Article:
Gamble, C.M., Sabico, S., Honvo, G. et al. (2026) Health-related quality of life instruments for older persons with osteoporosis: a WHO-BOHEG systematic review of measurement properties for use in clinical trials and routine practice. Age and Ageing, 55 (5). afag124. ISSN: 0002-0729

https://doi.org/10.1093/ageing/afag124

Age and Ageing 2026; **55:** afag124
https://doi.org/10.1093/ageing/afag124

SYSTEMATIC REVIEW

Health-related quality of life instruments for older persons with osteoporosis: a WHO-BOHEG systematic review of measurement properties for use in clinical trials and routine practice

Clair M. Gamble[1], Shaun Sabico[2], Germain Honvo[3], Salisu Abubakar[1], Jan R. Boehnke[1], Stijn De Baets[4], Nicola Veronese[5], Veerle Knoop[6], Cyrus Cooper[7], Maria Luisa Brandi[8], Nicholas C. Harvey[9], Radmila Matijevic[10], Charlotte Beaudart[11], Rosemary Ngozi[12], Tereza Hough[13], Claudia Campusano[14], Marise Lazaretti-Castro[15], Manju Chandran[16], Ambrish Mithal[17], Yousef Al-Saleh[18], Nasser Al-Daghri[19], Bess Dawson-Hughes[20], Nansa Burlet[21], Etienne Cavalier[22], Eugene McCloskey[23,24], Oliver Bruyere[25], Jean-Yves Reginster[26], Theresa Diaz[27], René Rizzoli[28], Jotheeswaran Amuthavalli Thiyagarajan[29]

[1]School of Health Sciences, University of Dundee, Dundee DD1 4HJ, United Kingdom
[2]Biochemistry, King Saud University, Riyadh 11451, Saudi Arabia
[3]Consultant to the Department of Sexual, Reproductive, Maternal, Child and Adolescent Health and Ageing, World Health Organization, Geneva 1211, Switzerland
[4]Ghent University Faculty of Sciences, 9000 Ghent, Flanders, Belgium
[5]Department of Internal Medicine,University of Palermo, Via del Vespro 133, 90127, Palermo, Sicily, Italy
[6]Vrije Universiteit Brussel, Pleinlaan 21050 Elsene, Brussel, Belgium
[7]MRC Lifecourse Epidemiology Centre, University of Southampton, SO16 6YD Southampton, United Kingdom
[8]Department of Endocrinology, Vita-Salute San Raffaele University, 20132 Milan, Lombardy, Italy
[9]MRC Lifecourse Epidemiology Unit, University of Southampton, SO16 6YD Southampton, United Kingdom
[10]Faculty of Medicine, University of Novi Sad, 21000 Novi Sad, Serbia
[11]Biomedical Sciences, University of Namur, B-5000 Namur, Belgium
[12]Department of Radiation Medicine, University of Nigeria, Ihe Nsukka 410001, Nsukka, Enugu, Nigeria
[13]National Osteoporosis Foundation of South Africa (NOFSA), Durbanville 7550, South Africa
[14]Department of Internal Medicine, Clínica Universidad de los Andes, Av Plaza 2501, Las Condes, Chile
[15]Metabolic Bone and Mineral Diseases Research Center, Federal University of Sao Paulo, São Paulo, CEP 01246904, Brazil
[16]Department of Endocrinology, Singapore General Hospital, Singapore 169608
[17]Endocrinology & Diabetes, Max Super Speciality Hospital, Dwarka, Delhi 110075, India
[18]Department of Medicine, Health Oasis Hospital, Riyadh 12451, Saudi Arabia
[19]Protein Research Chair, Biochemistry Department, King Saud University, Riyadh 11451, Saudi Arabia
[20]Jean Mayer Human Nutrition Research Center on Aging, Tufts University, Boston, MA 02111, Massachusetts, USA
[21]WHO Collaborating Centre for epidemiology of musculoskeletal health and ageing, University of Liege, 4000 Liège, Walloon Region, Belgium
[22]Department of Clinical Chemistry, University of Liege, 4000 Liège, Walloon Region, Belgium
[23]Mellanby Centre for Bone Research, The University of Sheffield, S10 2RX Sheffield, United Kingdom
[24]Department of Medicine, School of Clinical Sciences, Monash University, 3004 Melbourne, Victoria, Australia

[25]Department of Public Health, Epidemiology and Health Economics, University of Liege, 4000 Liège, Walloon Region, Belgium
[26]WHO Collaborating Center for Epidemiology of Musculoskeletal Health and Ageing, University of Liege, 4000 Liège, Walloon Region, Belgium
[27]Department of Maternal, Newborn, Child, Adolescent Health and Ageing, World Health Organization, Geneva 1211, Switzerland
[28]Division of Bone Diseases, Department of Internal Medicine Specialties, Geneva University Hospitals, Geneva 1205, Switzerland
[29]Science, Innovation, Monitoring Unit, Department of Sexual, Reproductive, Maternal, Child, Adolescent Health and Ageing, World Health Organization, 20 Avenue Appia, Geneva 1211, Switzerland

Address correspondence to: Jotheeswaran Amuthavalli Thiyagarajan, Science, Innovation, Monitoring Unit, Department of Sexual, Reproductive, Maternal, Child, Adolescent Health and Ageing, World Health Organization, 20 Avenue Appia, Geneva 1211, Switzerland. Email: amuthavallithiya@who.int

Abstract

***Background*:** Osteoporosis increases fracture risk in older adults and is associated with impaired health-related quality of life (HRQoL). Osteoporosis-specific HRQoL instruments are widely used, but their measurement performance in older populations has not been comprehensively synthesised.
***Objective*:** To systematically identify and synthesise development and validation studies of osteoporosis-specific HRQoL instruments for older adults, and to appraise their measurement properties using COSMIN criteria and a modified GRADE approach to inform instrument selection for clinical trials, routine care, and research.
***Design*:** Systematic review.
***Setting*:** Community-dwelling; Long-term care facility; Primary care facility.
***Subjects*:** Older persons aged 60 years or over.
***Methods*:** We searched Medline (Ovid), Embase, PsycINFO (Ovid), and AMED (Ovid) from inception to August 2024. We extracted and appraised evidence for COSMIN-defined measurement properties: content validity, structural validity, internal consistency, cross-cultural validity or measurement invariance, reliability, measurement error, criterion validity, hypothesis testing for construct validity, and responsiveness. Methodological quality was assessed using the COSMIN Risk of Bias checklist, measurement properties were rated against COSMIN criteria for good measurement properties, and certainty of evidence was graded using a modified GRADE approach.
***Results*:** Forty-three studies reported the development and/or validation of nine osteoporosis-specific HRQoL instruments; language and version adaptations resulted in 15 instrument variants. Content validity was the most prominent limitation: only OPTQoL and the English Mini-OQLQ demonstrated sufficient evidence for multiple content validity components, while most widely used instruments lacked adequate evidence on item relevance, comprehensiveness, or comprehensibility. Evidence for internal structure was limited, with structural validity sufficient for ECOS-16 and QoLOS-NVFX, insufficient for QUALEFFO-41, and indeterminate for most other instruments. Cross-cultural validity and measurement invariance were almost entirely unexamined. By contrast, reliability and hypothesis testing for construct validity were more consistently supported, often with moderate-to-high certainty. Measurement error, interpretability, and responsiveness were poorly reported across instruments.
***Conclusions*:** The evidence base supports purpose-driven selection of osteoporosis-specific HRQoL instruments rather than a single preferred instrument. Across included studies, ECOS-16 shows the most consistently supported measurement properties in older adults for longitudinal assessment, although important evidence gaps remain (notably measurement error and cross-cultural validity). When brevity is the primary feasibility constraint in routine care, the Mini-OQLQ may be considered; however, evidence for responsiveness is limited.

Keywords: osteoporosis; quality of life; measurement instruments; validity; reliability; responsiveness; ageing; older persons; systematic review

Key points

- No single instrument is fit-for-all: Evidence supports use-case–driven instrument selection rather than a universal choice.
- For routine care, Mini-OQLQ offers low respondent burden with strong reliability but limited evidence for responsiveness.
- For longitudinal applications, ECOS-16 has the most consistently supported measurement properties in older adults.

Introduction

Osteoporosis is a systemic skeletal disease characterised by low bone density and microarchitectural deterioration, leading to increased bone fragility and susceptibility to fracture [1]. Its burden extends beyond fractures. Health-related quality of life (HRQoL) can be reduced even before a fragility fracture occurs [2], and fractures are associated with further and often sustained decrements in HRQoL compared with osteoporosis without fracture [3].

HRQoL reflects how osteoporosis affects everyday life across physical, psychological, and social functioning [4, 5]. Pain, limited mobility, fear of falling, and loss of independence often drive this burden in ways that routine clinical endpoints do not fully capture. For that reason, HRQoL is now a core outcome in osteoporosis research and care, alongside fracture risk and bone density [6–8] Regulators and the WHO evidence-to-decision (EtD) framework increasingly rely on validated, condition-specific patient-reported outcomes to weigh benefits against harms and to support decisions that improve quality of life [9–11].

Selecting an HRQoL instrument is purpose-driven. Instruments intended for longitudinal evaluation in randomised trials, cohort follow-up, or clinical service audits should be responsive to change, minimise floor and ceiling effects, and use sufficiently granular response options to detect small but clinically important differences. By contrast, cross-sectional instruments used for population surveys or clinical screening should maximise discrimination across the severity spectrum and prioritise interpretability and feasibility, even if some floor or ceiling items are retained [12].

Despite a large literature on health-related quality of life (HRQoL) in osteoporosis, with more than 3000 publications since 1976, no comprehensive synthesis has evaluated osteoporosis-specific HRQoL instruments for older adults using contemporary COSMIN (COnsensus- based Standards for the selection of health Measurement INstruments) standards and the GRADE system (Grading of Recommendations Assessment, Development, and Evaluation). Existing reviews either predate these methods or assess only a subset of instruments [13, 14]. As a result, the comparative measurement performance of available tools in older populations remains unclear, limiting evidence-based instrument selection for clinical trials, routine care, and research.

This systematic review aims to close this gap by identifying all studies that developed or validated osteoporosis-specific HRQoL instruments for older adults, and by appraising their measurement properties and certainty of evidence [15]. The findings will inform purpose-driven instrument selection for clinical trials, routine care, and research, rather than produce a simple ranking.

Methods

Guidelines and protocol registration

The review followed the COSMIN guideline for systematic reviews of patient-reported outcome measures (PROMs) [15] and was reported according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis)-for OMIs (Outcome Measurement Instruments) [16]. The protocol was registered in PROSPERO as part of a broader review of QoL PROMs in older people with osteoporosis or sarcopenia (CRD42023458414). This manuscript reports osteoporosis instruments only.

Eligibility criteria

We included primary studies whose main objective was to develop and/or evaluate measurement properties of osteoporosis-specific QoL instruments in older adults. Studies were eligible if participants were aged $\geq$60 years, had a mean age $\geq$ 60 years, included at least 50% aged $\geq$60 years, or reported results separately for those aged $\geq$60 years. Osteoporosis was defined using the WHO criterion (bone mineral density T-score ≤ -2.5). Because validation studies often recruit from fracture services, we also included studies using a clinical definition of osteoporosis (e.g. low-trauma fragility fracture) when BMD data were unavailable [17]. Mixed osteoporosis and osteopenia samples were included only if participants meeting the WHO osteoporosis criterion were clearly defined or osteoporosis subgroup results were reported separately and could be extracted; otherwise, studies were excluded. We excluded studies focused on secondary osteoporosis (e.g. cancer-related osteoporosis, hypogonadism, or corticosteroid-induced osteoporosis), studies in which the target instrument was used only as a comparator, and non-English or non-French publications.

Information sources and search strategy

We searched MEDLINE, Embase, PsycINFO, and AMED (all via Ovid). Searches were conducted in June 2023 and updated in August 2024. We also screened reference lists of included articles and contacted members of the WHO Bone Health Expert Working Group to identify missing studies and relevant grey literature. Database-specific strategies combined controlled vocabulary and keywords for QoL, osteoporosis and older age, and measurement properties. We applied the validated measurement properties search filter developed by Terwee et al., as recommended by COSMIN [15, 18]. Full search strategies are provided in Appendix S1.

Study selection

Records were managed in Covidence software. After deduplication, pairs of reviewers independently screened titles and abstracts, then assessed full texts for eligibility (S.S./C.B. and S.B./V.K.). Disagreements were resolved by consensus.

Data extraction

A standardised extraction form (Microsoft Excel) was developed and pilot tested in five studies. Two reviewers extracted data independently and in duplicate (S.S./C.B. and S.B./V.K.). Extracted data included study, participant, and instrument characteristics, together with all results

required to evaluate measurement properties. These included participant characteristics, setting, instrument used, scoring method, sample size, study purpose and design, follow-up duration, missing data, use of multiple imputation, and information on feasibility and interpretability.

Risk of bias assessment

Methodological quality was assessed using the COSMIN Risk of Bias checklist [12, 19, 20]. Two reviewers independently rated each measurement property in each study as very good, adequate, doubtful, or inadequate using the COSMIN worst score counts principle. Disagreements were resolved by consensus and checked by additional reviewers, with unresolved issues discussed with a senior reviewer.

Assessment of measurement properties

For each study, results were judged against COSMIN criteria for good measurement properties and classified as sufficient, insufficient, or indeterminate [12, 21]. Because no hypotheses were specified a priori for criterion validity, hypothesis testing for construct validity, or responsiveness, these properties were interpreted using the original authors' stated hypotheses. Evidence was synthesised by instrument and measurement property using qualitative summaries rather than meta-analysis because of heterogeneity in designs, metrics, and statistical approaches, and because some instruments were evaluated in few studies. Synthesised results were then rated as sufficient (+), insufficient (−), inconsistent (±), or indeterminate (?). Subgroup analyses were not undertaken for inconsistent results because few studies were available per instrument.

Certainty of evidence

Certainty for each synthesised result was assessed using the COSMIN modified GRADE approach [12] and rated as high, moderate, low, or very low based on risk of bias, inconsistency, indirectness, and imprecision. Certainty was not graded for inconsistent or indeterminate results.

Feasibility and interpretability

We extracted and summarised feasibility and interpretability information for each instrument, following COSMIN guidance [12]. Feasibility included available languages and translations, and licensing or copyright requirements; interpretability included cut-off scores.

Results

Study selection

Electronic database searches identified 2744 records. After removal of duplicates, 2415 titles and abstracts were screened, and 107 full-text articles were assessed for eligibility. Seventy-three full texts were excluded. The full list of excluded studies and reasons is available in the Open Science Framework deposit (OSF, https://osf.io/u3h8q/). Manual searching identified nine additional eligible studies. In total, 43 studies were included in the review [22–64]. The PRISMA flow diagram is presented in Fig. 1.

Characteristics of included instruments

Nine core osteoporosis-specific health-related quality of life (HRQoL) instruments were developed and/or validated in older populations: QUALEFFO, ECOS-16, JOQOL, OPAQ, OPTQoL, OQLQ, QoLOS-NVFX, QUALIOST, and the Italian Triple-Q. All instruments were patient-reported outcome measures designed for use in clinical encounters. Most were self-administered, although some were interviewer-administered or could be completed with interviewer support.

Several instruments were adapted into shortened forms and or translated and validated in additional languages, resulting in 15 instrument variants with measurement property evidence within the included studies. Table 1 summarises the characteristics, administration features, domains, scoring, and availability of the included osteoporosis-specific HRQoL instruments.

Characteristics of included studies

Study characteristics are summarised in Appendix S2 (Table S3) [22–64]. Twenty-four studies were conducted in high-income countries, 18 in upper-middle-income countries, and one in a lower-middle-income country (Morocco). Several studies were conducted across more than one country, predominantly in high-income settings. Studies varied in sample size (30 to 1486 participants) and central age estimates (mean or median 59.9 to 76.2 years). Most studies recruited community-dwelling participants (37 studies), and most samples included women only (36 studies). The largest body of evidence related to QUALEFFO-41 (13 studies; n = 2301), followed by ECOS-16 (6 studies; n = 1303) and QUALEFFO-31 (5 studies; n = 832).

Methodological quality and risk of bias

Methodological quality and reporting practices varied considerably (full COSMIN Risk of Bias ratings are provided in Appendix S3). Content validity was frequently not rateable because studies did not report sufficient detail on how patients and professionals informed item generation, item refinement, and comprehensibility testing. Structural validity was often rated as inadequate because factor analytic evaluations were absent, limited, or insufficiently reported. Internal consistency was commonly reported using appropriate statistics, but ratings were frequently constrained by the lack of supporting structural validity evidence required to justify subscale scoring. Reliability and hypothesis testing for construct validity were more often rated as adequate or very good, although reporting gaps remained. Cross-cultural validity and measurement invariance were not evaluated because included studies did not test invariance across

Table 1. Characteristics of osteoporosis-specific HRQoL instruments

S.no	PROM	PROM full name	Year of develop.	Administration (mode of administration, length of administration)	Description of the scale (items, domains)	Description of scoring and recall time	Availability†	Ref.
1	ECOS-16	Short osteoporosis QoL questionnaire	*2000*	*Mode of administration:* self-administration *Length of administration:* 10 minutes	*Number of items:* 16 *Description of domains:* 4 domains: Pain (5 items); Physical functioning (5 items): Fear of illness (2 items); Psychosocial functioning (4 items)	*Rating:* Likert scale *Scoring:* Total score ranges from 1 (best HRQoL) to 5 (worst HRQoL). Subscale domain score and overall score. *Recall time:* last week for some items	*Costs:* free *Languages*: validated in 5 languages (Persian, Turkish, English, Italian, Korean)	[23, 35, 42, 49, 54, 60]
2	JOQOL	The Japanese Osteoporosis Quality of Life Questionnaire	*2001*	*Mode of administration:* self-administration *Length of administration:* NR	*Number of items:* 38 *Description of domains:* 6 domains: pain, activities of daily living, social activity and leisure, general health, postural awareness, psychological factors, and falls	*Rating:* Scale graded from 0 to 4 *Scoring:* 0–152, total score converted into percentage (higher score = higher HRQoL) *Recall time:* None	*Costs:* free *Languages*: Japanese	[45, 76]
3	OPAQ1	Osteoporosis Assessment Questionnaire version 1	*1998*	*Mode of administration:* self-administration or oral interview *Length of administration:* 30–45 minutes	Number of items: 84 *Description of domains:* 18 domains distributed into four dimensions: physical function, psychological status, symptoms, and social interaction.	Rating: 5-point Likert scale (1 = no impairment, 5 = constant impairment) *Scoring:* scoring algorithm, higher value = better health status *Recall time:* 4 weeks	*Costs:* free *Languages*: English, Brazilian Portuguese	[26, 52, 56]
4	OPAQ2	Osteoporosis Assessment Questionnaire version 2	*2014*	*Mode of administration:* self-administration *Length of administration:* 15–20 minutes	Number of items: 60 questions which includes 54 questions grouped into 14 domains, plus six additional questions on general health, overall HRQOL, change in HRQOL over the last year, and residential status. *Description of domains:* 14 domains distributed into four dimensions: physical function, psychological status, symptoms, and social interaction.	*Rating:* 5-point Likert scale (1 = no impairment, 5 = constant impairment) *Scoring:* scoring algorithm, higher value = better health status *Recall time:* 2 weeks	*Costs:* free *Languages*: English	[55]
5	OPAQ- SV	Osteoporosis Assessment Questionnaire Short Version	*2016*	*Mode of administration:* self-administration *Length of administration:* 15–20 minutes	*Number of items:* 34 *Description of domains:* three dimensions: physical function, emotional status, symptoms	*Rating:* 5-point Likert scale (1 = no impairment, 5 = constant impairment) *Scoring:* scoring algorithm, higher value = better health status *Recall time:* None	*Costs:* free *Languages*: Chinese	[63]

(continued)

Table 1. Continued.

S.no	PROM	PROM full name	Year of develop.	Administration (mode of administration, length of administration)	Description of the scale (items, domains)	Description of scoring and recall time	Availability†	Ref.
6	OPAQ-M	Osteoporosis Assessment Questionnaire for Men	*2012*	*Mode of administration:* self-administration *Length of administration:* NR	*Number of items:* 39 *Description of domains:* three dimensions: physical function, emotional status, symptoms	*Rating:* 5-point Likert scale (1 = no impairment, 5 = constant impairment) *Scoring:* scoring algorithm, higher value = better health status *Recall time:* 1 week	*Costs:* free *Languages*: English	[56]
7	OPTQoL	Osteoporosis-Targeted Quality of Life	*1997*	*Mode of administration:* self-administration *Length of administration:* NR	*Number of items: 37 Description of domains: Four domains:* Physical difficulty (10 items). Adaptations (10 items), Fears (7 items), Health-related (10 items)	*Rating:* Likert scale *Scoring:* 1–10 *Recall time: None*	*Costs:* free *Languages*: English	[27, 46]
8	OQLQ	Osteoporosis Quality of Life Questionnaire	*1997*	*Mode of administration:* interview *Length of administration:* 20 minutes	*Number of items:* 30 *Description of domains:* Five domains: symptoms, physical function, activities of daily living, emotional function, and leisure	*Rating*: 7-points Likert scale (7 = best possible function, 1 = worst possible function) *Scoring:* scoring algorithm *Recall time:* last 2 weeks	*Costs:* not mentioned *Languages*: English, German	[24, 28, 36]
9	Mini-OQLQ	Mini-Osteoporosis Quality of Life Questionnaire	*1999*	*Mode of administration:* self-administration *Length of administration:* 3 minutes	*Number of items:* 10 *Description of domains:* Five domains: symptoms, physical function, activities of daily living, emotional function, and leisure	*Rating:* 7-points Likert scale (7 = best possible function, 1 = worst possible function) *Scoring:* scoring algorithm *Recall time:* last 2 weeks	*Costs:* not mentioned *Languages*: English, Turkish	[29, 61]
10	QoLOS-NVFX	Quality of Life Osteoporosis Scale-Nonvertebral Fractures	2021	*Mode of administration:* self-administration *Length of administration:* 10 minutes	*Number of items:* 23 *Description of domains:* NR	*Rating*: 3-point Likert scale from 1 = insufficient or never, 2 = good or sometimes, and 3 = excellent or every day *Scoring:* 23–69, higher score, higher quality of life *Recall time:* None	*Costs:* not mentioned *Languages*: Italian	[62]
11	QUALEFFO-41	The 41-item Quality of Life Questionnaire of the European Foundation for Osteoporosis	*1996*	*Mode of administration:* self-administration or in presence of an interviewer *Length of administration:* 20 minutes	*Number of items:* 41 *Description of domains:* Five domains: pain, physical function, social function, general health perception, and mental function	*Rating:* Likert scale, five response option (0–4 points) *Scoring:* scoring algorithm (most five response options. Exceptions are items 27, 28 and 29 with four responses (0–3 points). Items 23, 24, 25 and 26 with 3 response options (0–2 points))—total 100 points *Recall time:* last week for some questions	*Costs:* Available upon request from IOF *Languages*: Malay, Arabic, Serbian, Portuguese, Danish, Turkish, Iranian, Korean, Polish, Mexican, English, French, German, Italian, Swedish and Dutch	[22, 24, 25, 30, 37, 39, 43, 45, 50, 51, 53, 57, 59]

(continued)

Table 1. Continued.

S.no	PROM	PROM full name	Year of develop.	Administration (mode of administration, length of administration)	Description of the scale (items, domains)	Description of scoring and recall time	Availability†	Ref.
12	QUALEFFO-31	The 31-item Quality of Life Questionnaire of the European Foundation for Osteoporosis	*2006*	*Mode of administration:* self-administration *Length of administration:* NR	*Number of items: 31* *Description of domains:* Three domains: pain, physical function, mental state	*Rating:* Likert scale, 5-point ordinal *Scoring:* scoring algorithm—total 100 points *Recall time:* last week for some questions	*Costs:* Available upon request from IOF *Languages*: Spanish, Taiwanese Chinese, Chinese, Turkish	[32, 34, 40, 44, 64]
13	QUALEFFO-26	The simplified 26-item Quality of Life Questionnaire of the European Foundation for Osteoporosis		*Mode of administration:* self-administration *Length of administration:* NR	*Number of items: 26* *Description of domains:* Three domains: pain, physical function, mental state	*Rating:* Likert scale, 5-point ordinal *Scoring:* scoring algorithm - total 100 points *Recall time:* last week for some questions	*Costs:* Available upon request from IOF *Languages*: Korean	[41]
14	QUALIOST	Quality of Life Questionnaire in Osteoporosis	*2001*	*Mode of administration:* self-administration *Length of administration:* NR	*Number of items:* 23 *Description of domains:* two dimensions (physical and emotional)	*Rating:* 0–100, 100 = highest impairments *Scoring:* total score + score per dimension *Recall time:* last four weeks	*Costs:* free *Languages*: French, English, Turkish, Polish, Spanish, Italian	[31, 33, 48]
15	Triple-Q questionnaire	Triple-Q questionnaire	*2010*	*Mode of administration:* self-administration *Length of administration:* 15 minutes	*Number of items: 25* *Description of domains:* Five domains: back pain, functional status, social and recreational activities, subject's overall assessment of her general health, mental status	*Rating:* 4–23 for the pain domain, 5–39 for the physical domain, 0–15 for the social domain, 4–16 for overall health perception, 5–20 for the mental domain, and 18–113 for all domains. *Scoring*: The answers to questions 1, 2, 6, 7, 11, 12, 15, 16, 19, 21, 23 and 24 were scored from 1 to 4, where 4 represents the worst possible impairment and 1 the best possible function. The remaining answers were scored from 1 to 5, in which 5 represents the worst function and 1 the best one. Items for which the answer was 'not applicable' were scored 0. *Recall time* depends on the items	*Costs:* free *Languages*: Italian	[47]

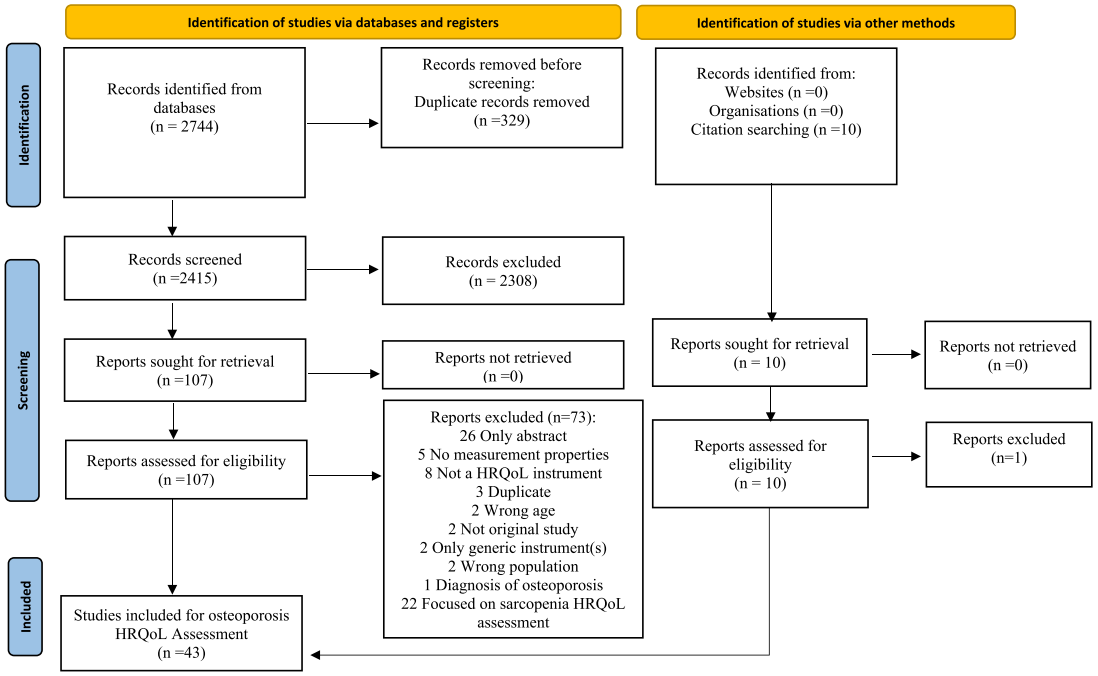

Figure 1. Flow diagram of study selection following PRISMA guidelines.

groups. Criterion validity was not assessed because there is no widely accepted reference standard for osteoporosis-specific HRQoL. Responsiveness was infrequently evaluated.

Overview of measurement property evidence

Across the 15 instrument variants, most COSMIN ratings were indeterminate because evidence was missing, incompletely reported, or not aligned with COSMIN evaluation requirements. Summary of findings for measurement properties and certainty of evidence are presented in Table 2. Detailed results are presented in Appendix S4. The following section provides a consolidated summary of the measurement properties of the included instruments.

Content validity

Only a small subset of instrument variants had any content validity components rated sufficient. OPTQoL had sufficient evidence for relevance and comprehensiveness, and the Mini-OQLQ had sufficient evidence for comprehensiveness and comprehensibility. In contrast, OPAQ1, OQLQ, and QoLOS-NVFX had insufficient ratings across content validity components. For the remaining instruments, content validity was rated indeterminate because studies did not provide enough information to judge item relevance, comprehensiveness, and comprehensibility using COSMIN criteria.

Internal structure

Evidence for internal structure was limited. Structural validity was rated sufficient for ECOS-16 and QoLOS-NVFX, and insufficient for QUALEFFO-41. For most remaining instruments, structural validity was indeterminate because appropriate analyses were not performed or not reported in sufficient detail. Internal consistency estimates were available for 13 of 15 instrument variants. ECOS-16, QoLOS-NVFX, and QUALEFFO-31 were rated sufficient overall, while OPAQ2 was rated insufficient. The remaining instruments were rated indeterminate, most commonly because internal consistency was reported without clear structural validity evidence confirming that relevant subscales were unidimensional. No study evaluated measurement invariance or cross-cultural validity for any instrument variant, despite multiple translated versions being used across settings.

Reliability and measurement error

Reliability was assessed for 13 of 15 instrument variants. Ten instrument variants demonstrated sufficient reliability, typically based on test–retest designs with acceptable intraclass correlation coefficients. Instrument variants rated indeterminate for reliability generally lacked COSMIN-preferred statistics or did not report suitable test–retest results. All instrument variants were rated indeterminate for measurement error. Across studies, the statistics needed to quantify measurement error and support interpretation of change, such as standard error of measurement, smallest detectable change, or limits of agreement, were rarely reported. This limits the ability to determine whether observed score changes reflect true clinical change or measurement variability.

Construct validity and responsiveness Construct validity evidence was the most consistently reported property across instruments. Most studies assessed convergent validity by hypothesis testing against generic HRQoL measures and osteoporosis-related constructs, most commonly SF-36 [65]

Table 2. Summary of findings for measurement properties of osteoporosis specific QoL instruments for older persons

Quality of the evidence for measurement properties of the PROMS			ECOS-16 Studies (N = 6) Population (N = 1303)	JOQOL Studies (N = 2) Population (N = 699)	OPAQ1 Studies (N = 2) Population (N = 70)	OPAQ2 Studies (N=) Population (N = 1477)	OPAQ-M Studies (N = 1) Population (N = 37)	OPAQ-SV Studies (N = 1) Population (N = 469)	OPTQoL Studies (N = 2) Population (N = 665)	OQLQ Studies (N = 3) Population (N = 968)	Mini-OQLQ Studies (N = 2) Population (N = 290)	QoLOS-NVFX Studies (N = 1) Population (N = 458)	QUALEFFO-41 Studies (N = 13) Population (N = 2301)	QUALEFFO-31 Studies (N = 5) Population (N = 832)	QUALEFFO-26 Studies (N = 1) Population (N = 127)	QUALIOST Studies (N = 3) Population (N = 1736)	Triple-Q Studies (N = 1) Population (N = 200)
Content validity	**Relevance**	Rating summarised result	(?)	(?)	(−)	(?)	(?)	(?)	(+)	(−)	(?)	(−)	(?)	(?)	(?)	(?)	(?)
		Quality of evidence	No GRADE	No GRADE	Low	No GRADE	No GRADE	No GRADE	High	Low	No GRADE	Low	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE
	Comprehensiveness	Rating summarised result	(?)	(?)	(−)	(?)	(?)	(?)	(+)	(−)	(+)	(−)	(?)	(?)	(?)	(?)	(?)
		Quality of evidence	No GRADE	No GRADE	Low	No GRADE	No GRADE	No GRADE	High	Low	Low	Low	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE
	Comprehensibility	Rating summarised result	(?)	(?)	(−)	(?)	(?)	(?)	(?)	(−)	(+)	(−)	(?)	(?)	(?)	(?)	(?)
		Quality of evidence	No GRADE	No GRADE	Low	No GRADE	No GRADE	No GRADE	No GRADE	Low	Low	Low	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE
Structural validity		Rating summarised result	(+)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(+)	(−)	(?)	(?)	(?)	(?)
		Quality of evidence	High	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	Moderate	Moderate	No GRADE	No GRADE	No GRADE	No GRADE
Internal consistency		Rating summarised result	(+)	(?)	(?)	(−)	(?)	(?)	(?)	(?)	(?)	(+)	(?)	(+)	(?)	(?)	(?)
		Quality of evidence	High	No GRADE	No GRADE	Low	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	Moderate	No GRADE	Low	No GRADE	No GRADE	No GRADE
Cross-cultural validity		Rating summarised result	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)
		Quality of evidence	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE
Measurement invariance		Rating summarised result	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)
		Quality of evidence	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE
Reliability		Rating summarised result	(+)	(?)	(?)	(?)	(+)	(+)	(+)	(+)	(+)	(?)	(+)	(+)	(+)	(+)	(?)
		Quality of evidence	High	No GRADE	No GRADE	No GRADE	Moderate	Moderate	Moderate	High	High	No GRADE	High	High	Moderate	High	No GRADE
Measurement error		Rating summarised result	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)
		Quality of evidence	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE
Criterion validity		Rating summarised result	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)	(?)
		Quality of evidence	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE
Hypothesis testing for construct validity		Rating summarised result	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)
		Quality of evidence	High	Moderate	High	Moderate	High	High	High	High	High	Moderate	High	High	Moderate	High	Moderate
Responsiveness		Rating summarised result	(+)	(+)	(?)	(+)	(?)	(?)	(?)	(?)	(−)	(?)	(?)	(?)	(?)	(?)	(?)
		Quality of evidence	Low	Moderate	No GRADE	Moderate	No GRADE	No GRADE	No GRADE	No GRADE	Low	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE	No GRADE

Measurement property rating: sufficient (+), insufficient (−), inconsistent (±), indeterminate (?); NR = No data/results reported in the included studies; No GRADE = No GRADE assessment performed.

and EQ-5D [66]. Overall, hypothesis testing for construct validity was rated sufficient across all instrument variants, with moderate to high certainty for many instruments.

Responsiveness was rated sufficient for ECOS-16, JOQOL, and OPAQ2, insufficient for the Mini-OQLQ, and indeterminate for the remaining instrument variants. Indeterminate ratings primarily reflected the absence of longitudinal designs suitable for responsiveness assessment or insufficient reporting of a priori hypotheses and change metrics.

Certainty of evidence

Certainty of evidence varied across instruments and properties. Evidence was most often graded as moderate or high for reliability and hypothesis testing for construct validity, reflecting repeated evaluations using appropriate methods. In contrast, certainty grading was frequently not possible for content validity, structural validity for most instruments, cross-cultural validity and measurement invariance, and measurement error because results were predominantly indeterminate.

Feasibility and interpretability

Feasibility and interpretability information is summarised in Appendix S5. Most instruments were self-administered, with administration time ranging from approximately 2 to 40 minutes when reported. Availability and licensing varied from free access to permission or purchase requirements. Interpretability information was consistently sparse. Studies rarely reported reference values, cut-off scores, or thresholds for meaningful within-person change. In combination with the absence of measurement error statistics, this constrains clinical interpretation of absolute scores and score change over time.

Discussion

This review synthesised development and validation studies of osteoporosis-specific HRQoL instruments in older adults and appraised measurement properties using COSMIN with certainty judged by a modified GRADE approach. Across 43 studies and 15 instrument variants, the main pattern was not poor performance, but missing or methodologically misaligned evidence, leading to many indeterminate ratings.

Summary of principal findings

Content validity was the most important limitation. Only OPTQoL and the English Mini-OQLQ had sufficient evidence for multiple COSMIN content validity components. For OPTQoL, evidence supported relevance and comprehensiveness, while comprehensibility remained uncertain. For the Mini-OQLQ, evidence supported comprehensiveness and comprehensibility, while relevance was uncertain. Several widely used instruments, including QUALEFFO variants and QUALIOST, had insufficient reporting to judge whether items are relevant, comprehensive, and comprehensible for older adults with osteoporosis. This matters because content validity is foundational. If the item set does not capture what older people consider important, then strong reliability or expected correlations with generic measures cannot fully compensate.

Evidence for internal structure was also limited. Structural validity was sufficient for ECOS-16 and QoLOS-NVFX, but insufficient for QUALEFFO-41, and indeterminate for most other instruments because dimensionality was not evaluated or not reported in sufficient detail. Internal consistency was frequently reported, but often without the structural validity evidence needed to justify interpreting domain scores. This pattern suggests that subscale interpretation has often outpaced the evidence needed to support it.

Cross-cultural validity and measurement invariance were almost entirely absent, despite extensive translation and use of several instruments across countries and languages. In the absence of invariance testing, scores from different language versions should not be assumed comparable, which is a critical constraint for multi-country trials and pooled analyses.

By contrast, reliability and hypothesis testing for construct validity were more commonly supported, often with moderate to high certainty. Responsiveness was demonstrated for a small subset of instruments but remained under-evaluated overall. Finally, measurement error and interpretability were major evidence gaps. Key indices such as standard error of measurement, smallest detectable change, and limits of agreement were rarely reported, and minimal important change thresholds and score interpretation aids were uncommon. These gaps constrain the use of osteoporosis-specific HRQoL instruments for monitoring change over time and for making clinically interpretable judgements at the individual level.

Implications for selecting instruments in clinical trials, routine care, and research

These findings support a purpose-driven approach to instrument selection rather than a search for a single 'best' osteoporosis-specific HRQoL tool. Instruments optimised for longitudinal detection of change differ from those optimised for cross-sectional discrimination and description, and selection should follow the intended inference and practical constraints of the setting. This logic is consistent with broader guidance on quality-of-life instrument selection, which emphasises matching the item set and scaling properties to study purpose and population [65].

The following statements summarise where the available evidence is comparatively stronger for specific use cases; they are provided to support context-specific instrument selection and should not be interpreted as formal recommendations or endorsements.

For longitudinal applications such as clinical trials and monitoring, ECOS-16 may be considered a pragmatic option, as it has the most consistently supported evidence profile in older adults. It is the only instrument with high certainty evidence of sufficient structural validity, internal consistency, reliability, and hypothesis testing for construct validity, with

additional evidence of responsiveness albeit with low certainty. JOQOL and OPAQ2 also show sufficient responsiveness with moderate certainty, but both have important limitations for use as primary longitudinal endpoints. JOQOL has indeterminate reliability, and OPAQ2 has insufficient internal consistency, raising concerns about score precision and coherence of domain scoring. The Mini-OQLQ has high certainty evidence for reliability, but insufficient responsiveness, and therefore is not well supported for detecting change over time.

In routine care, feasibility and interpretability often determine whether a measure can be used at all. Where a brief, low-burden cross-sectional assessment is the priority, the Mini-OQLQ may be considered, given its short administration time and supportive evidence for reliability and construct validity. However, limited evidence for responsiveness constrains its use for monitoring change over time. Where routine practice requires follow-up comparisons, ECOS-16 may be considered, but small score changes should be interpreted cautiously until measurement error and meaningful change thresholds are established. For research that is primarily cross-sectional or correlational, instrument choice depends on whether the priority is defensible scoring structure or stronger evidence that the content reflects patient-important impacts. ECOS-16 has comparatively stronger evidence for internal structure, supporting more defensible use of total and domain scores. OPTQoL has comparatively stronger evidence for key aspects of content validity and may be preferred when breadth and relevance of patient-important impacts is the dominant concern, while recognising that evidence for structural validity and internal consistency remain indeterminate. Across all uses, the lack of evidence for cross-cultural validity and measurement invariance means that translated versions should be treated as distinct instruments unless equivalence is demonstrated. This is particularly important for multinational trials and for comparative work across settings.

Implications for future research

This review highlights several measurement properties that remain inadequately explored and that can be addressed with targeted and feasible methodological work.

First, robust evidence for content validity is scarce. This likely reflects limited stakeholder engagement during development and poor transparency in reporting, particularly for older instruments developed before contemporary standards. Future work should prioritise well-documented qualitative studies with older adults, including those with and without fragility fractures, with representation across cultures, literacy levels, and care settings. A parallel priority is improving theoretical clarity about dimensionality, because uncertainty about whether an instrument is unidimensional or multidimensional undermines interpretation of subscales and weakens the evidential basis for internal consistency. Updating widely used instruments using contemporary COSMIN-recommended approaches, including systematic stakeholder engagement and clear documentation of item generation and refinement, may be more efficient than creating new tools, especially where resources are constrained [12, 19, 66].

Second, validation research should explicitly strengthen theoretical underpinnings and content validity in low- and middle-income countries. This is important for equity and generalisability, given demographic change and the growing global burden of osteoporosis in older populations [67].

Third, cross-cultural validity and measurement invariance are essential for global health applications and multi-country trials yet were rarely evaluated. Comprehensive testing in every context is unrealistic, but pragmatic steps include rigorous translation and cultural adaptation protocols, initial invariance testing across key groups, and development of shared reference datasets to support comparability over time.

Fourth, methodological shortcomings often reduced the usefulness of reported results. Test–retest studies frequently lacked evidence of participant stability, and structural validity analyses did not always match the measurement model. Better adherence to COSMIN guidance would reduce indeterminate ratings and improve interpretability [12, 66]. Fifth, reanalysis of existing datasets could efficiently address gaps, including measurement invariance across language versions [68–70].

Sixth, future studies should treat measurement error, interpretability, and responsiveness as core priorities rather than optional extras. Measurement error indices such as standard error of measurement, smallest detectable change, and limits of agreement should be routinely reported. Responsiveness studies should pre-specify hypotheses, report effect sizes, and estimate meaningful change thresholds where possible. Interpretability reporting should include score distributions, floor and ceiling effects, missingness, and guidance for clinical meaning.

Finally, study populations should be broadened and better characterised. Most evidence concerned women only, and studies often combined participants with densitometric osteoporosis and those with fragility fractures. Future research should include older men and consider stratification by fracture status where feasible. The near absence of psychometric evidence in men remains a critical gap for generalisability and equity in assessment [71, 72].

Implications for clinical practice and public health policy

Clinicians and trialists may wish to prioritise instruments with demonstrated content validity when the goal is to represent patient experience. In this review, OPTQoL and the English Mini-OQLQ provided the clearest signal for key content validity components, but the evidence for their internal consistency and other measurement properties requires further confirmation.. In multilingual settings, instruments should not be assumed comparable across translations unless cross-cultural validity and measurement invariance have been demonstrated, or unless adaptation and validation follow rigorous, well-documented protocols.

A further consideration is the distinction between condition-specific and generic HRQoL instruments [73]. This review focused on condition-specific HRQoL instruments, i.e. those that allow a standardised assessment of the impact of osteoporosis (specifically) on physical, psychological, and social domains of life. Generic instruments in turn, are designed to be relevant to all conditions, patient groups, and interventions. When selecting HRQoL instruments for a specific context, condition-specific instruments may be better at providing a clinically embedded characterisation of HRQoL and its changes [74], while generic measures allow for comparisons across all conditions and are advantageous when for example health state utility values are a required [75].

For policy and guideline development, these findings argue for investment in strengthening and updating existing osteoporosis-specific HRQoL instruments rather than proliferation of new tools without full validation. Standardised reporting expectations, shared translation resources, and reference datasets could materially improve comparability and usefulness of HRQoL outcomes in research and practice, particularly in low- and middle-income settings.

Strengths and limitations

Strengths include adherence to COSMIN and PRISMA-COSMIN standards, duplicate screening and extraction, and use of modified GRADE to make certainty transparent. Limitations mainly reflect the underlying literature: many studies lacked the reporting needed for COSMIN evaluation, feasibility data were inconsistent, and restricting inclusion to English and French may have missed some validation studies.

Conclusion

This review supports purpose-driven, evidence-informed selection rather than a single preferred instrument. ECOS-16 has the most consistently supported evidence for longitudinal assessment in older adults, whereas the Mini-OQLQ may be useful for brief routine assessment and OPTQoL may be preferred when content validity is the main priority. Across instruments, limited evidence on measurement error, interpretability, and cross-cultural validity constrains interpretation of change and comparability across translations. These statements are intended as evidence-informed use-case guidance rather than prescriptive recommendations.

Acknowledgements: The views expressed in this paper are those of the authors and do not necessarily reflect the views of the institutions or organisations with which they are affiliated.

Supplementary Data: Supplementary data is available at *Age and Ageing* online.

Declaration of Conflicts of Interest: None.

Declaration of Sources of Funding: European Society for Clinical and Economic Aspects of Osteoporosis, Osteoarthritis and Musculoskeletal Diseases (ESCEO).

Data Availability: All data relevant to the study are included in the article or uploaded as supplementary information.

References

1. World Health Organization. Assessment of fracture risk and its application to screening for postmenopausal osteoporosis : report of a WHO study group [meeting held in Rome from 22 to 25 June 1992]: World Health Organization; 1994. Available from: https://iris.who.int/handle/10665/39142.
2. Hopman WM, Berger C, Joseph L *et al.* Longitudinal assessment of health-related quality of life in osteoporosis: data from the population-based Canadian multicentre osteoporosis study. *Osteoporos Int* 2019;**30**:1635–44. https://doi.org/10.1007/s00198-019-05000-y.
3. Gao S, Zhao Y. Quality of life in postmenopausal women with osteoporosis: a systematic review and meta-analysis. *Qual Life Res* 2023;**32**:1551–65. https://doi.org/10.1007/s11136-022-03281-1.
4. Mayo NE, Figueiredo S, Ahmed S *et al.* Montreal accord on patient-reported outcomes (PROs) use series - paper 2: terminology proposed to measure what matters in health. *J Clin Epidemiol* 2017;**89**:119–24. https://doi.org/10.1016/j.jclinepi.2017.04.013.
5. Costa DSJ, Mercieca-Bebber R, Rutherford C *et al.* How is quality of life defined and assessed in published research? *Qual Life Res* 2021;**30**:2109–21. https://doi.org/10.1007/s11136-021-02826-0.
6. World Health Organization; European Observatory on Health Systems and Policies; I Papanicolas *et al.* Eds. *Health System Performance Assessment: A Framework for Policy Analysis*: World Health Organization; 2022. Available from: https://iris.who.int/handle/10665/352686
7. WHO. *Constitution of the World Health Organization*. Geneva: World Health Organization; 1946. Available from: https://apps.who.int/gb/bd/pdf/bd47/en/constitution-en.pdf.
8. Geng D, Li X, Shi Y. Effect of exercise intervention on health-related quality of life in middle-aged and older people with osteoporosis: a systematic review and meta-analysis. *PeerJ.* 2025;**13**:e18889. https://doi.org/10.7717/peerj.18889.
9. Kaplan RM, Hays RD. Health-related quality of life measurement in public health. *Annu Rev Public Health* 2022;**43**:355–73. https://doi.org/10.1146/annurev-publhealth-052120-012811.
10. Higginson IJ, Carr AJ. Measuring quality of life: using quality of life measures in the clinical setting. *BMJ.* 2001;**322**:1297–300. https://doi.org/10.1136/bmj.322.7297.1297.
11. World Health Organization, Rob MPM, Adam T *et al. Making Choices in Health: WHO Guide to Cost-Effectiveness Analysis / Edited by T. Tan-Torres Edejer. [Et al]*. World Health Organization; 2003. Available from: https://iris.who.int/handle/10665/42699.
12. LB Mokkink, CAC Prinsen, DL Patrick, J Alonso, LM Bouter, de HCW Vet *et al.* COSMIN methodology for systematic

reviews of Patient-Reported Outcome Measures (PROMs): user manual 2018 [02/02/2023]. Available from: https://cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf.

13. Choo YW, Mohd Tahir NA, Mohamed Said MS *et al.* Health-related quality of life in osteoporosis: a systematic review of measurement properties of the QUALEFFO-41. *Osteoporos Int* 2024;**35**:745–57. https://doi.org/10.1007/s00198-023-07005-0.
14. Madureira MM, Ciconelli RM, Pereira RM. Quality of life measurements in patients with osteoporosis and fractures. *Clinics (Sao Paulo)* 2012;**67**:1315–20. https://doi.org/10.6061/clinics/2012(11)16.
15. Prinsen CAC, Mokkink LB, Bouter LM *et al.* COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;**27**:1147–57. https://doi.org/10.1007/s11136-018-1798-3.
16. Elsman EBM, Mokkink LB, Terwee CB *et al.* Guideline for reporting systematic reviews of outcome measurement instruments (OMIs): PRISMA-COSMIN for OMIs 2024. *J Clin Epidemiol* 2024;**173**:111422. https://doi.org/10.1016/j.jclinepi.2024.111422.
17. WHO. *Prevention and Management of Osteoporosis*. Geneva, Switzerland, 2003.
18. Terwee CB, Jansma EP, Riphagen II *et al.* Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res* 2009;**18**:1115–23. https://doi.org/10.1007/s11136-009-9528-5.
19. Mokkink LB, de Vet HCW, Prinsen CAC *et al.* COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;**27**:1171–9. https://doi.org/10.1007/s11136-017-1765-4.
20. Terwee CB, Prinsen CAC, Chiarotto A *et al.* COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res* 2018;**27**:1159–70. https://doi.org/10.1007/s11136-018-1829-0.
21. Mokkink LB, Boers M, van der Vleuten CPM *et al.* COSMIN risk of bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. *BMC Med Res Methodol* 2020;**20**:293. https://doi.org/10.1186/s12874-020-01179-5.
22. Azimi P, Shahzadi S, Azhari S *et al.* An outcome measure of functionality and quality of life in Iranian women with osteoporotic vertebral fractures: a validation study of the QUALEFFO-41. *J Orthop Sci* 2014;**19**:860–7. https://doi.org/10.1007/s00776-014-0609-0.
23. Badia X, Díez-Pérez A, Lahoz R *et al.* The ECOS-16 questionnaire for the evaluation of health related quality of life in postmenopausal women with osteoporosis. *Health Qual Life Outcomes* 2004;**2**:41. https://doi.org/10.1186/1477-7525-2-41.
24. Badia X, Díez-Pérez A, Alvarez-Sanz C *et al.* Measuring quality of life in women with vertebral fractures due to osteoporosis: a comparison of the OQLQ and QUALEFFO. *Qual Life Res* 2001;**10**:307–17. https://doi.org/10.1023/a:1012200508847.
25. Bergland A, Thorsen H, Kåresen R. Association between generic and disease-specific quality of life questionnaires and mobility and balance among women with osteoporosis and vertebral fractures. *Aging Clin Exp Res* 2011;**23**:296–303. https://doi.org/10.1007/bf03324967.
26. Cantarelli FB, Szejnfeld VL, Oliveira LM *et al.* Quality of life in patients with osteoporosis fractures: cultural adaptation, reliability and validity of the osteoporosis assessment questionnaire. *Clin Exp Rheumatol* 1999;**17**:547–51.
27. Chandler JM, Martin AR, Girman C *et al.* Reliability of an osteoporosis-targeted quality of life survey instrument for use in the community: OPTQoL. *Osteoporos Int* 1998;**8**:127–35. https://doi.org/10.1007/bf02672508.
28. Osteoporosis Quality of Life Study G. Measuring quality of life in women with osteoporosis. *Osteoporos Int* 1997;**7**:478–87. https://doi.org/10.1007/PL00004151.
29. Cook DJ, Guyatt GH, Adachi JD *et al.* Development and validation of the mini-osteoporosis quality of life questionnaire (OQLQ) in osteoporotic women with back pain due to vertebral fractures. Osteoporosis quality of life study group. *Osteoporos Int* 1999;**10**:207–13. https://doi.org/10.1007/s001980050217.
30. de Oliveira FN, Arthuso M, da Silva RB *et al.* Validation of the Portuguese version of the quality of life questionnaire of the European foundation for osteoporosis (QUALEFFO-41) in Brazilian women with postmenopausal osteoporosis with vertebral fracture. *Clin Rheumatol* 2013;**32**:1585–92. https://doi.org/10.1007/s10067-013-2265-8.
31. de la Loge C, Sullivan K, Pinkney R *et al.* Cross-cultural validation and analysis of responsiveness of the QUALIOST®: QUAlity of life questionnaire In OSTeoporosis. *Health Qual Life Outcomes* 2005;**3**:69. https://doi.org/10.1186/1477-7525-3-69.
32. González Matarín PJ, Martínez-Amat A, Lomas-Vega R *et al.* Validation of the quality of life questionnaire of the European Foundation for Osteoporosis-31 in Spanish postmenopausal women. *Menopause.* 2014;**21**:469–76. https://doi.org/10.1097/GME.0b013e3182a6cc64.
33. Hepguler S, Atamaz FC, Pinar Y *et al.* Acceptability, validity and reliability of the Turkish QUALIOST(®) in postmenopausal women with osteoporosis. *Rheumatol Int* 2013;**33**:757–61. https://doi.org/10.1007/s00296-012-2443-z.
34. Ince B, Kucukakkas O. The reliability and validity of the Turkish version of the quality of life questionnaire of the European Foundation for Osteoporosis-31 (QUALEFFO-31). *Arch Osteoporos* 2021;**16**:128. https://doi.org/10.1007/s11657-021-00997-4.
35. Karadeniz PG, Bekiroglu N, Ercan I *et al.* Application of Rasch analysis to Turkish version of ECOS-16 questionnaire. *J Appl Meas* 2014;**15**:302–12.
36. Kerschan-Schindl K, Patsch J, Kudlacek S *et al.* Measuring quality of life with the German osteoporosis quality of life questionnaire in women with osteoporosis. *Wien Klin Wochenschr* 2012;**124**:532–7. https://doi.org/10.1007/s00508-012-0212-3.
37. Koçyigit H, Gülseren S, Erol A *et al.* The reliability and validity of the Turkish version of quality of life questionnaire of the European Foundation for osteoporosis (QUALEFFO). *Clin Rheumatol* 2003;**22**:18–23. https://doi.org/10.1007/s10067-002-0653-6.
38. Kumamoto K, Nakamura T, Suzuki T *et al.* Validation of the Japanese osteoporosis quality of life questionnaire. *J Bone Miner Metab* 2010;**28**:1–7. https://doi.org/10.1007/s00774-009-0125-z.
39. Lai PSM, Chua SS, Chan SP *et al.* Validation of the English version of the quality of life questionnaire of the

European Foundation for osteoporosis (QUALEFFO) in Malaysia. *Int J Rheum Dis* 2008;**11**:421–9. https://doi.org/10.1111/j.1756-185X.2008.00402.x.

40. Lai BM, Tsang SW, Lam CL *et al.* Validation of the quality of life questionnaire of the European Foundation for osteoporosis (QUALEFFO-31) in Chinese. *Clin Rheumatol* 2010;**29**:965–72. https://doi.org/10.1007/s10067-010-1495-2.
41. Lee JS, Shin JK, Son SM *et al.* Validation of the quality-of-life questionnaire of the European Foundation for osteoporosis (QUALEFFO-26) in Korean population. *Rheumatol Int* 2014;**34**:919–27. https://doi.org/10.1007/s00296-013-2942-6.
42. Lee JS, Son SM, Goh TS *et al.* Validation of the ECOS-16 questionnaire in Koreans with osteoporosis. *Asian Spine J* 2016;**10**:877–85. https://doi.org/10.4184/asj.2016.10.5.877.
43. Lee YK, Kim HJ, Park JW *et al.* Transcultural adaptation and psychometric properties of the Korean version of the quality of life questionnaire of the European Foundation for osteoporosis (QUALEFFO-41). *Arch Osteoporos* 2019;**14**:96. https://doi.org/10.1007/s11657-019-0647-5.
44. Li W-C, Chen Y-C, Yang R-S *et al.* Taiwanese Chinese translation and validation of the quality of life questionnaire of the European Foundation for osteoporosis 31 (QUALEFFO-31). *J Formos Med Assoc* 2013;**112**:621–9. https://doi.org/10.1016/j.jfma.2012.09.015.
45. Lips P, Cooper C, Agnusdei D *et al.* Quality of life in patients with vertebral fractures: validation of the quality of life questionnaire of the European Foundation for osteoporosis (QUALEFFO). Working Party for Quality of life of the European Foundation for osteoporosis. *Osteoporos Int* 1999;**10**:150–60. https://doi.org/10.1007/s001980050210.
46. Lydick E, Zimmerman SI, Yawn B *et al.* Development and validation of a discriminative quality of life questionnaire for osteoporosis (the OPTQoL). *J Bone Miner Res* 1997;**12**:456–63. https://doi.org/10.1359/jbmr.1997.12.3.456.
47. Maggio D, Ruggiero C, Ercolani S *et al.* A multi-dimensional questionnaire quantifying quality of life in elderly osteoporotic women: the Italian triple-Q osteoporosis study. *Aging Clin Exp Res* 2010;**22**:330–9. https://doi.org/10.1007/BF03324937.
48. Marquis P, Cialdella P, De la Loge C. Development and validation of a specific quality of life module in post-menopausal women with osteoporosis: the QUALIOST. *Qual Life Res* 2001;**10**:555–66. https://doi.org/10.1023/a:1013041206433.
49. Moradzadeh R, Moghimi N, Nadrian H *et al.* Validity and reliability of the Farsi version of the ECOS-16 questionnaire for females with osteoporosis. *East Mediterr Health J* 2018;**23**:729–33. https://doi.org/10.26719/2017.23.11.729.
50. Nagammai T, Mohazmi M, Liew SM *et al.* Validation of the Malay version of the quality of life questionnaire of the European Foundation for osteoporosis (QUALEFFO-41) in Malaysia. *Qual Life Res* 2015;**24**:2031–7. https://doi.org/10.1007/s11136-015-0933-7.
51. Ramírez Pérez E, Clark P, Wacher NH *et al.* Cultural adaptation and validation of the quality of life questionnaire of the European Foundation for osteoporosis (QUALEFFO) in a Mexican population. *Clin Rheumatol* 2008;**27**:151–61. https://doi.org/10.1007/s10067-007-0661-7.
52. Randell AG, Bhalerao N, Nguyen TV *et al.* Quality of life in osteoporosis: reliability, consistency, and validity of the osteoporosis assessment questionnaire. *J Rheumatol* 1998;**25**:1171–9.
53. Rostom S, Allali F, Bahiri R *et al.* Psychometric properties evaluation of the quality of life questionnaire of the European Foundation for osteoporosis in Arabic population. *Rheumatol Int* 2012;**32**:2037–49. https://doi.org/10.1007/s00296-011-1910-2.
54. Salaffi F, Malavolta N, Cimmino MA *et al.* Validity and reliability of the Italian version of the ECOS-16 questionnaire in postmenopausal women with prevalent vertebral fractures due to osteoporosis. *Clin Exp Rheumatol* 2007;**25**:390–403.
55. Shen W, Burge R, Naegeli AN *et al.* Psychometric properties of the osteoporosis assessment questionnaire (OPAQ) 2.0: results from the multiple outcomes of raloxifene evaluation (MORE) study. *BMC Musculoskelet Disord* 2014;**15**:374. https://doi.org/10.1186/1471-2474-15-374.
56. Solimeo SL, Silverman SL, Calderon AD *et al.* Measuring health-related quality of life (HRQOL) in osteoporotic males using the male OPAQ. *Osteoporos Int* 2012;**23**:841–52. https://doi.org/10.1007/s00198-011-1625-y.
57. Tadic I, Vujasinovic Stupar N, Tasic L *et al.* Validation of the osteoporosis quality of life questionnaire QUALEFFO-41 for the Serbian population. *Health Qual Life Outcomes* 2012;**10**:74. https://doi.org/10.1186/1477-7525-10-74.
58. Urushihara H, Yoh K, Hamaya E *et al.* Responsiveness of the Japanese osteoporosis quality of life questionnaire in women with postmenopausal osteoporosis. *Health Qual Life Outcomes* 2014;**12**:178. https://doi.org/10.1186/s12955-014-0178-0.
59. Yilmaz F, Doğu B, Sahin F *et al.* Investigation of responsiveness indices of generic and specific measures of health related quality of life in patients with osteoporosis. *J Back Musculoskelet Rehabil* 2014;**27**:391–7. https://doi.org/10.3233/bmr-140459.
60. Yilmaz F, Dogu B, Sahin F *et al.* Reliability and validity of the Turkish version of the ECOS 16 questionnaire in postmenopausal osteoporosis. *Eur J Phys Rehabil Med* 2009;**45**:521–6.
61. Yilmaz F, Sahin F, Dogu B *et al.* Reliability and validity of the Turkish version of the mini osteoporosis quality of life questionnaire. *J Back Musculoskelet Rehabil* 2012;**25**:89–94. https://doi.org/10.3233/bmr-2012-0314.
62. Basilici Zannetti E, 'Angelo D D, Cittadini N *et al.* Development and testing of the quality of life osteoporosis scale-nonvertebral fractures (QoLOS-NVFX). *Orthop Nurs* 2021;**40**:33–41. https://doi.org/10.1097/nor.0000000000000728.
63. Zhang YP, Wei HH, Wang W *et al.* Cross-cultural adaptation and validation of the osteoporosis assessment questionnaire short version (OPAQ-SV) for Chinese osteoporotic fracture females. *Clin Rheumatol* 2016;**35**:1003–10. https://doi.org/10.1007/s10067-015-3010-2.
64. Zhou C, Li Q, Huang S *et al.* Validation of the simplified Chinese version of the quality of life questionnaire of the European foundation for osteoporosis (QUALEFFO-31). *Eur Spine J* 2016;**25**:318–24. https://doi.org/10.1007/s00586-015-4066-z.
65. Hyland ME. A brief guide to the selection of quality of life instrument. *Health Qual Life Outcomes* 2003;**1**:24. https://doi.org/10.1186/1477-7525-1-24.
66. Gagnier JJ, Lai J, Mokkink LB *et al.* COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. *Qual Life Res* 2021;**30**:2197–218. https://doi.org/10.1007/s11136-021-02822-4.

67. WHO. Fact sheet: Ageing and health 2024 [21/07/2025]. Available from: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health.
68. Zumbo BD. Three generations of DIF analyses: considering where it has been, where it is now, and where it is going. *Lang Assess Q* 2007;**4**:223–33. https://doi.org/10.1080/15434300701375832.
69. Leitgob H, Seddig D, Asparouhov T *et al.* Measurement invariance in the social sciences: historical development, methodological challenges, state of the art, and future perspectives. *Soc Sci Res* 2023;**110**:102805. https://doi.org/10.1016/j.ssresearch.2022.102805.
70. Chua KC, Bohnke JR, Prince M *et al.* Health-related quality-of-life assessment in dementia: evidence of cross-cultural validity in Latin America. *Psychol Assess* 2019;**31**:1264–77. https://doi.org/10.1037/pas0000743.
71. Borgstrom F, Karlsson L, Ortsater G *et al.* Fragility fractures in Europe: burden, management and opportunities. *Arch Osteoporos* 2020;**15**:59. https://doi.org/10.1007/s11657-020-0706-y.
72. Adler RA. Osteoporosis in men: A review. *Bone Res* 2014;**2**:14001. https://doi.org/10.1038/boneres.2014.1.
73. Brazier JE, Rowen D, Mavranezouli I *et al.* Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). *Health Technol Assess* 2012;**16**:1–114. https://doi.org/10.3310/hta16320.
74. Beaudart C, Biver E, Bruyère O *et al.* Quality of life assessment in musculo-skeletal health. *Aging Clin Exp Res* 2018;**30**:413–8. https://doi.org/10.1007/s40520-017-0794-8.
75. Ara R, Brazier J, Peasgood T *et al.* The identification, review and synthesis of health state utility values from the literature. *PharmacoEconomics.* 2017;**35**:43–55. https://doi.org/10.1007/s40273-017-0547-8.
76. Tomaszewski KA, Henry BM, Paradowski J *et al.* Cross cultural adaptation of the English version of the IOF-QLQ to polish, to assess the health-related quality-of-life of patients after a distal radius fracture. *Health Qual Life Outcomes* 2015;**13**:158. https://doi.org/10.1186/s12955-015-0354-x.

Received 21 August 2025; accepted 16 March 2026