BMJ Open

Evaluation of the psychometric properties of patient-reported outcome measures of health-related quality of life across the European cancer continuum: a systematic review protocol using COSMIN methodology

Laurence Leysen ![ORCID],[1] Karla Marticic Giljevic,[1] Claire Piccinin,[1] Morena Shkodra,[2] Madeline Pe,[1] Morten Petersen,[3] Giovanni Apolone,[2] Cinzia Brunelli ![ORCID],[2] Claudio Lombardo,[4] Galina Velikova,[5] Alexandra Gilbert,[5] Gabriella Pravettoni,[6,7] Gennaro Ciliberto,[8] Mogens Groenvold,[9] Montserrat Ferrer,[10,11] Nanne Bos,[12] Ricardo Pietrobon,[13] Norbert Couespel,[14] Augusto Caraceni,[15] Marianne Guren,[16] Aude Sirven,[17] Hugo Vachon ![ORCID],[1] on behalf of the EUonQOL consortium

Correspondence to
Hugo Vachon;
hugo.vachon@eortc.org

ABSTRACT

Introduction Over the past decades, there has been increasing recognition that assessing patients with cancer's health-related quality of life (HRQoL) is pivotal to delivering optimal patient-centred healthcare. However, with the increasing number of patient-reported outcome measures (PROMs) available, it becomes more and more challenging to identify the most appropriate PROM to capture HRQoL. Therefore, the aim of this systematic review is to (1) identify all available PROMs assessing HRQoL across the European cancer continuum and (2) critically appraise, compare and summarise the psychometric properties of the identified PROMs.

Methods and analysis Bibliographic databases MEDLINE and PubMed Central (through PubMed) and EMBASE (through Scopus) will be comprehensively searched from database inception until March 2024. Studies reporting on the measurement properties of PROMs assessing HRQoL throughout the European cancer continuum will be included. The evaluation of the psychometric properties, data extraction and data synthesis will be conducted according to the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) methodology. Two reviewers will independently assess the methodological quality using the COSMIN risk of bias checklist and the COSMIN criteria for good measurement properties. Subsequently, findings will be qualitatively summarised. The Grading of Recommendations Assessment, Development and Evaluations (GRADE) guidelines will be used to grade and summarise the quality of the evidence.

Ethics and dissemination Ethical clearance for this research is not required, as the systematic review will only use information from previously published research. The results of this review will be submitted for publication in a peer-reviewed journal and will be used to provide a set of evidence-based recommendations for a European project (EUonQOL), which aims at developing a new PROM (EUonQOL toolkit) to assess HRQoL across the European cancer continuum. Moreover, findings will be disseminated to a clinical audience and policymakers through conferences, supporting researchers and clinicians in choosing the best measure to evaluate HRQoL in patients with cancer and survivors in Europe.

PROSPERO registration number CRD42023418616.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This systematic review will report on the psychometric properties of the patient-reported outcome measures (PROMs) used for the assessment of health-related quality of life across the European cancer continuum at a subscale level.

⇒ This systematic review will follow the highest methodological standards, that is, the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) Protocols, the Consensus-based Standards for the selection of health Measurement Instruments and the PRISMA 2020 statement.

⇒ Due to the expected heterogeneity in study samples and reporting of PROMs' psychometric properties within the literature, no quantitative pooling of data (ie, meta-analysis) will be performed and the findings will be summarised qualitatively.

⇒ The systematic review will specifically include PROMs validated in European patients with cancer and survivors; thus, PROMs which have only been validated outside of Europe will not be covered.

⇒ The systematic review will only consider articles published in English, which may introduce an information bias and limit the comprehensiveness of the results.

INTRODUCTION

Health-related quality of life (HRQoL) can be defined as 'how well a person functions in their life and his or her perceived well-being in physical, mental and social domains of health'.[1] Functioning refers here to a patient's ability to carry out some pre-defined activities and well-being to their subjective feelings.[1] More specifically, the framework developed by Wilson and Cleary, which is currently the most applied theoretical model of HRQoL,[2] conceives HRQoL as a multidimensional construct encompassing five components: symptom status, functional status, biological and psychological variables, general health perceptions and overall quality of life.

Over the past decades, there has been increasing recognition that assessing patients with cancer's HRQoL is pivotal to delivering optimal patient-centred healthcare.[3 4] HRQoL is now perceived as a meaningful endpoint throughout the cancer continuum[5 6] and can serve as a valuable source of information to guide healthcare policies (eg, Europe's Beating Cancer plan[7]). However, HRQoL is often inaccurately assessed by healthcare providers and poorly captured by medical procedures or tests, highlighting the need for patient involvement in reporting their outcomes.[3 4 8 9]

Patient-reported outcomes (PROs) are defined by the Food and Drug Administration as 'a measurement based on a report that comes directly from the patient about the status of a patient's health condition, without amendment or interpretation of the patient's response by a clinician or anyone else'.[10] Patient-reported outcome measures (PROMs) refer to the tools used to measure PROs and are now systematically used for the assessment of HRQoL in cancer care. To assess the HRQoL of patients with cancer, a wide array of PROMs is currently available, ranging from generic (eg, 36-Item Short-Form Survey Instrument [SF-36], 5-level EQ-5D [EQ-5D-5L]) to cancer-specific (eg, EORTC Core Quality of Life questionnaire [QLQ-C30], Functional Assessment of Cancer Therapy - General [FACT-G]) and tumour-specific tools (eg, EORTC Breast Cancer Module [QLQ-BR23], Functional Assessment of Cancer Therapy - Breast [FACT-B]). However, this diversity made it more and more challenging to select the most appropriate PROM. This choice should be made with regard to the target population, the target construct and, importantly, the PROM measurement properties.[11]

Over the past years, many systematic reviews comparing PROMs for the assessment of HRQoL in patients with cancer have been published. Most of them focused on PROMs measuring HRQoL in a specific type of cancer (eg, breast cancer, prostate cancer)[12–23] or cancer population (eg, cancer survivors, advanced cancer, palliative patients).[14 24–26] Several of these reviews focused on PROMs evaluating one specific HRQoL-related construct (eg, depression, fatigue, pain),[12 13 27–29] and the majority did not report the psychometric properties of the PROMs under investigation per subscale.[13–17 19–22 24 25 27 28 30] For the reviews reporting on the psychometric properties of PROMs, the methods used to assess both the quality of the studies and the results differed significantly.[31]

Currently, the highest methodological standards for the conduct of systematic reviews on the psychometric properties of PROMS are provided by the COnsensus-based Standards for the selection of health Measurement INstruments initiative (COSMIN[32]). However, among the reviews published to date, only half relied on the COSMIN methodology, and most of them did not apply it fully. For instance, in several reviews, the rating of the overall results per PROM was unclear or not performed,[12 16 20 27 33] and the risk of bias assessment or the grading of the evidence was not conducted.[12 13 24 27 30 33] As such, a comprehensive overview of the psychometric properties of PROMs used for the assessment of HRQoL across the cancer continuum is still needed and missing. Therefore, this study aims to systematically review the measurement properties of PROMs assessing the multidimensional construct of HRQoL in European patients with cancer and survivors to make objective recommendations on the most suitable PROM to use in these populations.

METHODS AND ANALYSIS

The protocol of this systematic review is based on the PRISMA-P (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols) guidelines[34] and has been prospectively registered in the International Prospective Register of Systematic Reviews database (PROSPERO 2023, CRD42023418616). In case of protocol amendments, modifications will be reported in the publication reporting the results of the systematic review as supplementary material.

The systematic review will be conducted according to the COSMIN guidelines for systematic reviews[32] and will use the COSMIN taxonomy of measurement properties (table 1). All steps of the screening process will be performed using Rayyan.[35]

Search strategy

A systematic search will be performed in the bibliographic databases MEDLINE and PubMed Central (through PubMed) and EMBASE (through Scopus) without a publication date restriction up to February 2023 (updated up to March 2024). The search strategy will be based on the Population Intervention Comparator acronym[36] (PICOM) in which the population will be represented by patients with cancer and survivors, the outcome by health-related quality of life and the methods by psychometric properties. No comparator or intervention will be used. Both Medical Subject Headings (MeSH) terms and text words will be used.

Original research articles published in English (including erratum and correction articles) will be considered for inclusion. Reference lists of included articles will be manually searched by hand to ensure all relevant studies will be considered. Additionally, the exclusion

Table 1 COSMIN definitions of measurement properties

Measurement property	Definition
Content validity	The degree to which a PROM measures the construct(s) it purports to measure
Structural validity	The degree to which the scores of a PROM are an adequate reflection of the dimensionality of the construct to be measured
Internal consistency	The degree of inter-relatedness among the items
Cross-cultural validity	The degree to which the performance of the items on a translated or culturally adapted PROM are an adequate reflection of the performance of the items of the original version of the PROM
Measurement invariance	The proportion of the total variance in the measurements which is due to 'true' differences between patients
Reliability	The degree to which the measurement is free from measurement error
Reliability (extended definition)	The extent to which scores for patients who have not changed are the same for repeated measurement under several conditions: eg, using different sets of items for the same PROM (internal consistency); over time (test–retest); by different persons on the same occasion (inter-rater); or by the same persons (ie, raters or responders) on different occasions (intra-rater)
Measurement error	The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured
Construct validity	The degree to which the scores of a PROM are consistent with hypotheses (for instance with regard to internal relationships, relationships to scores of other instruments or differences between relevant groups) based on the assumption that the PROM validly measures the construct to be measured
Responsiveness	The ability of a PROM to detect change over time in the construct to be measured

COSMIN, Consensus-based Standards for the selection of health Measurement Instruments; PROM, patient-reported outcome measure.

filter of Terwee et al[37] will be used. The grey literature will not be considered.

The respective search strategies that will be used for PubMed and Scopus are provided in online supplemental appendix 1.

Selection process
The selection process will be twofold. First, it will be determined whether the PROMs captured by the search should be included or excluded. Second, all titles and abstracts will be screened for eligibility in a blinded standardised manner. If the study seems relevant or in case of doubt, the full-text article will be retrieved and screened. Both the abstract and full-text screening will be done independently by a minimum of two reviewers. For both steps, a pilot screening will be performed on a random subsample of studies and the screening methodology will be clarified within the review team if deemed necessary. Discrepancies will be resolved by discussion and/or consultation of a third reviewer. Inter-rater reliability will be assessed and reported.

PROM selection
To be included, PROMs will need to meet the following criteria:
1. PROMs must be self-administered questionnaires (paper-pencil or electronic). This excludes assessment tools based (fully or partially) on hetero-assessment, interactive voice response systems, talking touch screens, drawings or nomograms. An interview format is allowed when the study population is not able to complete the PROM independently.
2. PROMs must assess HRQoL as a multi-domain concept (ie, based on a multidimensional model of HRQoL) and be applicable across cancer types. This excludes tools designed to assess a specific HRQoL subdomain (eg, exclusively assessing physical functioning) or cancer site (eg, assessing HRQoL following breast reconstruction). Preference-based measures that are used to calculate quality-adjusted life years within the field of health economics will not be considered for the scope of this review.
3. PROMs must be validated (ie, evidence of content validity, structural validity or construct validity) for use in the target population of European patients with cancer or survivors (online supplemental appendix 2). In case no European validation can be found for a PROM identified through the initial search, an additional search will be performed in PubMed (online supplemental appendix 3). If no evidence of validity among European patients with cancer or survivors can be retrieved after the additional search, the PROM and its related articles will be excluded.

Study selection
Studies will be included when the following criteria are met:
1. Studies must provide information on the measurement properties of the included PROMs. For this review, the development, content validity, structural validity/

unidimensionality, internal consistency, cross-cultural validity, measurement invariance, reliability, measurement error and construct validity will be considered. Studies reporting on criterion validity will be considered to inform construct validity due to the absence of a gold standard for PROMs.[32] Responsiveness will not be assessed in this review since the content and the number of hypotheses to assess responsiveness are inexhaustible and arbitrary, and the quality of comparator instruments (in the absence of gold standard) cannot be proven.[38]

2. Studies must provide original research data (including erratum and correction articles) and be published in English. Articles written in other languages or case studies, protocols, conference abstracts, conference reports, commentaries, opinion articles and reviews will not be considered.

3. Studies must be performed in adult European patients with cancer or survivors (mean age ≥21 years and not defined as adolescents and young adults). Articles including 'mixed samples' (ie, European patients with cancer and patients without cancer) will only be included if separate results are provided for the patients with cancer group. Studies involving both European and non-European patients with cancer will be included. Studies only reporting results within a non-European cancer sample will be excluded (except for development and content validity studies). Articles reporting on patients with benign tumours or including less than 15 patients with cancer will also be excluded.

Detailed information on the selection process will be reported in a PRISMA flowchart (PRISMA 2020 flow diagram[39]).

Data extraction

During the data extraction, it will be determined which measurement properties will be evaluated for every included study. Extracted data will be entered into a customised xls file using Microsoft Excel. Data extraction will be performed independently by two reviewers, and discrepancies will be resolved by discussion and/or consultation of a third reviewer. Data extraction will be piloted on a random subsample of studies, and the extraction methodology will be clarified within the review team if deemed necessary. When available, data will be extracted as follows:

1. Study characteristics: authors, title, publication year, design.

2. Study sample characteristics: sample size, age, gender, EU/non-EU, clinical status (general population, patients without cancer, patients with cancer undergoing curative treatment, patients with cancer undergoing palliative treatment, cancer survivors), cancer stage and cancer site.

3. PROM characteristics: PROM specimen, original development paper, original language in which the PROM was developed, target population for whom the PROM was developed, number of subscales and items,

content coverage, recall period, response options, type of scale(s), scoring and estimated duration of assessment. In case of missing data, additional information will be retrieved by searching Google and ePROVIDE (https://eprovide.mapi-trust.org) or by contacting the PROM developers.

4. PROM measurement properties: development and content validity, structural validity/undimensionality, internal consistency, cross-cultural validity and measurement invariance, reliability, measurement error and construct validity. Detailed information on the data that will be extracted for these measurement properties is provided in online supplemental appendix 4.

Following data extraction, all PROMs and related studies will be included in the next phase of the review process for quality assessment.

PROM quality assessment

A scoring manual based on the procedures mentioned hereafter will be built and piloted on a random subsample of studies to enhance the inter-rater homogeneity of PROM quality assessment. The assessment will be performed independently by two reviewers. Discrepancies will be solved by consensus. In case of disagreement, a third reviewer will be involved to solve the discrepancy. As per COSMIN guidelines,[32] quality assessment will be conducted sequentially for each PROM in the following order: development/content validity, internal structure (ie, structural validity, internal consistency and cross-cultural validity/measurement invariance), reliability, measurement error and construct validity (ie, criterion validity and hypotheses testing). The COSMIN group defines content validity as the most important measurement property and recommends assessing it first and excluding PROMs with high-quality evidence of inadequate content validity.[32 40] However, studies that would report on the poor content validity of a PROM are unlikely to be published and this requirement is unlikely to be met, which does not allow for differentiating between PROMs based on the quality of content validity. Therefore, it was decided that the remaining psychometric properties will not be assessed if PROMs demonstrated inadequate content validity at any level of evidence or no evidence of content validity could be found, as PROMs should be relevant, comprehensive and comprehensible with respect to HRQoL and the European cancer population. Studies assessing structural validity based on a Multi-Trait Multimethod approach[41] will be considered to inform construct validity as this method is not appropriate for the assessment of structural validity.[32]

For all psychometric properties, the assessment will be performed at a subscale level (when applicable). Quality assessment will be performed for each study and measurement property as follows:

Table 2 COSMIN criteria for good measurement properties

Measurement property	Rating	Criteria
Structural validity	+	CTT CFA: CFI or TLI or comparable measure>0.95 OR RMSEA<0.06 OR SRMR<0.082 IRT/Rasch ▶ No violation of <u>unidimensionality</u>: CFI or TLI or comparable measure>0.95 OR RMSEA<0.06 OR SRMR<0.08 AND ▶ No violation of <u>local independence</u>: residual correlations among the items after controlling for the dominant factor<0.20 OR Q3's<0.37 AND ▶ No violation of <u>monotonicity</u>: adequate-looking graphs OR item scalability >0.30 AND ▶ Adequate <u>model fit</u>: IRT: $\chi 2 > 0.01$ Rasch: infit and outfit mean squares ≥ 0.5 and ≤ 1.5 OR Z-standardised values > -2 and <2
	±	Results are inconsistent across studies
	–	Criteria for (+) are not met
	?	CTT: Not all information for (+) is reported IRT/Rasch: Model fit not reported OR only EFA was performed
Internal consistency	+	At least low evidence for sufficient structural validity AND reliability coefficient(s) $\geq$ 0.70 for each unidimensional scale or subscale
	±	Results are inconsistent across studies
	–	At least low evidence for sufficient structural validity AND reliability coefficient(s) < 0.70 for each unidimensional scale or subscale
	?	Criteria for 'At least low evidence for sufficient structural validity' are not met: ▶ There is only very low evidence for sufficient structural validity (eg, because there was only one study on structural validity with a very low sample size) ▶ There was (any) evidence for insufficient structural validity ▶ There are inconsistent results for structural validity which cannot be explained ▶ There is no information on the structural validity available
Cross-cultural validity/measurement invariance	+	No important differences found between group factors (such as age, gender, language) in multiple group factor analysis OR no important DIF for group factors (McFadden's $R^2 < 0.02$)
	±	Results are inconsistent across studies
	–	Important differences between group factors OR DIF were found
	?	No multiple group factor analysis OR DIF analysis performed
Reliability	+	Correlation coefficient ≥ 0.70
	±	Results are inconsistent across studies
	–	Correlation coefficient <0.70
	?	Correlation coefficient not reported
Measurement error	+	SDC or LoA<MIC The MIC is defined as the smallest measured change score that patients perceive to be important. If the SDC is smaller than the MIC, it is possible to distinguish a clinically important change from measurement error with a large amount of certainty
	±	Results are inconsistent across studies
	–	SDC or LoA > MIC If the SDC is larger than the MIC, there is a considerable chance that the observed change is caused by measurement error
	?	MIC not defined

Table 2 Continued

Measurement property	Rating	Criteria
Construct validity	+	The result is in accordance with the hypothesis
	±	Results are inconsistent across studies
	–	The result is not in accordance with the hypothesis
	?	No hypotheses were formulated a priori

+, sufficient results; –, insufficient results; ±, inconsistent results; ?, indeterminate results; CFA, confirmatory factor analysis; CFI, Comparative Fit Index; COSMIN, Consensus-based Standards for the selection of health Measurement Instruments; CTT, classical test theory; DIF, differential item functioning; EFA, exploratory factor analysis; IRT, item response theory; LoA, limits of agreement; MCID, minimal clinical important difference; MIC, minimal important change; MID, minimal important difference; RMSEA, root mean square error of approximation; SDC, smallest detectable change; SRMR, standardised root mean residuals; TLI, Tucker-Lewis Index.

Risk of bias assessment

The methodological quality of each study will be evaluated using the COSMIN Risk of Bias Checklist,[42] which provides a set of standards for design requirements and preferred statistical analyses per measurement property. These standards provide a framework to assess whether the results, based on the methodological quality of a given study, are trustworthy. Each standard will be rated on a four-point rating scale as 'very good', 'adequate', 'doubtful' or 'inadequate'. Each assessment of a measurement property is considered to be a separate study. For development/content validity, the quality of each standard will first be determined by retaining the highest rating across the identified studies before taking the lowest rating of each standard to determine the overall quality of the PROM development and content validity. For all other measurement properties, the overall rating of the quality of each study will be determined separately by taking the lowest rating of each standard. Several adjustments were made to the ratings of the COSMIN Risk of Bias Checklist, which are all listed in online supplemental appendix 5.

Criteria for good measurement properties

These criteria are recommendations from COSMIN for which PROMs are assessed as appropriate to be used in research or clinical practice.[32]

Development and content validity

The overall content validity scoring will comprise four steps.[40] First, the results of both the PROM development and content validity studies will be rated by two reviewers independently (online supplemental appendix 6). Each criterion will be scored as 'sufficient' (+), 'insufficient' (−) or 'indeterminate' (?). Reviewers will rate the content of the PROM of interest with 'sufficient' (+) or 'insufficient' (−), using the same criteria. When there is no content validity study available, content validity criteria will be rated 'insufficient' (−). The scoring 'indeterminate' (?) will only be used when there is evidence that some aspects of content validity were assessed, but the authors did not provide enough information to score the criterion appropriately. Second, an overall 'sufficient' (+), 'insufficient' (−), 'indeterminate' (?) or 'inconsistent' (±) rating will be calculated for relevance, comprehensiveness and comprehensibility per study[40] (online supplemental appendix 7). Third, an overall rating per PROM will be calculated for relevance, comprehensiveness and comprehensibility by jointly considering the results of the PROM development and content validity studies, and the reviewer's ratings. The evidence from the content validity will be weighted higher than the evidence from the development study and the reviewer's rating. Online supplemental appendix 8 provides a detailed overview of this overall rating process. Last, an overall 'sufficient'

Table 3 A priori hypotheses for construct validity

Type of construct validity (subtype)	Hypothesis
Between-PROM (convergent validity)	Correlations with instruments measuring similar constructs should be ≥0.50
Between-PROM (convergent/divergent validity)	Correlations with instruments measuring related, but dissimilar constructs should be ≥0.30
Between-PROM (divergent validity)	Correlations with instruments measuring unrelated constructs should be <0.30
Within-PROM (convergent validity)	Correlations between an item and its own scale (corrected for overlap) should be ≥0.40
Within-PROM (divergent validity)	Correlation between an item and its hypothesised subscale (corrected for overlap) is higher than its correlation with the other subscales

PROM, patient-reported outcome measure.

Table 4   COSMIN adapted the GRADE approach for development/content validity

		QUALITY OF EVIDENCE: the starting point is always HIGH	
	⬇	**High**	
		Moderate	
		Low	
		Very Low	
Risk of bias	1: Serious	The content validity study is of doubtful quality. The content validity rating of the content validity study is insufficient (−) OR indeterminate (?) OR inconsistent (±)	
	2: Very serious	No content validity study OR content validity study of insufficient quality (−) **AND** The development study is of doubtful quality. The content validity rating of the development study is indeterminate (?) OR inconsistent (±)	
	3: Very serious	No content validity study OR content validity study of insufficient quality (−) **AND** No development study or development study is of inadequate quality. The content validity rating of the development study is insufficient (−)	
Inconsistency	1: Serious	The combination of the scores for development study, content validity study and reviewer's rating is rated inconsistent (±)	
Indirectness	1: Serious	A content validity study was performed in a cancer population but not representative of the population of interest (eg, patients with head and neck cancer vs patients with cancer, palliative questionnaire assessed in patients with non-palliative cancer)	
	2: Very serious	A content validity study was performed in a non-cancer population.	

COSMIN, Consensus-based Standards for the selection of health Measurement Instruments; GRADE, Grading of Recommendations Assessment, Development and Evaluation.

(+), 'insufficient' (−) or 'inconsistent' (±) content validity rating will be calculated by aggregating the overall relevance, comprehensiveness and comprehensibility rating. Online supplemental appendix 9 provides a detailed overview of the overall content validity rating process.

Other psychometric properties
Criteria for good measurement properties will be applied for each individual study, resulting in a 'sufficient' (+), 'insufficient' (−), or 'indeterminate' (?) rating. The

Table 5   COSMIN adapted the GRADE approach for other psychometric properties

		QUALITY OF EVIDENCE: the starting point is always HIGH	
	⬇	**High**	
		Moderate	
		Low	
		Very low	
Risk of bias (Consider the ratings of the individual studies in STEP 1)	1	There are multiple studies of doubtful (D) quality OR there is only 1 study of adequate (A) quality available	
	2	There are multiple studies of inadequate (I) quality OR there is only 1 study of doubtful quality (D) available	
	3	There is only 1 study of inadequate (I) quality available	
Inconsistency	1	Overall rating across studies is scored with (±)	
Imprecision	1	Total sample size of the pooled or summarised studies <**100**	
	2	Total sample size of the pooled or summarised studies <**50**	
Indirectness*	1	Psychometric properties were assessed in a cancer population but not representative of the target population (eg, patients with head and neck cancer vs patients with cancer, palliative questionnaire assessed in patients with non-palliative cancer)	

*To assess the indirectness, one should look at the characteristics of the pooled population across studies.
COSMIN, Consensus-based Standards for the selection of health Measurement Instruments; GRADE, Grading of Recommendations Assessment, Development and Evaluation.

evidence across studies will be summarised qualitatively, and it will be decided whether the results per psychometric property are consistent. Consistency is defined as at least 75% of individual studies being rated similarly for a given PROM and measurement property. If the threshold of 75% is not reached for any of the rating options and studies with exclusively '+' or '−' ratings are available in combination with '?' ratings, studies with a '?' will be ignored and not included when summarising the results. In all other cases, the overall rating will be scored as 'inconsistent' (±). If the results are inconsistent, possible explanations will be explored and the results will be summarised per subgroup when applicable. If no explanation for the inconsistency can be found, the overall rating will remain 'inconsistent' (±). A detailed overview of the criteria for good measurement properties, incorporating the inconsistency rating, can be found in table 2. For construct validity, a priori hypotheses were formulated to evaluate the results (table 3).

Quality of evidence
The quality of the evidence will be graded per measurement property using a modified Grading of Recommendations Assessment, Development and Evaluation approach (GRADE[32 43]) resulting in four quality levels: 'high', 'moderate', 'low' or 'very low'. Starting from high-quality level, the quality of evidence will be downgraded if applicable according to the following factors: risk of bias (methodological quality of the studies), inconsistency (of results across studies), imprecision (total sample size of the studies) and indirectness (evidence comes from a different target population). For some factors, the original COSMIN modified GRADE approach does not provide clear guidance on the criteria to be used for the risk assessment; therefore, the GRADE approach was further adapted. The adapted GRADE approach that will be used is reported in tables 4 and 5 for development/content validity and the remaining psychometric properties respectively. The quality of evidence for internal consistency will start at the level of structural validity.[32]

Reporting of results
The reporting of the results will follow the PRISMA 2020 statement and a PRISMA checklist will be provided.[44] Considering the expected high heterogeneity of the results, no quantitative pooling of the studies' results per PROM will be performed and no meta-analysis will be planned. In line with the COSMIN guidelines,[32] summary tables describing the PROMs' characteristics, including feasibility and interpretability, and study populations will be produced. The reporting of the results will include the individual ratings on PROM development and content validity, PROM measurement properties and quality of evidence per study. The findings will then be qualitatively summarised as follows.

For content validity, an overall rating per PROM will be calculated for relevance, comprehensiveness and comprehensibility by jointly considering the results of the PROM development and content validity studies, and the reviewer's ratings. The overall content validity will be rated as 'sufficient' (+), 'insufficient' (−) or 'inconsistent' (±), by aggregating the overall relevance, comprehensiveness and comprehensibility rating.

For the remaining psychometric properties, the evidence across studies will be summarised and it will be decided whether the results per psychometric property are consistent. Consistency will be defined as at least 75% of studies being rated similarly for a given PROM and measurement property. If the threshold of 75% is not reached for any of the rating options and studies with exclusively '+' or '−' ratings are available in combination with '?' ratings, studies with a '?' will be ignored and excluded from the summary. In all other cases, the overall rating will be scored as 'inconsistent' (±). For construct validity, a priori hypotheses will be formulated to evaluate the results.

Recommendations
PROMs with sufficient content validity (ie, rated '±' or higher) and at least low-quality evidence (ie, GRADE)[43] for sufficient structural validity and internal consistency will be recommended.[32] On the other hand, PROMs will not be recommended when there is high-quality evidence for any insufficient measurement property. As with the quality assessment, the formulation of recommendations will be made at a subscale level.

Patient and public involvement
Currently, it is expected that researchers actively involve patients, healthcare professionals and the public in their research. Within systematic reviews, active patient and public involvement has been proposed as a way to enhance the actual and perceived usefulness of the summarised evidence, hence addressing barriers to the uptake of evidence in practice.[45] Patient involvement will be ensured at key stages of the systematic review and peer reviewing the academic papers. The results of the review will be discussed with a representative panel of stakeholders, including patients and healthcare professionals to ensure the co-design approach throughout the entire EUonQoL project. It is essential that the PROMs selected to serve as a basis for the development of the EUonQOL toolkit are supported by evidence of content validity, that is, the items constituting these PROMs should be relevant, comprehensive and comprehensible with respect to HRQoL and the European cancer population.

ETHICS AND DISSEMINATION
Ethical clearance for this research is not required, as the systematic review will only use information from previously published research. The results will be disseminated to clinicians, researchers and health policymakers by presenting at relevant conferences and by publication in a peer-reviewed journal. Besides that, the findings will be used to identify the most appropriate PROMs for the

assessment of HRQoL throughout the European cancer continuum, to serve as a basis for the development of the EUonQOL toolkit and to provide evidence-based recommendations to the EUonQOL consortium.

Author affiliations
¹Quality of Life Department, European Organisation for Research and Treatment of Cancer, Brussels, Belgium
²Fondazione IRCCS Istituto Nazionale dei Tumori, Milano, Italy
³The Research Unit, Department of Palliative Medicine, Bispebjerg Hospital, Copenhagen, Denmark
⁴Organisation of European Cancer Institutes – European Economic Interest Grouping, Rome, Italy
⁵Leeds Institute of Medical Research, University of Leeds, Leeds, UK
⁶Applied Research Division for Cognitive and Psychological Science, Istituto Europeo di Oncologia, Milano, Italy
⁷Department of Oncology and Hemato-oncology, Universita degli Studi di Milano, Milano, Lombardia, Italy
⁸Regina Elena Institute, Roma, Lazio, Italy
⁹Palliative Medicine, University of Copenhagen, Kobenhavn, Region Hovedstaden, Denmark
¹⁰Hospital del Mar Research Institute, Barcelona, Spain
¹¹CIBERESP, Madrid, Spain
¹², Netherlands Institute for Health Services Research (Nivel), Utrecht, The Netherlands
¹³SporeData OU, Tallinn, Estonia
¹⁴European Cancer Organisation, Brussels, Belgium
¹⁵Palliative Care, Pain Therapy and Rehabilitation, Fondazione IRCCS Istituto Nazionale dei Tumori, Milano, Italy
¹⁶Oncology, Oslo University Hospital, Oslo, Norway
¹⁷Unicancer, Paris, Île-de-France, France

ORCID iDs
Laurence Leysen http://orcid.org/0000-0003-0246-7002
Cinzia Brunelli http://orcid.org/0000-0003-3905-1289
Hugo Vachon http://orcid.org/0000-0003-1259-649X

REFERENCES

1. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA* 1995;273:59–65.
2. Bakas T, McLennon SM, Carpenter JS, *et al*. Systematic review of health-related quality of life models. *Health Qual Life Outcomes* 2012;10:134.
3. Fiteni F, Cuenant A, Favier M, *et al*. Clinical Relevance of Routine Monitoring of Patient-reported Outcomes *Versus* Clinician-reported Outcomes in Oncology. *In Vivo* 2019;33:17–21.
4. Deshpande PR, Rajan S, Sudeepthi BL, *et al*. Patient-reported outcomes: A new era in clinical research. *Perspect Clin Res* 2011;2:137–44.
5. LeBlanc TW, Abernethy AP. Patient-reported outcomes in cancer care - hearing the patient voice at greater volume. *Nat Rev Clin Oncol* 2017;14:763–72.
6. Graupner C, Kimman ML, Mul S, *et al*. Patient outcomes, patient experiences and process indicators associated with the routine use of patient-reported outcome measures (PROMs) in cancer care: a systematic review. *Support Care Cancer* 2021;29:573–93.
7. European Commission. EUR-Lex 2021. Europe's beating cancer plan. 2021. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52021DC0044
8. Laugsand EA, Sprangers MAG, Bjordal K, *et al*. Health care providers underestimate symptom intensities of cancer patients: a multicenter European study. *Health Qual Life Outcomes* 2010;8:104.
9. Di Maio M, Gallo C, Leighl NB, *et al*. Symptomatic toxicities experienced during anticancer treatment: agreement between patient and physician reporting in three randomized trials. *J Clin Oncol* 2015;33:910–5.
10. U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research. *Guidance for industry on patient-reported outcome measures: use in medical product development to support labeling claims*. 2009.
11. Gotay CC. Assessing cancer-related quality of life across a spectrum of applications. *J Natl Cancer Inst Monogr* 2004;126–33.
12. Amarsheda S, Bhise AR. Systematic review of cancer-related fatigue instruments in breast cancer patients. *Palliat Support Care* 2022;20:122–8.
13. Afseth J, Neubeck L, Karatzias T, *et al*. Holistic needs assessment in brain cancer patients: A systematic review of available tools. *Eur J Cancer Care (Engl)* 2019;28:e12931.
14. Bryant AL, Walton A, Shaw-Kokot J, *et al*. A Systematic Review of Psychometric Properties of Health-Related Quality-of-Life and Symptom Instruments in Adult Acute Leukemia Survivors. *Cancer Nurs* 2016;39:375–82.
15. Harrington S, Lee J, Colon G, *et al*. Oncology Section EDGE Task Force on Prostate Cancer: A Systematic Review of Outcome Measures for Health-Related Quality of Life. *Rehabil Oncol* 2016;34:27–35.
16. Gondivkar SM, Gadbail AR, Sarode SC, *et al*. Measurement properties of oral health related patient reported outcome measures in patients with oral cancer: A systematic review using COSMIN checklist. *PLoS One* 2019;14:e0218833.
17. Mason SJ, Catto JWF, Downing A, *et al*. Evaluating patient-reported outcome measures (PROMs) for bladder cancer: a systematic review using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist. *BJU Int* 2018;122:760–73.
18. Parkar S, Sharma A. Validation of European Organization for Research and Treatment of Cancer Head and Neck Cancer Quality of Life Questionnaire (EORTC QLQ-H&N35) Across Languages: A Systematic Review. *Indian J Otolaryngol Head Neck Surg* 2022;74:6100–7.
19. Preston NJ, Wilson N, Wood NJ, *et al*. Patient-reported outcome measures for use in gynaecological oncology: a systematic review. *BJOG* 2015;122:615–22.
20. Ratti MM, Gandaglia G, Sisca ES, *et al*. A Systematic Review to Evaluate Patient-Reported Outcome Measures (PROMs) for Metastatic Prostate Cancer According to the COnsensus-Based Standard for the Selection of Health Measurement INstruments (COSMIN) Methodology. *Cancers (Basel)* 2022;14:5120.
21. Salas M, Mordin M, Castro C, *et al*. Health-related quality of life in women with breast cancer: a review of measures. *BMC Cancer* 2022;22:66.
22. Tax C, Steenbergen ME, Zusterzeel PLM, *et al*. Measuring health-related quality of life in cervical cancer patients: a systematic review of the most used questionnaires and their validity. *BMC Med Res Methodol* 2017;17:15.
23. van der Hout A, Neijenhuijs KI, Jansen F, *et al*. Measuring health-related quality of life in colorectal cancer patients: systematic review

of measurement properties of the EORTC QLQ-CR29. *Support Care Cancer* 2019;27:2395–412.

24 van Roij J, Fransen H, van de Poll-Franse L, *et al*. Measuring health-related quality of life in patients with advanced cancer: a systematic review of self-administered measurement instruments. *Qual Life Res* 2018;27:1937–55.

25 Pearce NJM, Sanson-Fisher R, Campbell HS. Measuring quality of life in cancer survivors: a methodological review of existing scales. *Psychooncology* 2008;17:629–40.

26 Treanor C, Donnelly M. A methodological review of the Short Form Health Survey 36 (SF-36) and its derivatives among breast cancer survivors. *Qual Life Res* 2015;24:339–62.

27 Al Maqbali M, Hughes C, Gracey J, *et al*. Quality assessment criteria: psychometric properties of measurement tools for cancer related fatigue. *Acta Oncol* 2019;58:1286–97.

28 Abahussin AA, West RM, Wong DC, *et al*. PROMs for Pain in Adult Cancer Patients: A Systematic Review of Measurement Properties. *Pain Pract* 2019;19:93–117.

29 Curcio KR. Instruments for Assessing Chemotherapy-Induced Peripheral Neuropathy: A Review of the Literature. *Clin J Oncol Nurs* 2016;20:144–51.

30 Albach CA, Wagland R, Hunt KJ. Cross-cultural adaptation and measurement properties of generic and cancer-related patient-reported outcome measures (PROMs) for use with cancer patients in Brazil: a systematic review. *Qual Life Res* 2018;27:857–70.

31 Mokkink LB, Terwee CB, Stratford PW, *et al*. Evaluation of the methodological quality of systematic reviews of health status measurement instruments. *Qual Life Res* 2009;18:313–33.

32 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.

33 Luckett T, King MT, Butow PN, *et al*. Choosing between the EORTC QLQ-C30 and FACT-G for measuring health-related quality of life in cancer clinical research: issues, evidence and recommendations. *Ann Oncol* 2011;22:2179–90.

34 Moher D, Shamseer L, Clarke M, *et al*. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* 2015;4:1.

35 Ouzzani M, Hammady H, Fedorowicz Z, *et al*. Rayyan-a web and mobile app for systematic reviews. *Syst Rev* 2016;5:210.

36 Higgins JPT, Thomas J, Chandler J, *et al*. Cochrane handbook for systematic reviews of interventions version 6.3. 2023. Available: www.training.cochrane.org/handbook

37 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.

38 Mokkink L, Terwee C, de Vet H. Key concepts in clinical epidemiology: Responsiveness, the longitudinal aspect of validity. *J Clin Epidemiol* 2021;140:159–62.

39 Page MJ, McKenzie JE, Bossuyt PM, *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71.

40 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.

41 Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull* 1959;56:81–105.

42 Mokkink LB, Terwee CB, Patrick DL, *et al*. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010;19:539–49.

43 Schünemann HJ. GRADE: from grading the evidence to developing recommendations. A description of the system and a proposal regarding the transferability of the results of clinical research to clinical practice. *Z Evid Fortbild Qual Gesundhwes* 2009;103:391–400.

44 Shamseer L, Moher D, Clarke M, *et al*. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015;350:g7647.

45 Wallace J, Nwosu B, Clarke M. Barriers to the uptake of evidence from systematic reviews and meta-analyses: a systematic review of decision makers' perceptions. *BMJ Open* 2012;2:e001220.