	+	-	-	NA	-
		H	H	H	H		High	High	High	High		High
	Summarized	+				NA	+		-	-	NA	-
		High					High		High	High		High
Responsiveness	Individual	+				NA	NA		NA	NA	NA	
		High										
	Summarized	+				NA					NA	
		High										

Summarized results for psychometric properties are shown, as well as rating of summarized result (+ = sufficient, ± = inconsistent, - = insufficient, ? = indeterminate) and quality of evidence (H = High, M = Moderate, L = Low, VL = Very Low) based on COSMIN criteria. NA = not assessed. Colors are based on summarized result (+ = green, ± = yellow, - = orange, ? = grey) and shadings are based on quality of evidence (low quality evidence has lighter shading, high quality evidence has darker shading). Abbreviations: ICECAP-O: ICEpop CAPability measure for Older people; ASCOT-SCT4: Adult Social Care Outcomes Toolkit Self-Completion; ASCOT-Proxy: Adult Social Care Outcomes Toolkit Proxy-Completion; DEMQOL-U: Dementia Quality of Life Measure Utility version; DEMQOL-Proxy-U: Dementia Quality of Life Measure Proxy Utility version; AD-5D: Alzheimer's Disease Five Dimensions; DQI: Dementia Quality of Life Instrument.