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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2024-088716
Article Type:	Protocol
Date Submitted by the Author:	13-May-2024
Complete List of Authors:	Leysen, Laurence; European Organisation for Research and Treatment of Cancer, Quality of Life Marticic Giljevic, Karla; European Organisation for Research and Treatment of Cancer, Quality of Life Piccinin, Claire; European Organisation for Research and Treatment of Cancer, Quality of Life Shkodra, Morena; Fondazione IRCCS Istituto Nazionale dei Tumori Pe, Madeline; European Organisation for Research and Treatment of Cancer, Petersen, Morten; Bispebjerg Hospital, The Research Unit, Department of Palliative Medicine Apolone, Giovanni; Fondazione IRCCS Istituto Nazionale dei Tumori Brunelli, Cinzia; Fondazione IRCCS Istituto Nazionale dei Tumori Lombardo, Claudio; Organisation of European Cancer Institutes – European Economic Interest Grouping Velikova, Galina; University of Leeds, Leeds Institute of Medical Research Gilbert, Alexandra; University of Leeds, Leeds Institute of Medical Research Pravettoni, Gabriella; Istituto Europeo di Oncologia, Applied Research Division for Cognitive and Psychological Science; Universita degli Studi di Milano, Department of Oncology and Hemato-oncology Ciliberto, Gennaro; Regina Elena Institute Groenvold, Mogens; University of Copenhagen, Palliative Medicine Ferrer, Montserrat; Hospital del Mar Research Institute; CIBERESP Bos, Nanne; Nederlands Instituut voor Onderzoek van de Gezondheidszorg, Zorg vanuit patiëntperspectief Pietrobon, Ricardo; SporeData OU Couespel, Norbert; European CanCer Organisation Caraceni, A; Fondazione IRCCS Istituto Nazionale dei Tumori, Palliative Care, Pain Therapy and Rehabilitation Guren, Marianne; Oslo University Hospital, Oncology Sirven, Aude; UNICANCER Vachon, Hugo; European Organisation for Research and Treatment of Cancer, Quality of Life Department
Keywords:	Quality of Life, ONCOLOGY, Patient Reported Outcome Measures, Psychometrics, QUALITATIVE RESEARCH, Systematic Review

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

Leysen, L.[1], Martičić Giljević, K.[1], Piccinin, C.[1], Shkodra, M.[2], Pe., M.[1], Petersen, M.[3], Apolone, G.[4], Brunelli, C.[4], Lombardo, C.[5], Velikova, G.[6], Gilbert, A.[6], Pravettoni, G.[7], Ciliberto, G.[8], Groenvold, M.[3], Ferrer, M.[9], Bos, N.[10], Pietrobon, R.[11], Couespel, N.[12], Caraceni, A.[13], Guren, M.G.[14], Sirven, A.[15], & Vachon, H. on behalf of the EUonQOL consortium

Author affiliations:

[1] Quality of Life Department, European Organisation for Research and Treatment of Cancer, Brussels, Belgium

[2] Palliative Care, Pain Therapy and Rehabilitation Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milano, Italy

[3] The Research Unit, Department of Palliative Medicine, Bispebjerg Hospital, University of Copenhagen, Copenhagen, Denmark

[4] Fondazione IRCCS Istituto Nazionale Dei Tumori di Milano, Milan, Italy

[5] Organisation of European Cancer Institutes, Brussels, Belgium

[6] Leeds Institute of Cancer and Pathology, University of Leeds, St James's Hospital, Leeds, UK

[7] Applied Research Division for Cognitive and Psychological Science, IEO, European Institute of Oncology IRCCS, Milan, Italy

[8] IRCCS National Cancer Institute Regina Elena, Rome, Italy

[9] Hospital del Mar Research Institute, Barcelona, Spain

[10] Netherlands Institute for Health Services Research (Nivel), Utrecht, The Netherlands

[11] SporeData OÜ, Tallinn, Estonia

[12] European Cancer Organisation (ECO), Brussels, Belgium

[13] Università Degli Studi Di Milano, Milan, Italy

[14] Department of Oncology, Oslo University Hospital, Oslo, Norway

[15] Unicancer, Paris, France

Correspondence: Laurence Leysen, Quality of Life Department, European Research for Treatment and Research of Cancer, laurence.leysen@eortc.org, address: Avenue E. Mounier 83/11, 1200 Brussels, Belgium.

Protocol registration number: CRD42023418616

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

Abstract

INTRODUCTION: Over the past decades, there has been increasing recognition that assessing cancer patients' 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 (through PubMed) and ELSEVIER (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 summarized and if possible, data will be pooled using meta-analyses. The Grading of Recommendations, Assessment, Development and Evaluations (GRADE) guidelines will be used to grade and summarize the quality of the evidence.

ETHICS AND DISSEMINATION: The results of this review will not only be submitted for publication in a peer-reviewed journal, but will also 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 cancer patients and survivors.

Protocol registration number: CRD42023418616

Strengths and limitations of this study

- This systematic review will be the first to provide a comprehensive overview of the psychometric properties on a subscale level of the patient-reported outcome measures (PROMs) used for the assessment of health-related quality of life (HRQoL) across the European cancer continuum.
- This systematic review will follow the highest methodological standards, i.e., the Preferred Reporting Items for Systematic Review and Meta-Analyses protocols (PRISMA-P) and the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN).
- The results of this systematic review will be used to provide a set of evidence-based recommendations for a European project (EUonQOL), which aims at developing a new PROM (i.e., EUonQOL toolkit) to assess the multidimensional construct of HRQoL across a wide variety of cancer patients over the European Union and its associated countries.
- The systematic review will specifically include PROMs validated in European cancer patients and survivors, thus PROMs which have only been validated outside of Europe will not be covered.
- The heterogenous nature of both the clinical population and the methodologic approaches to assess psychometric properties may prevent quantitative pooling of data (i.e. meta-analysis).

1. 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 cancer patients’ 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 (e.g., Europe’s Beating Cancer plan,(7)). However, HRQoL is often inaccurately assessed by health care providers (HCPs) 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 cancer patients, a wide array of PROMs is currently available, ranging from generic (e.g., SF-36, EQ-5D-5L) to cancer-specific (e.g., EORTC QLQ-C30, FACT-G) and tumour-specific tools (e.g., EORTC QLQ-BR23, 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 (20).

Over the past years, many systematic reviews comparing PROMs for the assessment of HRQoL in cancer patients have been published. Most of them focused on PROMs measuring HRQoL in a specific type of cancer (e.g., breast cancer, prostate cancer, etc.) (21–32) or cancer population (e.g., cancer survivors, advanced cancer, palliative patients, etc.) (23,33–35). Several of these reviews focused on PROMs evaluating one specific HRQoL-related construct (e.g., depression, fatigue, pain, etc.) (21,22,36–38) and the majority did not report the psychometric properties of the PROMs under investigation per subscale (22–26,28–31,33,34,36,37,39). 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 (40).

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,(41)). 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 (21,25,29,36,42) and the risk of bias assessment or the grading of the evidence were not conducted (21,22,33,36,39,42). 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 throughout the European cancer continuum, to make objective recommendations on the most suitable PROM to use in this population.

2. 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 (43) and has been prospectively registered in the International Prospective Register of Systematic Reviews database (PROSPERO 2023 - CRD42023418616).

The systematic review will be conducted according to the COSMIN guidelines for systematic reviews (41) and will use the COSMIN taxonomy of measurement properties (Table 1). All steps of the screening process will be performed using RAYYAN (44).

2.1 Search strategy

A systematic search will be performed in the bibliographic databases MEDLINE (through PubMed) and ELSEVIER (through Scopus) without publication date restriction. The search strategy will be based on the PICOM acronym (45) in which the population will be represented by cancer patients and survivors, the outcome by health-related quality of life and the methods by psychometric properties. No comparator or intervention will be used. Both 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 filter of Terwee et al. (46) 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 Appendix 1.

2.2 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 standardized matter. 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. Discrepancies will be resolved by discussion and/or consultation of a third reviewer.

2.2.1 PROM selection

To be included PROMs will need to meet following criteria:

1) PROMs must be self-administered questionnaire (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 (i.e., based on a multidimensional model of HRQoL) and be applicable across cancer types. This excludes tools designed to assess a specific HRQoL subdomain (e.g., exclusively assessing physical functioning) or cancer site (e.g., 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 (i.e., evidence of content validity, structural validity, or construct validity) for use in the target population of European cancer patients or survivors. In case no European validation[1] can be found for a PROM identified through the initial search, an additional search will be performed in PubMed (Appendix 3). If no evidence of validity among European cancer patients

[1] European Union and associated countries (for the full list of countries, please see Appendix 2)

or survivors can be retrieved after the additional search, the PROM and its related articles will be excluded.

2.2.2 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 gold standard for PROMs (41). 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 (47).
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 article and reviews will not be considered.
3) Studies must be performed in adult European cancer patients or survivors (mean age ≥ 21 years and not defined as Adolescents and Young Adults (AYA)). Articles including "mixed samples" (i.e., European cancer patients and non-cancer patients) will only be included if separate results are provided for the cancer patients group. Studies involving both European and non-European cancer patients, 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 cancer patients will also be excluded.

Detailed information on the selection process will be reported in a PRISMA flowchart.

2.3 Data extraction

During the data extraction, it will be determined which measurement properties will be evaluated for every included study. Data extraction will be done by one reviewer and checked by a second reviewer. 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, non-cancer patients, cancer patients undergoing curative treatment, cancer patients 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 Appendix 4.

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

2.4 PROM quality assessment

Quality 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 (41), quality assessment will be conducted sequentially for each PROM in the following order: development/content validity[2], internal structure (i.e., structural validity, internal consistency, and cross-cultural validity/measurement invariance), reliability, measurement error and construct validity (i.e., 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 (41,48). 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 (49) will be considered to inform construct validity as this method is not appropriate for the assessment of structural validity (41).

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:

2.4.1 Risk of Bias assessment

The methodological quality of each study will be evaluated using the COSMIN Risk of Bias Checklist (50), 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 Appendix 5.

2.4.2 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 (41).

Development and content validity

The overall content validity scoring will comprise four steps (48). First, the results of both the PROM development and content validity studies will be rated by two reviewers independently (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

[2] PROM development is not a measurement property, but is taken into account when evaluating content validity as per COSMIN guidelines

for relevance, comprehensiveness and comprehensibility per study (48) (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. Appendix 8 provides a detailed overview of this overall rating process. Last, an overall "sufficient" (+), "insufficient" (−) or "inconsistent" (±) content validity rating will be calculated, by aggregating the overall relevance, comprehensiveness and comprehensibility rating. 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 evidence across studies will be summarized 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 summarizing 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 summarized 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).

2.4.3 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) (41,51) resulting in 4 quality levels: "high", "moderate", "low", or "very low". Starting from high-quality level, 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[3] (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 (41).

2.5 Meta-analysis

For this review, conventional meta-analytic techniques will be applied. Fisher's Z will be calculated, ranging from −∞ to +∞ and can be interpreted similar to a correlation coefficient, as the standardised common effect size. Depending on the measurement property and the reported estimates, intraclass correlations, as well as Pearson and Spearman correlations, will be converted to Fisher's Z using Fisher's variance stabilizing transformation (52,53). Due to the potential heterogeneity of the study samples, methodological approaches and estimates to assess psychometric properties, pooling of data in a meta-analysis may not be possible.

Meta-analyses will be performed applying the random-effect model, which assumes the average effect size varies between studies, hence heterogeneity is to be expected (45). Heterogeneity will be assessed

[3] Imprecision is not taken into account when grading the quality of evidence for content validity

by the I^2 statistics using the method proposed by Higgens et al. (54). The I^2 statistic represents an estimation of variability in effect estimates due to heterogeneity as opposed to sampling error (54). The Cochrane Handbook provides a rough interpretation of the I^2 statistic in which 0%–40% represents no important heterogeneity; 30%–60%: moderate heterogeneity; 50%–90%: substantial heterogeneity and 75%–100%: considerable heterogeneity (45). If significant heterogeneity is demonstrated, subsequent sensitivity analyses will be conducted to verify the robustness of the results (45). If pooling of the results is not possible, the findings from the different studies on each measurement property per subscale of each PROM will be only qualitatively summarized with, if possible, the quantitative range of values across studies.

2.6 Recommendations

PROMs with sufficient content validity (i.e., rated "±" or higher) and at least low-quality evidence (i.e., GRADE) (32) for sufficient structural validity and internal consistency will be recommended (41). 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.

2.7 Patient and public involvement

Currently, it is expected that researchers actively involve patients, healthcare professionals and 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 summarized evidence, hence addressing barriers to the uptake of evidence in practice (55). 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, i.e., the items constituting these PROMs should be relevant, comprehensive, and comprehensible with respect to HRQoL and the European cancer population.

3. 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.

4. Additional information

Contributors: HV conceived the idea. HV, KM and LL planned and designed the study protocol, the data extraction and the data assessment. HV, KM and LL wrote the first draft of the protocol. CP, MP, MS and MP provided critical insights. All authors have approved and contributed to the final written manuscript.

Funding: this work was supported by the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement Number 101096362.

Competing interests: None to declare.

Patient and public involvement: Patients and/or public will be involved in the conduction, reporting or dissemination plans of this research. Refer to Methods section for further details.

Patient consent for publication: Not required.

5. References

1. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. JAMA. 1995 Jan;273(1):59–65.
2. Bakas T, McLennon SM, Carpenter JS, Buelow JM, Otte JL, Hanna KM, et al. Systematic review of health-related quality of life models. Health Qual Life Outcomes. 2012;10:1–12.
3. Fiteni F, Cuenant A, Favier M, Cousin C, Houede N. Clinical Relevance of Routine Monitoring of Patient-reported Outcomes Versus Clinician-reported Outcomes in Oncology. In Vivo. 2019;33(1):17–21.
4. Deshpande PR, Rajan S, Sudeepthi BL, Abdul Nazir CP. Patient-reported outcomes: A new era in clinical research. Perspect Clin Res. 2011 Oct;2(4):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 Dec;14(12):763–72.
6. Graupner C, Kimman ML, Mul S, Slok AHM, Claessens D, Kleijnen J, 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 Feb;29(2):573–93.
7. European Commission. EUR-Lex. 2021 [cited 2023 Sep 5]. Europe's Beating Cancer plan. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52021DC0044
8. Laugsand EA, Sprangers MAG, Bjordal K, Skorpen F, Kaasa S, Klepstad P. Health care providers underestimate symptom intensities of cancer patients: a multicenter European study. Health Qual Life Outcomes. 2010 Sep;8:104.
9. Di Maio M, Gallo C, Leighl NB, Piccirillo MC, Daniele G, Nuzzo F, et al. Symptomatic toxicities experienced during anticancer treatment: agreement between patient and physician reporting in three randomized trials. J Clin Oncol. 2015 Mar;33(8):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; Availability. 2009.
11. LeBlanc TW, Abernethy AP. Patient-reported outcomes in cancer care - hearing the patient voice at greater volume. Nat Rev Clin Oncol. 2017 Dec;14(12):763–72.
12. Trautmann F, Hentschel L, Hornemann B, Rentsch A, Baumann M, Ehninger G, et al. Electronic real-time assessment of patient-reported outcomes in routine care-first findings and experiences from the implementation in a comprehensive cancer center. Support Care Cancer. 2016 Jul;24(7):3047–56.
13. Kotronoulas G, Kearney N, Maguire R, Harrow A, Di Domenico D, Croy S, et al. What is the value of the routine use of patient-reported outcome measures toward improvement of patient outcomes, processes of care, and health service outcomes in cancer care? A systematic review of controlled trials. J Clin Oncol. 2014 May;32(14):1480–501.

14. Graupner C, Kimman ML, Mul S, Slok AHM, Claessens D, Kleijnen J, 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 Feb;29(2):573–93.

15. Detmar SB, Muller MJ, Schornagel JH, Wever LD V, Aaronson NK. Health-related quality-of-life assessments and patient-physician communication: a randomized controlled trial. JAMA. 2002 Dec;288(23):3027–34.

16. Boyes A, Newell S, Girgis A, McElduff P, Sanson-Fisher R. Does routine assessment and real-time feedback improve cancer patients' psychosocial well-being? Eur J Cancer Care (Engl). 2006 May;15(2):163–71.

17. Trautmann F, Hentschel L, Hornemann B, Rentsch A, Baumann M, Ehninger G, et al. Electronic real-time assessment of patient-reported outcomes in routine care-first findings and experiences from the implementation in a comprehensive cancer center. Support Care Cancer. 2016 Jul;24(7):3047–56.

18. Nelson EC, Eftimovska E, Lind C, Hager A, Wasson JH, Lindblad S. Patient reported outcome measures in practice. BMJ. 2015 Feb;350:g7818.

19. Higginson IJ, Carr AJ. Measuring quality of life: Using quality of life measures in the clinical setting. BMJ. 2001 May;322(7297):1297–300.

20. Gotay CC. Assessing cancer-related quality of life across a spectrum of applications. J Natl Cancer Inst Monogr. 2004;(33):126–33.

21. Amarsheda S, Bhise AR. Systematic review of cancer-related fatigue instruments in breast cancer patients. Palliat Support Care. 2022 Feb;20(1):122–8.

22. Afseth J, Neubeck L, Karatzias T, Grant R. Holistic needs assessment in brain cancer patients: A systematic review of available tools. Eur J Cancer Care. 2019 May;28(3):e12931.

23. Bryant AL, Walton A, Shaw-Kokot J, Mayer DK, Reeve BB. A Systematic Review of Psychometric Properties of Health-Related Quality-of-Life and Symptom Instruments in Adult Acute Leukemia Survivors. Cancer Nurs. 2016;39(5):375–82.

24. Harrington S, Lee J, Colon G, Alappattu M. Oncology Section EDGE Task Force on Prostate Cancer: A Systematic Review of Outcome Measures for Health-Related Quality of Life. Rehabil Oncol. 2016 Jan;34(1):27–35.

25. Gondivkar SM, Gadbail AR, Sarode SC, Gondivkar RS, Yuwanati M, Sarode GS, 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(6):e0218833.

26. Mason SJ, Catto JWF, Downing A, Bottomley SE, Glaser AW, Wright P. 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 Nov;122(5):760–73.

27. 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 Dec;74(Suppl 3):6100–7.

28. Preston NJ, Wilson N, Wood NJ, Brine J, Ferreira J, Brearley SG. Patient-reported outcome measures for use in gynaecological oncology: a systematic review. BJOG. 2015 Apr;122(5):615–22.

29. Ratti MM, Gandaglia G, Sisca ES, Derevianko A, Alleva E, Beyer K, 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 Oct;14(20).

30. Salas M, Mordin M, Castro C, Islam Z, Tu N, Hackshaw MD. Health-related quality of life in women with breast cancer: a review of measures. BMC Cancer. 2022 Jan;22(1):66.

31. Tax C, Steenbergen ME, Zusterzeel PLM, Bekkers RLM, Rovers MM. 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 Jan;17(1):15.

32. van der Hout A, Neijenhuijs KI, Jansen F, van Uden-Kraan CF, Aaronson NK, Groenvold M, 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 Jul;27(7):2395–412.

33. van Roij J, Fransen H, van de Poll-Franse L, Zijlstra M, Raijmakers N. Measuring health-related quality of life in patients with advanced cancer: a systematic review of self-administered measurement instruments. Qual Life Res. 2018 Aug;27(8):1937–55.

34. Pearce NJM, Sanson-Fisher R, Campbell HS. Measuring quality of life in cancer survivors: a methodological review of existing scales. Psychooncology. 2008 Jul;17(7):629–40.

35. 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 Feb;24(2):339–62.

36. Al Maqbali M, Hughes C, Gracey J, Rankin J, Dunwoody L, Hacker E. Quality assessment criteria: psychometric properties of measurement tools for cancer related fatigue. Acta Oncol. 2019 Sep;58(9):1286–97.

37. Abahussin AA, West RM, Wong DC, Ziegler LE. PROMs for Pain in Adult Cancer Patients: A Systematic Review of Measurement Properties. Pain Pract. 2019 Jan;19(1):93–117.

38. Curcio KR. Instruments for Assessing Chemotherapy-Induced Peripheral Neuropathy: A Review of the Literature. Clin J Oncol Nurs. 2016 Apr;20(2):144–51.

39. 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 Apr;27(4):857–70.

40. Mokkink LB, Terwee CB, Stratford PW, Alonso J, Patrick DL, Riphagen I, et al. Evaluation of the methodological quality of systematic reviews of health status measurement instruments. Qual Life Res. 2009 Apr;18(3):313–33.

41. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018 May;27(5):1147–57.

42. Luckett T, King MT, Butow PN, Oguchi M, Rankin N, Price MA, 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 Oct;22(10):2179–90.

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

44. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev. 2016 Dec;5(1):210.

45. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T PMW. Cochrane Handbook for Systematic Reviews of Interventions version 6.3 [Internet]. 2023. Available from: www.training.cochrane.org/handbook

46. Terwee CB, Jansma EP, Riphagen II, de Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res. 2009 Oct;18(8):1115–23.

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

48. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res. 2018 May;27(5):1159–70.

49. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychol Bull. 1959 Mar;56(2):81–105.

50. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol 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 May;19(4):539–49.

51. 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(6):391–400.

52. Rosenthal R. Parametric measures of effect size. In: The handbook of research synthesis. New York, NY, US: Russell Sage Foundation; 1994. p. 231–44.

53. Metawin 2.0. User's manual: statistical software for meta-analysis. Sinauer Associates; 2000.

54. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003 Sep;327(7414):557–60.

55. 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(5).

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 interrelatedness 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: e.g., 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 (i.e., raters or responders) on different occasions (intra-rater)
Measurement error	The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured
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

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 unidimensionality: CFI or TLI or comparable measure >0.95 OR RMSEA <0.06 OR SRMR <0.08 AND - No violation of local independence: residual correlations among the items after controlling for the dominant factor <0.20 OR Q3's < 0.37 AND - No violation of monotonicity: adequate looking graphs OR item scalability >0.30 AND - Adequate model fit: IRT: $\chi 2$ >0.01 Rasch: infit and outfit mean squares ≥ 0.5 and ≤ 1.5 OR Z-standardized 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) ≥ 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 (e.g., because there was only 1 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 R2 < 0.02)
	±	Results are inconsistent across studies
	-	Important differences between group factors OR DIF was 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
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

Abbreviations: + = sufficient results; - = insufficient results; ± = inconsistent results; ? = indeterminate results; CFA = Confirmatory Factor Analysis; CFI = Comparative Fit Index; CTT = Classical Test Theory; DIF = Differential Item Functioning; LoA = Limits of Agreement; IRT = Item Response Theory; MIC = Minimal Important Change; MID: Minimal Important Difference; MCID = Minimal Clinical Important Difference; RMSEA = Root Mean Square Error of Approximation; SDC = Smallest Detectable Change; SRMR: Standardized Root Mean Residuals; TLI: Tucker-Lewis Index.

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 hypothesized subscale (corrected for overlap) is higher than its correlation with the other subscales

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

	QUALITY OF EVIDENCE: starting point is always HIGH ↓ HIGH MODERATE LOW VERY LOW	
Risk of bias	- 1: Serious	Content validity study is of doubtful quality. The content validity rating of content validity study is insufficient (-) OR indeterminate (?) OR inconsistent (±)
	- 2: Very serious	No content validity study OR content validity study of insufficient quality (-) **AND** 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 (±) **(see scoring table below)**
Indirectness	- 1: Serious	Content validity study was performed in a cancer population but not representative of the population of interest (e.g. head & neck cancer patients versus cancer patients, palliative questionnaire assessed in non-palliative cancer patients)
	- 2: Very serious	Content validity study was performed in a non-cancer population.

Table 5 COSMIN adapted GRADE approach for other psychometric properties

		QUALITY OF EVIDENCE: starting point is always HIGH ↓ HIGH MODERATE LOW VERY LOW
Risk of bias (Consider the ratings of the individual studies in STEP 1)	-1	The 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 summarized studies **<100**
	-2	Total sample size of the pooled or summarized studies **<50**
Indirectness*	-1	Psychometric properties were assessed in a cancer population but not representative of the target population (e.g. head & neck cancer patients versus cancer patients, palliative questionnaire assessed in non-palliative cancer patients)
* To assess the indirectness one should look at the characteristics of the pooled population across studies.		

Appendix 1. Detailed overview of the search strategy for PubMed and Scopus

	PubMed	Scopus
Population: cancer patients & survivors	("patient*"[MeSH Terms] OR "Survivors"[MeSH Terms] OR "Palliative Care"[MeSH Terms]) AND ("Neoplasms"[MeSH Terms] OR "Carcinoma"[MeSH Terms] OR "post-cancer"[Title/Abstract] OR "postcancer"[Title/Abstract])	((TITLE-ABS-KEY ("tumor*")) OR (TITLE-ABS-KEY ("neoplasm*")) OR (TITLE-ABS-KEY ("neoplasia*")) OR (TITLE-ABS-KEY ("cancer*")) OR (TITLE-ABS-KEY ("malignanc*")) OR (TITLE-ABS-KEY ("carcinoma*")) OR (TITLE-ABS-KEY ("postcancer")) OR (TITLE-ABS-KEY ("post-cancer"))) AND ((TITLE-ABS-KEY ("palliative care")) OR (TITLE-ABS-KEY ("palliative treatment*")) OR (TITLE-ABS-KEY ("palliative therap*")) OR (TITLE-ABS-KEY ("palliative surger*")) OR (TITLE-ABS-KEY ("palliative supportive care*")) OR (TITLE-ABS-KEY ("survivor*")) OR (TITLE-ABS-KEY ("patient*")))
Exposure: psychometric properties	AND ("instrument*"[Title/Abstract] OR "questionnaire*"[Title/Abstract] OR "measur*"[Title/Abstract] OR "rating*"[Title/Abstract] OR "computer*"[Title/Abstract] OR "digital*"[Title/Abstract] OR "computer-adaptive test*"[Title/Abstract] OR "computer adaptive test*"[Title/Abstract] OR "computer adaptive"[Title/Abstract] OR "computer-adaptive"[Title/Abstract] OR "computerized adaptive test*"[Title/Abstract] OR "computerised adaptive test*"[Title/Abstract] OR "CAT"[Title/Abstract]) AND ("chronbach*"[Title/Abstract] OR "cronbach*"[Title/Abstract] OR "psychometric properties"[Title/Abstract] OR "psychometr*"[Title/Abstract] OR "factor analysis"[Title/Abstract] OR "develop*"[Title/Abstract] OR "reliab*"[Title/Abstract] OR "valid*"[Title/Abstract] OR "translat*"[Title/Abstract] OR "cross-cultural"[Title/Abstract] OR "minimal clinically important difference*"[Title/Abstract] OR "minimal important change*"[Title/Abstract] OR "minimal important difference*"[Title/Abstract] OR "clinically meaningful change*"[Title/Abstract] OR "clinically meaningful difference*"[Title/Abstract] OR "responsiveness"[Title/Abstract])	AND ((TITLE-ABS-KEY ("questionnaire")) OR (TITLE-ABS-KEY ("questionnaires")) OR (TITLE-ABS-KEY ("instrument")) OR (TITLE-ABS-KEY ("instruments")) OR (TITLE-ABS-KEY ("rating")) OR (TITLE-ABS-KEY ("outcome measure")) OR (TITLE-ABS-KEY ("outcome measures")) OR (TITLE-ABS-KEY ("measurement tool")) OR (TITLE-ABS-KEY ("measurement tools")) OR (TITLE-ABS-KEY ("computer-based")) OR (TITLE-ABS-KEY ("digital")) OR (TITLE-ABS-KEY ("computer adaptive test*")) OR (TITLE-ABS-KEY ("computer-adaptive test*")) OR (TITLE-ABS-KEY ("computer-adaptive")) OR (TITLE-ABS-KEY ("computer adaptive")) OR (TITLE-ABS-KEY ("computerized adaptive test*")) OR (TITLE-ABS-KEY ("computerised adaptive test*")) OR (TITLE-ABS-KEY ("cat"))) AND ((TITLE-ABS-KEY ("chronbach*")) OR (TITLE-ABS-KEY ("cronbach*")) OR (TITLE-ABS-KEY ("psychometric properties")) OR (TITLE-ABS-KEY ("psychometric analysis")) OR (TITLE-ABS-KEY ("psychometric evaluation")) OR (TITLE-ABS-KEY ("psychometric characteristics")) OR (TITLE-ABS-KEY ("factor analysis")) OR (TITLE-ABS-KEY ("reliability")) OR (TITLE-ABS-KEY ("reliable")) OR (TITLE-ABS-KEY ("validity")) OR (TITLE-ABS-KEY ("valid")) OR (TITLE-ABS-KEY ("validation")) OR (TITLE-ABS-KEY ("minimal clinically important difference*")) OR (TITLE-ABS-KEY ("clinically meaningful change*")) OR (TITLE-ABS-KEY ("clinically meaningful difference*")) OR (TITLE-ABS-KEY ("responsiveness")) OR (TITLE-ABS-KEY ("minimal important change*")) OR (TITLE-ABS-KEY ("minimal important difference*")) OR (TITLE-ABS-KEY ("translation")) OR (TITLE-ABS-KEY ("translated")) OR (TITLE-ABS-KEY ("cross-cultural")) OR (TITLE-ABS-KEY ("development")))
Outcome: Health-related Quality of Life	AND ("quality of life"[MeSH Terms] OR "perceived health"[Text Word] OR "health status"[Text Word] OR "life satisfaction"[Text Word] OR "well-being"[Text Word] OR "wellbeing"[Text Word] OR "patient reported outcome measures"[MeSH Terms])	AND (TITLE-ABS-KEY ("quality of life")) OR (TITLE-ABS-KEY ("life quality")) OR (TITLE-ABS-KEY ("patient-reported outcome*")) OR (TITLE-ABS-KEY ("hrqol")) OR (TITLE-ABS-KEY ("patient reported outcome*")) OR (TITLE-ABS-KEY ("perceived health")) OR (TITLE-ABS-KEY ("health status")) OR (TITLE-ABS-KEY ("well-being")) OR (TITLE-ABS-KEY ("wellbeing"))
Exclusion string Terwee et al. 2009 + English filter	AND (english[Filter]) NOT (‘‘addresses’’[Publication Type] OR ‘‘biography’’[Publication Type] OR ‘‘case reports’’[Publication Type] OR ‘‘comment’’[Publication Type] OR ‘‘directory’’[Publication Type] OR ‘‘editorial’’[Publication Type] OR ‘‘festschrift’’[Publication Type] OR ‘‘interview’’[Publication Type] OR ‘‘lectures’’[Publication Type] OR ‘‘legal cases’’[Publication Type] OR ‘‘legislation’’[Publication Type] OR ‘‘letter’’[Publication Type] OR ‘‘news’’[Publication Type] OR ‘‘newspaper article’’[Publication Type] OR ‘‘patient education handout’’[Publication Type] OR ‘‘popular works’’[Publication Type] OR ‘‘congresses’’[Publication Type] OR ‘‘consensus development conference’’[Publication Type] OR ‘‘consensus development conference, nih’’[Publication Type] OR ‘‘practice guideline’’[Publication Type]) NOT	AND (LIMIT TO (LANGUAGE , "english")) AND (EXCLUDE (DOCTYPE , "le") OR EXCLUDE (DOCTYPE , "ed")) AND (EXCLUDE (DOCTYPE , "cp"))

	(‘‘animals’’[MeSH Terms] NOT ‘‘humans’’[MeSH Terms])	

Appendix 2. List of European and associated countries

European and associated countries		
Albania	Germany	North-Macedonia
Andorra	Greece	Norway
Armenia	Hungary	Portugal
Austria	Iceland	Romania
Azerbaijan	Ireland	Russia
Belarus	Italy	San Marino
Belgium	Kazakhstan	Serbia
Bulgaria	Latvia	Slovenia
Croatia	Liechtenstein	Slovakia
Cyprus	Lithuania	Spain
Czechia	Luxembourg	Sweden
Denmark	Malta	Switzerland
Estonia	Moldavia	Turkey
Finland	Monaco	Ukraine
France	Montenegro	United Kingdom
Georgia	Netherlands	Vatican City

Appendix 3. Additional search strategy for European validation papers

1. STEP 1:

- Define entry terms for the SPECIFIC QUESTIONNAIRE:
  - Full name (make sure to enter all the different spelling options)
  - Acronym (make sure to enter all the different spelling options)

Example:

EORTC-QLQ-C30	"European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30"
	EORTC-QLQ-C30
	EORTC QLQ-C30
	EORTC QLQ C30
	QLQ C30

- Combine all the entry terms with OR-function:
  - ("European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30") OR (eortc-qlq-c30)) OR (eortc qlq-c30)) OR (eortc qlq c30)) OR (qlq c30)

2. STEP 2:

- Enter search string for POPULATION:
  - ("patient*"[MeSH Terms] OR "Survivors"[MeSH Terms] OR "Palliative Care"[MeSH Terms]) AND ("Neoplasms"[MeSH Terms] OR "Carcinoma"[MeSH Terms] OR "post-cancer"[Title/Abstract] OR "postcancer"[Title/Abstract])
- Enter search string for PSYCHOMETRIC PROPERTIES:
  - ("instrument*"[Title/Abstract] OR "questionnaire*"[Title/Abstract] OR "measur*"[Title/Abstract] OR "rating*"[Title/Abstract] OR "computer*"[Title/Abstract] OR "digital*"[Title/Abstract] OR "computer-adaptive test*"[Title/Abstract] OR "computer adaptive test*"[Title/Abstract] OR "computer adaptive"[Title/Abstract] OR "computer-adaptive"[Title/Abstract] OR "computerized adaptive test*"[Title/Abstract] OR "computerised adaptive test*"[Title/Abstract] OR "CAT"[Title/Abstract]) AND ("chronbach*"[Title/Abstract] OR "cronbach*"[Title/Abstract] OR "psychometric properties"[Title/Abstract] OR "psychometr*"[Title/Abstract] OR "factor analysis"[Title/Abstract] OR "develop*"[Title/Abstract] OR "reliab*"[Title/Abstract] OR "valid*"[Title/Abstract] OR "translat*"[Title/Abstract] OR "cross-cultural"[Title/Abstract] OR "minimal clinically important difference*"[Title/Abstract] OR "minimal important change*"[Title/Abstract] OR "minimal important difference*"[Title/Abstract] OR "clinically meaningful change*"[Title/Abstract] OR "clinically meaningful difference*"[Title/Abstract] OR "responsiveness"[Title/Abstract])

3. STEP 3:

- Combine search strings of POPULATION, PSYCHOMETRIC PROPERTIES and SPECIFIC QUESTIONNAIRE with the AND-function:
  - ((((("European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30") OR (eortc-qlq-c30)) OR (eortc qlq-c30)) OR (eortc qlq c30)) OR (qlq c30)) AND (("instrument*"[Title/Abstract] OR "questionnaire*"[Title/Abstract] OR "measur*"[Title/Abstract] OR "rating*"[Title/Abstract] OR "computer*"[Title/Abstract] OR "digital*"[Title/Abstract] OR "computer-adaptive test*"[Title/Abstract] OR "computer adaptive test*"[Title/Abstract] OR "computer adaptive"[Title/Abstract] OR "computer-

adaptive"[Title/Abstract] OR "computerized adaptive test*"[Title/Abstract] OR "computerised adaptive test*"[Title/Abstract] OR "CAT"[Title/Abstract]) AND ("chronbach*"[Title/Abstract] OR "cronbach*"[Title/Abstract] OR "psychometric properties"[Title/Abstract] OR "psychometr*"[Title/Abstract] OR "factor analysis"[Title/Abstract] OR "develop*"[Title/Abstract] OR "reliab*"[Title/Abstract] OR "valid*"[Title/Abstract] OR "translat*"[Title/Abstract] OR "cross-cultural"[Title/Abstract] OR "minimal clinically important difference*"[Title/Abstract] OR "minimal important change*"[Title/Abstract] OR "minimal important difference*"[Title/Abstract] OR "clinically meaningful change*"[Title/Abstract] OR "clinically meaningful difference*"[Title/Abstract] OR "responsiveness"[Title/Abstract]))) AND (("patient*"[MeSH Terms] OR "Survivors"[MeSH Terms] OR "Palliative Care"[MeSH Terms]) AND ("Neoplasms"[MeSH Terms] OR "Carcinoma"[MeSH Terms] OR "post-cancer"[Title/Abstract] OR "postcancer"[Title/Abstract]))

4. ***STEP 4:***

- Find search string (which is used to gather the articles for our systematic review but remove English filter)
    - (((((("instrument*"[Title/Abstract] OR "questionnaire*"[Title/Abstract] OR "measur*"[Title/Abstract] OR "rating*"[Title/Abstract] OR "computer*"[Title/Abstract] OR "digital*"[Title/Abstract] OR "computer-adaptive test*"[Title/Abstract] OR "computer adaptive test*"[Title/Abstract] OR "computer adaptive"[Title/Abstract] OR "computer-adaptive"[Title/Abstract] OR "computerized adaptive test*"[Title/Abstract] OR "computerised adaptive test*"[Title/Abstract] OR "CAT"[Title/Abstract]) AND ("chronbach*"[Title/Abstract] OR "cronbach*"[Title/Abstract] OR "psychometric properties"[Title/Abstract] OR "psychometr*"[Title/Abstract] OR "factor analysis"[Title/Abstract] OR "develop*"[Title/Abstract] OR "reliab*"[Title/Abstract] OR "valid*"[Title/Abstract] OR "translat*"[Title/Abstract] OR "cross-cultural"[Title/Abstract] OR "minimal clinically important difference*"[Title/Abstract] OR "minimal important change*"[Title/Abstract] OR "minimal important difference*"[Title/Abstract] OR "clinically meaningful change*"[Title/Abstract] OR "clinically meaningful difference*"[Title/Abstract] OR "responsiveness"[Title/Abstract])) AND ("quality of life"[MeSH Terms] OR "perceived health"[Text Word] OR "health status"[Text Word] OR "life satisfaction"[Text Word] OR "well-being"[Text Word] OR "wellbeing"[Text Word] OR "patient reported outcome measures"[MeSH Terms])) AND (("patient*"[MeSH Terms] OR "Survivors"[MeSH Terms] OR "Palliative Care"[MeSH Terms]) AND ("Neoplasms"[MeSH Terms] OR "Carcinoma"[MeSH Terms] OR "post-cancer"[Title/Abstract] OR "postcancer"[Title/Abstract])) AND (english[Filter])) NOT ((‘‘animals’’[MeSH Terms] NOT ‘‘humans’’[MeSH Terms]))) NOT ((((("instrument*"[Title/Abstract] OR "questionnaire*"[Title/Abstract] OR "measur*"[Title/Abstract] OR "rating*"[Title/Abstract] OR "computer*"[Title/Abstract] OR "digital*"[Title/Abstract] OR "computer-adaptive test*"[Title/Abstract] OR "computer adaptive test*"[Title/Abstract] OR "computer adaptive"[Title/Abstract] OR "computer-adaptive"[Title/Abstract] OR "computerized adaptive test*"[Title/Abstract] OR "computerised adaptive test*"[Title/Abstract] OR "CAT"[Title/Abstract]) AND ("chronbach*"[Title/Abstract] OR "cronbach*"[Title/Abstract] OR "psychometric properties"[Title/Abstract] OR "psychometr*"[Title/Abstract] OR "factor analysis"[Title/Abstract] OR "develop*"[Title/Abstract] OR "reliab*"[Title/Abstract] OR "valid*"[Title/Abstract] OR "translat*"[Title/Abstract] OR "cross-cultural"[Title/Abstract] OR "minimal clinically important difference*"[Title/Abstract] OR "minimal important change*"[Title/Abstract] OR "minimal important difference*"[Title/Abstract] OR "clinically meaningful change*"[Title/Abstract] OR "clinically meaningful difference*"[Title/Abstract] OR "responsiveness"[Title/Abstract])) AND ("quality of life"[MeSH Terms] OR "perceived health"[Text Word] OR "health status"[Text Word] OR "life satisfaction"[Text Word] OR "well-being"[Text Word] OR "wellbeing"[Text Word] OR "patient reported outcome measures"[MeSH Terms])) AND (("patient*"[MeSH Terms] OR

"Survivors"[MeSH Terms] OR "Palliative Care"[MeSH Terms]) AND ("Neoplasms"[MeSH Terms] OR "Carcinoma"[MeSH Terms] OR "post-cancer"[Title/Abstract] OR "postcancer"[Title/Abstract])) AND ((address[Filter] OR biography[Filter] OR casereports[Filter] OR comment[Filter] OR congress[Filter] OR consensusdevelopmentconference[Filter] OR consensusdevelopmentconferencenih[Filter] OR directory[Filter] OR editorial[Filter] OR festschrift[Filter] OR interview[Filter] OR lecture[Filter] OR legalcase[Filter] OR legislation[Filter] OR letter[Filter] OR news[Filter] OR newspaperarticle[Filter] OR patienteducationhandout[Filter] OR practiceguideline[Filter]))

5. ***STEP 5:***

- Combine search string of STEP 3 (POPULATION AND PSYCHOMETRIC PROPERTIES AND SPECIFIC QUESTIONNAIRE) and STEP 4 (ENTIRE search string) with NOT-function:
    - (((((("European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30") OR (eortc-qlq-c30)) OR (eortc qlq-c30)) OR (eortc qlq c30)) OR (qlq c30)) AND (("instrument*"[Title/Abstract] OR "questionnaire*"[Title/Abstract] OR "measur*"[Title/Abstract] OR "rating*"[Title/Abstract] OR "computer*"[Title/Abstract] OR "digital*"[Title/Abstract] OR "computer-adaptive test*"[Title/Abstract] OR "computer adaptive test*"[Title/Abstract] OR "computer adaptive"[Title/Abstract] OR "computer-adaptive"[Title/Abstract] OR "computerized adaptive test*"[Title/Abstract] OR "computerised adaptive test*"[Title/Abstract] OR "CAT"[Title/Abstract]) AND ("chronbach*"[Title/Abstract] OR "cronbach*"[Title/Abstract] OR "psychometric properties"[Title/Abstract] OR "psychometr*"[Title/Abstract] OR "factor analysis"[Title/Abstract] OR "develop*"[Title/Abstract] OR "reliab*"[Title/Abstract] OR "valid*"[Title/Abstract] OR "translat*"[Title/Abstract] OR "cross-cultural"[Title/Abstract] OR "minimal clinically important difference*"[Title/Abstract] OR "minimal important change*"[Title/Abstract] OR "minimal important difference*"[Title/Abstract] OR "clinically meaningful change*"[Title/Abstract] OR "clinically meaningful difference*"[Title/Abstract] OR "responsiveness"[Title/Abstract]))) AND (("patient*"[MeSH Terms] OR "Survivors"[MeSH Terms] OR "Palliative Care"[MeSH Terms]) AND ("Neoplasms"[MeSH Terms] OR "Carcinoma"[MeSH Terms] OR "post-cancer"[Title/Abstract] OR "postcancer"[Title/Abstract]))) NOT ((((((("instrument*"[Title/Abstract] OR "questionnaire*"[Title/Abstract] OR "measur*"[Title/Abstract] OR "rating*"[Title/Abstract] OR "computer*"[Title/Abstract] OR "digital*"[Title/Abstract] OR "computer-adaptive test*"[Title/Abstract] OR "computer adaptive test*"[Title/Abstract] OR "computer adaptive"[Title/Abstract] OR "computer-adaptive"[Title/Abstract] OR "computerized adaptive test*"[Title/Abstract] OR "computerised adaptive test*"[Title/Abstract] OR "CAT"[Title/Abstract]) AND ("chronbach*"[Title/Abstract] OR "cronbach*"[Title/Abstract] OR "psychometric properties"[Title/Abstract] OR "psychometr*"[Title/Abstract] OR "factor analysis"[Title/Abstract] OR "develop*"[Title/Abstract] OR "reliab*"[Title/Abstract] OR "valid*"[Title/Abstract] OR "translat*"[Title/Abstract] OR "cross-cultural"[Title/Abstract] OR "minimal clinically important difference*"[Title/Abstract] OR "minimal important change*"[Title/Abstract] OR "minimal important difference*"[Title/Abstract] OR "clinically meaningful change*"[Title/Abstract] OR "clinically meaningful difference*"[Title/Abstract] OR "responsiveness"[Title/Abstract])) AND ("quality of life"[MeSH Terms] OR "perceived health"[Text Word] OR "health status"[Text Word] OR "life satisfaction"[Text Word] OR "well-being"[Text Word] OR "wellbeing"[Text Word] OR "patient reported outcome measures"[MeSH Terms])) AND (("patient*"[MeSH Terms] OR "Survivors"[MeSH Terms] OR "Palliative Care"[MeSH Terms]) AND ("Neoplasms"[MeSH Terms] OR "Carcinoma"[MeSH Terms] OR "post-cancer"[Title/Abstract] OR "postcancer"[Title/Abstract])) AND (english[Filter])) NOT ((‘‘animals’’[MeSH Terms] NOT ‘‘humans’’[MeSH Terms]))) NOT ((((("instrument*"[Title/Abstract] OR "questionnaire*"[Title/Abstract] OR "measur*"[Title/Abstract] OR "rating*"[Title/Abstract] OR "computer*"[Title/Abstract] OR "digital*"[Title/Abstract] OR "computer-adaptive

test*"[Title/Abstract] OR "computer adaptive test*"[Title/Abstract] OR "computer adaptive"[Title/Abstract] OR "computer-adaptive"[Title/Abstract] OR "computerized adaptive test*"[Title/Abstract] OR "computerised adaptive test*"[Title/Abstract] OR "CAT"[Title/Abstract]) AND ("chronbach*"[Title/Abstract] OR "cronbach*"[Title/Abstract] OR "psychometric properties"[Title/Abstract] OR "psychometr*"[Title/Abstract] OR "factor analysis"[Title/Abstract] OR "develop*"[Title/Abstract] OR "reliab*"[Title/Abstract] OR "valid*"[Title/Abstract] OR "translat*"[Title/Abstract] OR "cross-cultural"[Title/Abstract] OR "minimal clinically important difference*"[Title/Abstract] OR "minimal important change*"[Title/Abstract] OR "minimal important difference*"[Title/Abstract] OR "clinically meaningful change*"[Title/Abstract] OR "clinically meaningful difference*"[Title/Abstract] OR "responsiveness"[Title/Abstract])) AND ("quality of life"[MeSH Terms] OR "perceived health"[Text Word] OR "health status"[Text Word] OR "life satisfaction"[Text Word] OR "well-being"[Text Word] OR "wellbeing"[Text Word] OR "patient reported outcome measures"[MeSH Terms])) AND (("patient*"[MeSH Terms] OR "Survivors"[MeSH Terms] OR "Palliative Care"[MeSH Terms]) AND ("Neoplasms"[MeSH Terms] OR "Carcinoma"[MeSH Terms] OR "post-cancer"[Title/Abstract] OR "postcancer"[Title/Abstract])) AND ((address[Filter] OR biography[Filter] OR casereports[Filter] OR comment[Filter] OR congress[Filter] OR consensusdevelopmentconference[Filter] OR consensusdevelopmentconferencenih[Filter] OR directory[Filter] OR editorial[Filter] OR festschrift[Filter] OR interview[Filter] OR lecture[Filter] OR legalcase[Filter] OR legislation[Filter] OR letter[Filter] OR news[Filter] OR newspaperarticle[Filter] OR patienteducationhandout[Filter] OR practiceguideline[Filter])))

6. ***STEP 6: Apply "English" filter***

7. ***STEP 7: Assess and screen articles for the predefined in- and exclusion criteria***

Appendix 4. Overview of the data extraction for the PROMs measurement properties

Measurement property	Data extracted
Development/ Content validity	- Level of analysis: scale/subscale - Methodological approach for concept elicitation, PROM design, relevance, comprehensiveness and comprehensibility
Structural validity/ Unidimensionality	- Level of analysis: scale/subscale - Statistical approach and related sample size: EFA, CFA or IRT - Final model and fit indexes: CFI, TLI, RMSEA (90%CI) SRMR or WRMR
Internal consistency	- Level of analysis: scale/subscale - Statistical approach and related sample size - Internal consistency reliability coefficients: Cronbach alpha, McDonald Omega, KR-20, SE(θ)
Cross-cultural validity/ Measurement invariance	- Level of analysis: scale/subscale - Statistical approach and related sample size - Group variable under investigation (e.g. country, age, gender,...) with its observed differences
Reliability	- Level of analysis: scale/subscale - Statistical approach and related sample size - Type of reliability: test-retest, inter-rater, intra-rater, parallel forms - Correlation coefficients: ICC, Spearman, Pearson, Kappa or weighted Kappa
Measurement error	- Level of analysis: scale/subscale - Statistical approach and related sample size - Standard Error of Measurement, Limits of Agreement, Smallest Detectable Change, Minimal Important Change
Construct validity with other PROM	- Level of analysis: scale/subscale - Statistical approach and related sample size - Comparator + formulated hypotheses - Correlation coefficients or effect sizes
Convergent/ divergent validity within PROM	- Level of analysis: scale/subscale - Statistical approach and related sample size - Formulated hypotheses - Correlation coefficients
Known-group comparison	- Level of analysis: scale/subscale - Statistical approach and related sample size - Formulated hypotheses - Group variable + defined subgroups with observed differences

Abbreviations: CFA = Confirmatory Factor Analysis; CFI = Comparative Fit Index; IRT = Item Response Theory; RMSEA = Root Mean Square Error of Approximation; SDC = Smallest Detectable Change; SRMR: Standardized Root Mean Residuals; TLI: Tucker-Lewis Index; WRMR: Weighted Root Mean Residuals

Appendix 5: Overview of adjustments made to the Risk of Bias rating of COSMIN Guidelines

Psychometric property	Criteria	Adjustment made
PROM development (Box 1)	23	Inadequate rating was removed from the response options.
	25	Adequate and doubtful rating were removed from the response options.
	26	Doubtful rating was removed and inadequate was defined as “NO or not clear (SKIP items 27-35)”.
	35	Adequate and doubtful rating were removed from the response options.
Content validity (Box 2)	6	Inadequate rating was removed from the response options.
	13	Inadequate rating was removed from the response options.
	20	Inadequate rating was removed from the response options.
	25	Inadequate rating was removed from the response options.
	30	Inadequate rating was removed from the response options.
Structural validity (Box 3)	2	Adequate rating was removed from the response options.
Internal consistency (Box 4)	5	Criteria 5 was removed from the Risk of bias assessment.
Cross-cultural validity & Measurement invariance (Box 5)	4	Criteria 4 was removed from the Risk of bias assessment.
Reliability (Box 6)	1-3	Not applicable rating was added to the response options.
Measurement error (Box 7)	6	Adequate rating was removed from the response options.
Construct validity (with other PROM) (hypothesis testing) (Box 9.a)	4	Inadequate rating was removed from the response options.
	1-4	Not applicable rating was added to the response options.
Construct validity (Known-group comparison) (Box 9.b)	7	Inadequate rating was removed from the response options.
	5-7	Not applicable rating was added to the response options.
Construct validity (convergent & divergent validity)	1	Criteria 3 of Box 9.a was introduced.

Appendix 6: The 10 criteria for good content validity

		PROM development study		Content validity study		Rating of reviewers
1	**+**	Construct of interest is clearly described **(criterion 1 of box 1A = very good) AND** origin of construct is clear **(criterion 2 of box 1A = very good) AND** the is evidence from concept elicitation, literature or professionals that ≥85% of the items refer to construct of interest	**+**	Professionals rated the relevance of items for the construct of interest as sufficient **(criteria 22-26 of box 2D = very good, adequate or doubtful)** and found ≥85% of the items relevant for the construct		Reviewers consider ≥85% of the items relevant for the construct of interest
	-	Quality is inadequate (≥ 1 of the 3 (+)-criteria is not fulfilled)	**-**	Professionals were not involved in the content validity study OR rated <85% of the items of the PROM relevant for the construct		Reviewers consider <85% of the items relevant for the construct of interest
	?	No(t enough) information available to score a (+) or (-)	**?**	No(t enough) information available to score a (+) or (-)		
2	**+**	Target population of interest is clearly described **(criterion 3 of box 1A = very good) AND** representative patients were involved in the elicitation of relevant items **(criterion 5 of box 1A = very good or adequate) AND** concept elicitation was not inadequate **(criteria 6-13 of box 1A = very good, adequate or doubtful)**	**+**	Patients rated the relevance of items for the construct of interest as sufficient **(criteria 1-7 of box 2A = very good, adequate or doubtful)** and found ≥85% of the items relevant for them		Reviewers consider ≥85% of the items relevant for the population of interest
	-	Quality is inadequate (≥ 1 of the 3 (+)-criteria is not fulfilled)	**-**	Patients were not involved in the content validity study OR rated <85% of the items of the PROM relevant for them		Reviewers consider <85% of the items relevant for the population of interest
	?	No(t enough) information available to score a (+) or (-) OR doubtful whether study was performed in a sample representing the target population	**?**	No(t enough) information available to score a (+) or (-)		
3	**+**	The context of use of interest is clearly described **(criterion 4 of box 1A = very good)**	**+**	Professionals rated the relevance of items for the context of use as sufficient **(criteria 22-26 of box 2D = very good, adequate or doubtful)** and found ≥85% of the items relevant for the context of use		Reviewers consider ≥85% of the items relevant for the context of use of interest
	-	The context of use of interest is not clearly described **(criterion 4 of box 1A = doubtful)**	**-**	Professionals were not involved in the content validity study OR rated <85% of the items of the PROM relevant for the context of use		Reviewers consider <85% of the items relevant for the context of use of interest
	?	No(t enough) information available to score a (+) or (-)	**?**	No(t enough) information available to score a (+) or (-)		
4	**+**	A justification is provided for the response options	**+**	Patients or professionals rated the appropriateness of the response options as sufficient **(criteria 1-7 of box 2A or criteria 22-26 of box 2D = very good, adequate or doubtful)** and found ≥85% of the response options relevant		Reviewers consider ≥85% of the response options appropriate for the construct, population, and context of use of interest
	-	No justification was provided for the response options	**-**	Patients or professionals were not involved in the content validity study OR rated <85% of the response options of the PROM relevant		Reviewers consider <85% of the response options appropriate for the construct, population, and context of use of interest
	?	No(t enough) information available to score a (+) or (-)	**?**	No(t enough) information available to score a (+) or (-)		
5	**+**	A justification is provided for the recall period	**+**	Patients or professionals rated the appropriateness of the recall period as sufficient **(criteria 1-7 of box 2A or criteria 22-26 of box 2D = very good, adequate or doubtful)** and found the recall period relevant	**+**	Reviewers consider the recall period appropriate for the construct, population and context of use of interest for **≥85% of the items**.
	-	No justification is provided for the recall period	**-**	Patients or professionals were not involved in the content validity study OR rated the recall period for <85% of the items of the PROM relevant	**-**	Reviewers consider the recall period **only partially** (<85% of the items) **OR not** appropriate for the construct, population and context of use of interest.
	?	No(t enough) information available to score a (+) or (-)	**?**	No(t enough) information available to score a (+) or (-)		

6	**+**	Patients were asked about the comprehensiveness of the PROM in concept elicitation phase or cognitive interview **(criteria 6-13 of box 1A or criteria 26-35 of box 1B = very good, adequate or doubtful) AND** no key concepts were missing	**+**	Patients or professionals were asked about the comprehensiveness of the PROM **(criteria 8-14 of box 2B or criteria 27-31 of box 2E = very good, adequate or doubtful) AND** no key concepts were missing		Reviewers consider the PROM **comprehensive** for the construct, population and context of use of interest for **≥85% of the items**.
	-	Quality is inadequate (≥ 1 of the 2 (+)-criteria is not fulfilled)	**-**	Patients or professionals were not involved in the content validity study OR quality is inadequate (≥ 1 of the 2 (+)-criteria is not fulfilled)		Reviewers consider the PROM **only partially** (<85% of the items) **OR not comprehensive** for the construct, population and context of use of interest comprehensive (<85% of the items)
	?	No(t enough) information available to score a (+) or (-)	**?**	No(t enough) information available to score a (+) or (-)		
7	**+**	Patients were asked about the comprehensibility of the instructions (including recall period) in cognitive interview **(criteria 16-25 of box 1B = very good, adequate or doubtful) AND** problems were adequately addressed	**+**	Patients were asked about the comprehensibility of the instructions (including recall period) **(criteria 15-21 of box 2C = very good, adequate or doubtful) AND** no important problems were found		
	-	Quality is inadequate (≥ 1 of the 2 (+)-criteria is not fulfilled)	**-**	Patients were not involved in the content validity study OR quality is inadequate (≥ 1 of the 2 (+)-criteria is not fulfilled)		
	?	No(t enough) information available to score a (+) or (-)	**?**	No(t enough) information available to score a (+) or (-)		
8	**+**	Patients were asked about the comprehensibility of the items and response options (including wording of items and response options) in cognitive interview **(criteria 16-25 of box 1B = very good, adequate or doubtful) AND** problems were adequately addressed	**+**	Patients were asked about the comprehensibility of the items and response options (including wording of items and response options) **(criteria 15-21 of box 2C = very good, adequate or doubtful) AND** no important problems were found for ≥85% of the items and response options		
	-	Quality is inadequate (≥ 1 of the 2 (+)-criteria is not fulfilled)	**-**	Patients were not involved in the content validity study OR quality is inadequate (≥ 1 of the 2 (+)-criteria is not fulfilled)		
	?	No(t enough) information available to score a (+) or (-)	**?**	No(t enough) information available to score a (+) or (-)		
9	**+**		**+**			Reviewers consider ≥85% of the items and response options appropriately worded
	-		**-**			Reviewers consider <85% of the items and response options appropriately worded
	?		**?**			
10	**+**		**+**			Reviewers consider ≥85% of the response options matching the questions
	-		**-**			Reviewers consider <85% of the response options matching the questions
	?		**?**			

Appendix 7: Calculation of the overall relevance, comprehensiveness and comprehensibility rating per study.

		PROM development		Content validity	Reviewer rating
Relevance rating	**+**	Criteria 1 and 2 are rated sufficient (+) AND ≥2 of remaining 3 items are rated sufficient (+)	**+**	Criteria 1 and 2 are rated sufficient (+) AND ≥2 of remaining 3 items are rated sufficient (+)	Criteria 1 and 2 are rated sufficient (+) AND ≥2 of remaining 3 items are rated sufficient (+)
	-	Criteria 1 and 2 are rated insufficient (-) AND ≥2 of remaining 3 items are rated insufficient (-)	**-**	Criteria 1 and 2 are rated insufficient (-) AND ≥2 of remaining 3 items are rated insufficient (-)	Criteria 1 and 2 are rated insufficient (-) AND ≥2 of remaining 3 items are rated insufficient (-)
	?	≥2 criteria are rated indeterminate (?)	**?**	≥2 criteria are rated indeterminate (?)	
	±	All other situations	**±**	All other situations	All other situations
Comprehensiveness rating		Rating of criterion 6		Rating of criterion 6	Rating of criterion 6
Comprehensibility rating	**+**	Criterion 8 = sufficient (+) AND criterion 7 = sufficient (+) or indeterminate (?)	**+**	Criterion 8 = sufficient (+) AND criterion 7 = sufficient (+) or indeterminate (?)	Criteria 9 and 10 are rated sufficient (+)
	-	Criterion 8 = insufficient (-)	**-**	Criterion 8 = insufficient (-)	Criteria 9 and 10 are rated insufficient (-)
	?	Criterion 8 = indeterminate (?)	**?**	Criterion 8 = indeterminate (?)	
	±	Criterion 8 = sufficient (+) AND criterion 7 = insufficient (-)	**±**	Criterion 8 = sufficient (+) AND criterion 7 = insufficient (-)	One criterion = sufficient (+) AND one criterion = insufficient (-)

Appendix 8: Calculation of the overall relevance, comprehensiveness and comprehensibility rating per PROM

PROM development	Content validity	Rating reviewer	Overall RELEVANCE COMPREHENSIVENES COMPREHENSIBILITY rating
+	+	+	+
+	+	±	+
+	+	-	+
+	-	+	±
+	-	±	±
+	-	-	-
+	?	+	+
+	?	±	±
+	?	-	±
+	±	+	±
+	±	±	±
+	±	-	±
-	+	+	+
-	+	±	±
-	+	-	±
-	-	+	-
-	-	±	-
-	-	-	-
-	?	+	±
-	?	±	±
-	?	-	-
-	±	+	±
-	±	±	±
-	±	-	±
?	+	+	+
?	+	±	±
?	+	-	±
?	-	+	±
?	-	±	±
?	-	-	-
?	?	+	+
?	?	±	±
?	?	-	-
?	±	+	±
?	±	±	±
?	±	-	±
±	+	+	+
±	+	±	+
±	+	-	±
±	-	+	±
±	-	±	-
±	-	-	-
±	?	+	±
±	?	±	±
±	?	-	±
±	±	+	±
±	±	±	±
±	±	-	-

Appendix 9: Calculation of the overall content validity rating

Overall RELEVANCE rating	Overall COMPREHENSIVENESS rating	Overall COMPREHENSIBILITY rating	Overall CONTENT VALIDITY rating
+	+	+	+
+	+	±	+
+	+	-	±
+	-	+	±
+	-	±	±
+	-	-	±
+	±	+	+
+	±	±	±
+	±	-	±
-	+	+	±
-	+	±	±
-	+	-	±
-	-	+	±
-	-	±	-
-	-	-	-
-	±	+	±
-	±	±	±
-	±	-	-
±	+	+	+
±	+	±	±
±	+	-	±
±	-	+	±
±	-	±	±
±	-	-	-
±	±	+	±
±	±	±	±
±	±	-	±

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2024-088716.R1
Article Type:	Protocol
Date Submitted by the Author:	03-Mar-2025
Complete List of Authors:	Leysen, Laurence; European Organisation for Research and Treatment of Cancer, Quality of Life Marticic Giljevic, Karla; European Organisation for Research and Treatment of Cancer, Quality of Life Piccinin, Claire; European Organisation for Research and Treatment of Cancer, Quality of Life Shkodra, Morena; Fondazione IRCCS Istituto Nazionale dei Tumori Pe, Madeline; European Organisation for Research and Treatment of Cancer, Petersen, Morten; Bispebjerg Hospital, The Research Unit, Department of Palliative Medicine Apolone, Giovanni; Fondazione IRCCS Istituto Nazionale dei Tumori Brunelli, Cinzia; Fondazione IRCCS Istituto Nazionale dei Tumori Lombardo, Claudio; Organisation of European Cancer Institutes – European Economic Interest Grouping Velikova, Galina; University of Leeds, Leeds Institute of Medical Research Gilbert, Alexandra; University of Leeds, Leeds Institute of Medical Research Pravettoni, Gabriella; Istituto Europeo di Oncologia, Applied Research Division for Cognitive and Psychological Science; Universita degli Studi di Milano, Department of Oncology and Hemato-oncology Ciliberto, Gennaro; Regina Elena Institute Groenvold, Mogens; University of Copenhagen, Palliative Medicine Ferrer, Montserrat; Hospital del Mar Research Institute; CIBERESP Bos, Nanne; Nederlands Instituut voor Onderzoek van de Gezondheidszorg, Zorg vanuit patiëntperspectief Pietrobon, Ricardo; SporeData OU Couespel, Norbert; European CanCer Organisation Caraceni, A; Fondazione IRCCS Istituto Nazionale dei Tumori, Palliative Care, Pain Therapy and Rehabilitation Guren, Marianne; Oslo University Hospital, Oncology Sirven, Aude; UNICANCER Vachon, Hugo; European Organisation for Research and Treatment of Cancer, Quality of Life Department
Primary Subject Heading:	Oncology

Secondary Subject Heading:	Global health, Research methods, Qualitative research, Patient-centred medicine
Keywords:	ONCOLOGY, Patient Reported Outcome Measures, Psychometrics, QUALITATIVE RESEARCH, Systematic Review, Quality of Life

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

Leysen, L.[1], Martičić Giljević, K.[1], Piccinin, C.[1], Shkodra, M.[2], Pe., M.[1], Petersen, M.[3], Apolone, G.[4], Brunelli, C.[4], Lombardo, C.[5], Velikova, G.[6], Gilbert, A.[6], Pravettoni, G.[7], Ciliberto, G.[8], Groenvold, M.[3], Ferrer, M.[9], Bos, N.[10], Pietrobon, R.[11], Couespel, N.[12], Caraceni, A.[13], Guren, M.G.[14], Sirven, A.[15], & Vachon, H. on behalf of the EUonQOL consortium

Author affiliations:

[1] Quality of Life Department, European Organisation for Research and Treatment of Cancer, Brussels, Belgium

[2] Palliative Care, Pain Therapy and Rehabilitation Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milano, Italy

[3] The Research Unit, Department of Palliative Medicine, Bispebjerg Hospital, University of Copenhagen, Copenhagen, Denmark

[4] Fondazione IRCCS Istituto Nazionale Dei Tumori di Milano, Milan, Italy

[5] Organisation of European Cancer Institutes, Brussels, Belgium

[6] Leeds Institute of Cancer and Pathology, University of Leeds, St James's Hospital, Leeds, UK

[7] Applied Research Division for Cognitive and Psychological Science, IEO, European Institute of Oncology IRCCS, Milan, Italy

[8] IRCCS National Cancer Institute Regina Elena, Rome, Italy

[9] Hospital del Mar Research Institute, Barcelona, Spain

[10] Netherlands Institute for Health Services Research (Nivel), Utrecht, The Netherlands

[11] SporeData OÜ, Tallinn, Estonia

[12] European Cancer Organisation (ECO), Brussels, Belgium

[13] Università Degli Studi Di Milano, Milan, Italy

[14] Department of Oncology, Oslo University Hospital, Oslo, Norway

[15] Unicancer, Paris, France

Correspondence: Hugo Vachon, Quality of Life Department, European Research for Treatment and Research of Cancer, hugo.vachon@eortc.org, address: Avenue E. Mounier 83/11, 1200 Brussels, Belgium.

Protocol registration number: CRD42023418616

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

Abstract

INTRODUCTION: Over the past decades, there has been increasing recognition that assessing cancer patients' 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 summarized. The Grading of Recommendations, Assessment, Development and Evaluations (GRADE) guidelines will be used to grade and summarize 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 cancer patients and survivors in Europe.

Protocol 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 (HRQoL) across the European cancer continuum at a subscale level.
- This systematic review will follow the highest methodological standards, i.e., the Preferred Reporting Items for Systematic Review and Meta-Analyses protocols (PRISMA-P), the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) 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 (i.e. meta-analysis) will be performed and the findings will be summarized qualitatively.
- The systematic review will specifically include PROMs validated in European cancer patients 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.

1 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 cancer patients’ 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 (e.g., Europe’s Beating Cancer plan [7]). However, HRQoL is often inaccurately assessed by health care providers (HCPs) 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 cancer patients, a wide array of PROMs is currently available, ranging from generic (e.g., SF-36, EQ-5D-5L) to cancer-specific (e.g., EORTC QLQ-C30, FACT-G) and tumour-specific tools (e.g., EORTC QLQ-BR23, 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 cancer patients have been published. Most of them focused on PROMs measuring HRQoL in a specific type of cancer (e.g., breast cancer, prostate cancer, etc.) [12-23] or cancer population (e.g., cancer survivors, advanced cancer, palliative patients, etc.) [14,24-26]. Several of these reviews focused on PROMs evaluating one specific HRQoL-related construct (e.g., depression, fatigue, pain, etc.) [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 were 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 cancer patients and survivors, to make objective recommendations on the most suitable PROM to use in these populations.

2 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 a 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].

2.1 Search strategy

A systematic search will be performed in the bibliographic databases MEDLINE and PubMed Central (through PubMed) and EMBASE (through Scopus) without publication date restriction up to March 2024. The search strategy will be based on the PICOM acronym [36] in which the population will be represented by cancer patients and survivors, the outcome by health-related quality of life and the methods by psychometric properties. No comparator or intervention will be used. Both 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 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 Appendix 1.

2.2 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 standardized matter. 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.

2.2.1 PROM selection

To be included PROMs will need to meet following criteria:

1) PROMs must be self-administered questionnaire (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 (i.e., based on a multidimensional model of HRQoL) and be applicable across cancer types. This excludes tools designed to assess a specific HRQoL subdomain (e.g., exclusively assessing physical functioning) or cancer site (e.g., 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 (i.e., evidence of content validity, structural validity, or construct validity) for use in the target population of European cancer patients or survivors (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 (Appendix 3). If no evidence of validity among European cancer patients or survivors can be retrieved after the additional search, the PROM and its related articles will be excluded.

2.2.2 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 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 article and reviews will not be considered.
3) Studies must be performed in adult European cancer patients or survivors (mean age ≥ 21 years and not defined as Adolescents and Young Adults (AYA)). Articles including "mixed samples" (i.e., European cancer patients and non-cancer patients) will only be included if separate results are provided for the cancer patients group. Studies involving both European and non-European cancer patients, 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 cancer patients will also be excluded.

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

2.3 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 customized 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, non-cancer patients, cancer patients undergoing curative treatment, cancer patients 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 Appendix 4.

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

2.4 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 (i.e., structural validity, internal consistency, and cross-cultural validity/measurement invariance), reliability, measurement error and construct validity (i.e., 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:

2.4.1 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 Appendix 5.

2.4.2 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 (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] (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. Appendix 8 provides a detailed overview of this overall rating process. Last, an overall “sufficient” (+), “insufficient” (−) or “inconsistent” (±) content validity rating will be calculated, by aggregating the overall relevance, comprehensiveness and comprehensibility rating. 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 evidence across studies will be summarized 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 summarizing 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 summarized 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).

2.4.3 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 4 quality levels: “high”, “moderate”, “low”, or “very low”. Starting from high-quality level, 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].

2.5 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 summarized 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 summarized 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.

2.6 Recommendations

PROMs with sufficient content validity (i.e., rated "±" or higher) and at least low-quality evidence (i.e., 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.

2.7 Patient and public involvement

Currently, it is expected that researchers actively involve patients, healthcare professionals and 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 summarized 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, i.e., 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.

Additional information

Contributors: HV conceived the idea. HV, KM and LL planned and designed the study protocol, the data extraction and the data assessment. HV, KM and LL wrote the first draft of the protocol. CP, MaP, MS and MoP provided critical insights. GA, CB, CL, GV, AG, GP, GC, MG, MF, NB, RP, NC, AC, MG & AS have approved and contributed to the final written manuscript.

Funding: This work was supported by the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement Number 101096362.

Competing interests: None to declare.

Patient and public involvement: Patients and/or public will be involved in the conduction, reporting or dissemination plans of this research. Refer to Methods section for further details.

Patient consent for publication: Not required.

References

1. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA* 1995;273(1):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;10:1–12.

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(1):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(4):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(12):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(2):573–93.

7. European Commission. *EUR-Lex 2021* [cited 2023 Sep 5]. Europe's Beating Cancer plan. Available from: 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(8):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*; Availability. 2009.

11. Gotay CC. Assessing cancer-related quality of life across a spectrum of applications. *J Natl Cancer Inst Monogr* 2004;(33):126–33.

12. Amarsheda S, Bhise AR. Systematic review of cancer-related fatigue instruments in breast cancer patients. *Palliat Support Care* 2022;20(1):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* 2019;28(3):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(5):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(1):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(6):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(5):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(Suppl 3):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(5):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(20).

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(1):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(1):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(7):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(8):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(7):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(2):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(9):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(1):93–117.

29. Curcio KR. Instruments for Assessing Chemotherapy-Induced Peripheral Neuropathy: A Review of the Literature. *Clin J Oncol Nurs* 2016;20(2):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(4):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(3):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(5):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(10):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):1.

35. Ouzzani M, Hammady H, Fedorowicz Z, et al. Rayyan-a web and mobile app for systematic reviews. *Syst Rev* 2016;5(1):210.

36. Higgins JPT, Thomas J, Chandler J, et al. *Cochrane Handbook for Systematic Reviews of Interventions version 6.3* [Internet]. 2023. Available from: 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(8):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.

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(5):1159–70.

41. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull* 1959;56(2):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(4):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(6):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;349.

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(5).

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 interrelatedness 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: e.g., 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 (i.e., raters or responders) on different occasions (intra-rater)
Measurement error	The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured
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

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 unidimensionality: CFI or TLI or comparable measure >0.95 OR RMSEA <0.06 OR SRMR <0.08 AND • No violation of local independence: residual correlations among the items after controlling for the dominant factor <0.20 OR Q3's < 0.37 AND • No violation of monotonicity: adequate looking graphs OR item scalability >0.30 AND • Adequate model fit: IRT: $\chi 2$ >0.01 Rasch: infit and outfit mean squares ≥ 0.5 and ≤ 1.5 OR Z-standardized 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) ≥ 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 (e.g., because there was only 1 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 R2 < 0.02)
	±	Results are inconsistent across studies
	-	Important differences between group factors OR DIF was 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
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

Abbreviations: + = sufficient results; - = insufficient results; ± = inconsistent results; ? = indeterminate results; CFA = Confirmatory Factor Analysis; CFI = Comparative Fit Index; CTT = Classical Test Theory; DIF = Differential Item Functioning; LoA = Limits of Agreement; IRT = Item Response Theory; MIC = Minimal Important Change; MID: Minimal Important Difference; MCID = Minimal Clinical Important Difference; RMSEA = Root Mean Square Error of Approximation; SDC = Smallest Detectable Change; SRMR: Standardized Root Mean Residuals; TLI: Tucker-Lewis Index.

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 hypothesized subscale (corrected for overlap) is higher than its correlation with the other subscales

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

	QUALITY OF EVIDENCE: starting point is always HIGH HIGH MODERATE LOW VERY LOW	
Risk of bias	- 1: Serious	Content validity study is of doubtful quality. The content validity rating of content validity study is insufficient (-) OR indeterminate (?) OR inconsistent (±)
	- 2: Very serious	No content validity study OR content validity study of insufficient quality (-) **AND** 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	Content validity study was performed in a cancer population but not representative of the population of interest (e.g. head & neck cancer patients versus cancer patients, palliative questionnaire assessed in non-palliative cancer patients)
	- 2: Very serious	Content validity study was performed in a non-cancer population.

Table 5 COSMIN adapted GRADE approach for other psychometric properties

		QUALITY OF EVIDENCE: starting point is always HIGH ↓ HIGH MODERATE LOW VERY LOW
Risk of bias (Consider the ratings of the individual studies in STEP 1)	-1	The 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 summarized studies **<100**
	-2	Total sample size of the pooled or summarized studies **<50**
Indirectness*	-1	Psychometric properties were assessed in a cancer population but not representative of the target population (e.g. head & neck cancer patients versus cancer patients, palliative questionnaire assessed in non-palliative cancer patients)
* To assess the indirectness one should look at the characteristics of the pooled population across studies.		

Appendix 1. Detailed overview of the search strategy for PubMed and Scopus

	PubMed	Scopus
Population: cancer patients & survivors	("patient*"[MeSH Terms] OR "Survivors"[MeSH Terms] OR "Palliative Care"[MeSH Terms]) AND ("Neoplasms"[MeSH Terms] OR "Carcinoma"[MeSH Terms] OR "post-cancer"[Title/Abstract] OR "postcancer"[Title/Abstract])	((TITLE-ABS-KEY ("tumor*")) OR (TITLE-ABS-KEY ("neoplasm*")) OR (TITLE-ABS-KEY ("neoplasia*")) OR (TITLE-ABS-KEY ("cancer*")) OR (TITLE-ABS-KEY ("malignanc*")) OR (TITLE-ABS-KEY ("carcinoma*")) OR (TITLE-ABS-KEY ("postcancer")) OR (TITLE-ABS-KEY ("post-cancer"))) AND ((TITLE-ABS-KEY ("palliative care")) OR (TITLE-ABS-KEY ("palliative treatment*")) OR (TITLE-ABS-KEY ("palliative therap*")) OR (TITLE-ABS-KEY ("palliative surger*")) OR (TITLE-ABS-KEY ("palliative supportive care*")) OR (TITLE-ABS-KEY ("survivor*")) OR (TITLE-ABS-KEY ("patient*")))
Exposure: psychometric properties	AND ("instrument*"[Title/Abstract] OR "questionnaire*"[Title/Abstract] OR "measur*"[Title/Abstract] OR "rating*"[Title/Abstract] OR "computer*"[Title/Abstract] OR "digital*"[Title/Abstract] OR "computer-adaptive test*"[Title/Abstract] OR "computer adaptive test*"[Title/Abstract] OR "computer adaptive"[Title/Abstract] OR "computer-adaptive"[Title/Abstract] OR "computerized adaptive test*"[Title/Abstract] OR "computerised adaptive test*"[Title/Abstract] OR "CAT"[Title/Abstract]) AND ("chronbach*"[Title/Abstract] OR "cronbach*"[Title/Abstract] OR "psychometric properties"[Title/Abstract] OR "psychometr*"[Title/Abstract] OR "factor analysis"[Title/Abstract] OR "develop*"[Title/Abstract] OR "reliab*"[Title/Abstract] OR "valid*"[Title/Abstract] OR "translat*"[Title/Abstract] OR "cross-cultural"[Title/Abstract] OR "minimal clinically important difference*"[Title/Abstract] OR "minimal important change*"[Title/Abstract] OR "minimal important difference*"[Title/Abstract] OR "clinically meaningful change*"[Title/Abstract] OR "clinically meaningful difference*"[Title/Abstract] OR "responsiveness"[Title/Abstract])	AND ((TITLE-ABS-KEY ("questionnaire")) OR (TITLE-ABS-KEY ("questionnaires")) OR (TITLE-ABS-KEY ("instrument")) OR (TITLE-ABS-KEY ("instruments")) OR (TITLE-ABS-KEY ("rating")) OR (TITLE-ABS-KEY ("outcome measure")) OR (TITLE-ABS-KEY ("outcome measures")) OR (TITLE-ABS-KEY ("measurement tool")) OR (TITLE-ABS-KEY ("measurement tools")) OR (TITLE-ABS-KEY ("computer-based")) OR (TITLE-ABS-KEY ("digital")) OR (TITLE-ABS-KEY ("computer adaptive test*")) OR (TITLE-ABS-KEY ("computer-adaptive test*")) OR (TITLE-ABS-KEY ("computer-adaptive")) OR (TITLE-ABS-KEY ("computer adaptive")) OR (TITLE-ABS-KEY ("computerized adaptive test*")) OR (TITLE-ABS-KEY ("computerised adaptive test*")) OR (TITLE-ABS-KEY ("cat"))) AND ((TITLE-ABS-KEY ("chronbach*")) OR (TITLE-ABS-KEY ("cronbach*")) OR (TITLE-ABS-KEY ("psychometric properties")) OR (TITLE-ABS-KEY ("psychometric analysis")) OR (TITLE-ABS-KEY ("psychometric evaluation")) OR (TITLE-ABS-KEY ("psychometric characteristics")) OR (TITLE-ABS-KEY ("factor analysis")) OR (TITLE-ABS-KEY ("reliability")) OR (TITLE-ABS-KEY ("reliable")) OR (TITLE-ABS-KEY ("validity")) OR (TITLE-ABS-KEY ("valid")) OR (TITLE-ABS-KEY ("validation")) OR (TITLE-ABS-KEY ("minimal clinically important difference*")) OR (TITLE-ABS-KEY ("clinically meaningful change*")) OR (TITLE-ABS-KEY ("clinically meaningful difference*")) OR (TITLE-ABS-KEY ("responsiveness")) OR (TITLE-ABS-KEY ("minimal important change*")) OR (TITLE-ABS-KEY ("minimal important difference*")) OR (TITLE-ABS-KEY ("translation")) OR (TITLE-ABS-KEY ("translated")) OR (TITLE-ABS-KEY ("cross-cultural")) OR (TITLE-ABS-KEY ("development")))
Outcome: Health-related Quality of Life	AND ("quality of life"[MeSH Terms] OR "perceived health"[Text Word] OR "health status"[Text Word] OR "life satisfaction"[Text Word] OR "well-being"[Text Word] OR "wellbeing"[Text Word] OR "patient reported outcome measures"[MeSH Terms])	AND (TITLE-ABS-KEY ("quality of life")) OR (TITLE-ABS-KEY ("life quality")) OR (TITLE-ABS-KEY ("patient-reported outcome*")) OR (TITLE-ABS-KEY ("hrqol")) OR (TITLE-ABS-KEY ("patient reported outcome*")) OR (TITLE-ABS-KEY ("perceived health")) OR (TITLE-ABS-KEY ("health status")) OR (TITLE-ABS-KEY ("well-being")) OR (TITLE-ABS-KEY ("wellbeing"))
Exclusion string Terwee et al. 2009 + English filter	AND (english[Filter]) NOT (‘‘addresses’’[Publication Type] OR ‘‘biography’’[Publication Type] OR ‘‘case reports’’[Publication Type] OR ‘‘comment’’[Publication Type] OR ‘‘directory’’[Publication Type] OR ‘‘editorial’’[Publication Type] OR ‘‘festschrift’’[Publication Type] OR ‘‘interview’’[Publication Type] OR ‘‘lectures’’[Publication Type] OR ‘‘legal cases’’[Publication Type] OR ‘‘legislation’’[Publication Type] OR ‘‘letter’’[Publication Type] OR ‘‘news’’[Publication Type] OR ‘‘newspaper article’’[Publication Type] OR ‘‘patient education handout’’[Publication Type] OR ‘‘popular works’’[Publication Type] OR ‘‘congresses’’[Publication Type] OR ‘‘consensus development conference’’[Publication Type] OR ‘‘consensus development conference, nih’’[Publication Type] OR ‘‘practice guideline’’[Publication Type]) NOT	AND (LIMIT TO (LANGUAGE , "english")) AND (EXCLUDE (DOCTYPE , "le") OR EXCLUDE (DOCTYPE , "ed")) AND (EXCLUDE (DOCTYPE , "cp"))

	(‘‘animals’’[MeSH Terms] NOT ‘‘humans’’[MeSH Terms])	

Appendix 2. List of European and associated countries

European and associated countries		
Albania	Germany	North-Macedonia
Andorra	Greece	Norway
Armenia	Hungary	Portugal
Austria	Iceland	Romania
Azerbaijan	Ireland	Russia
Belarus	Italy	San Marino
Belgium	Kazakhstan	Serbia
Bulgaria	Latvia	Slovenia
Croatia	Liechtenstein	Slovakia
Cyprus	Lithuania	Spain
Czechia	Luxembourg	Sweden
Denmark	Malta	Switzerland
Estonia	Moldavia	Turkey
Finland	Monaco	Ukraine
France	Montenegro	United Kingdom
Georgia	Netherlands	Vatican City

Appendix 3. Additional search strategy for European validation papers

1. STEP 1:

- Define entry terms for the SPECIFIC QUESTIONNAIRE:
  - Full name (make sure to enter all the different spelling options)
  - Acronym (make sure to enter all the different spelling options)

Example:

EORTC-QLQ-C30	"European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30"
	EORTC-QLQ-C30
	EORTC QLQ-C30
	EORTC QLQ C30
	QLQ C30

- Combine all the entry terms with OR-function:
  - ("European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30") OR (eortc-qlq-c30)) OR (eortc qlq-c30)) OR (eortc qlq c30)) OR (qlq c30)

2. STEP 2:

- Enter search string for POPULATION:
  - ("patient*"[MeSH Terms] OR "Survivors"[MeSH Terms] OR "Palliative Care"[MeSH Terms]) AND ("Neoplasms"[MeSH Terms] OR "Carcinoma"[MeSH Terms] OR "post-cancer"[Title/Abstract] OR "postcancer"[Title/Abstract])
- Enter search string for PSYCHOMETRIC PROPERTIES:
  - ("instrument*"[Title/Abstract] OR "questionnaire*"[Title/Abstract] OR "measur*"[Title/Abstract] OR "rating*"[Title/Abstract] OR "computer*"[Title/Abstract] OR "digital*"[Title/Abstract] OR "computer-adaptive test*"[Title/Abstract] OR "computer adaptive test*"[Title/Abstract] OR "computer adaptive"[Title/Abstract] OR "computer-adaptive"[Title/Abstract] OR "computerized adaptive test*"[Title/Abstract] OR "computerised adaptive test*"[Title/Abstract] OR "CAT"[Title/Abstract]) AND ("chronbach*"[Title/Abstract] OR "cronbach*"[Title/Abstract] OR "psychometric properties"[Title/Abstract] OR "psychometr*"[Title/Abstract] OR "factor analysis"[Title/Abstract] OR "develop*"[Title/Abstract] OR "reliab*"[Title/Abstract] OR "valid*"[Title/Abstract] OR "translat*"[Title/Abstract] OR "cross-cultural"[Title/Abstract] OR "minimal clinically important difference*"[Title/Abstract] OR "minimal important change*"[Title/Abstract] OR "minimal important difference*"[Title/Abstract] OR "clinically meaningful change*"[Title/Abstract] OR "clinically meaningful difference*"[Title/Abstract] OR "responsiveness"[Title/Abstract])

3. STEP 3:

- Combine search strings of POPULATION, PSYCHOMETRIC PROPERTIES and SPECIFIC QUESTIONNAIRE with the AND-function:
  - ((((("European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30") OR (eortc-qlq-c30)) OR (eortc qlq-c30)) OR (eortc qlq c30)) OR (qlq c30)) AND (("instrument*"[Title/Abstract] OR "questionnaire*"[Title/Abstract] OR "measur*"[Title/Abstract] OR "rating*"[Title/Abstract] OR "computer*"[Title/Abstract] OR "digital*"[Title/Abstract] OR "computer-adaptive test*"[Title/Abstract] OR "computer adaptive test*"[Title/Abstract] OR "computer adaptive"[Title/Abstract] OR "computer-

adaptive"[Title/Abstract] OR "computerized adaptive test*"[Title/Abstract] OR "computerised adaptive test*"[Title/Abstract] OR "CAT"[Title/Abstract]) AND ("chronbach*"[Title/Abstract] OR "cronbach*"[Title/Abstract] OR "psychometric properties"[Title/Abstract] OR "psychometr*"[Title/Abstract] OR "factor analysis"[Title/Abstract] OR "develop*"[Title/Abstract] OR "reliab*"[Title/Abstract] OR "valid*"[Title/Abstract] OR "translat*"[Title/Abstract] OR "cross-cultural"[Title/Abstract] OR "minimal clinically important difference*"[Title/Abstract] OR "minimal important change*"[Title/Abstract] OR "minimal important difference*"[Title/Abstract] OR "clinically meaningful change*"[Title/Abstract] OR "clinically meaningful difference*"[Title/Abstract] OR "responsiveness"[Title/Abstract]))) AND (("patient*"[MeSH Terms] OR "Survivors"[MeSH Terms] OR "Palliative Care"[MeSH Terms]) AND ("Neoplasms"[MeSH Terms] OR "Carcinoma"[MeSH Terms] OR "post-cancer"[Title/Abstract] OR "postcancer"[Title/Abstract]))

4. ***STEP 4:***

- Find search string (which is used to gather the articles for our systematic review but remove English filter)
    - (((((("instrument*"[Title/Abstract] OR "questionnaire*"[Title/Abstract] OR "measur*"[Title/Abstract] OR "rating*"[Title/Abstract] OR "computer*"[Title/Abstract] OR "digital*"[Title/Abstract] OR "computer-adaptive test*"[Title/Abstract] OR "computer adaptive test*"[Title/Abstract] OR "computer adaptive"[Title/Abstract] OR "computer-adaptive"[Title/Abstract] OR "computerized adaptive test*"[Title/Abstract] OR "computerised adaptive test*"[Title/Abstract] OR "CAT"[Title/Abstract]) AND ("chronbach*"[Title/Abstract] OR "cronbach*"[Title/Abstract] OR "psychometric properties"[Title/Abstract] OR "psychometr*"[Title/Abstract] OR "factor analysis"[Title/Abstract] OR "develop*"[Title/Abstract] OR "reliab*"[Title/Abstract] OR "valid*"[Title/Abstract] OR "translat*"[Title/Abstract] OR "cross-cultural"[Title/Abstract] OR "minimal clinically important difference*"[Title/Abstract] OR "minimal important change*"[Title/Abstract] OR "minimal important difference*"[Title/Abstract] OR "clinically meaningful change*"[Title/Abstract] OR "clinically meaningful difference*"[Title/Abstract] OR "responsiveness"[Title/Abstract])) AND ("quality of life"[MeSH Terms] OR "perceived health"[Text Word] OR "health status"[Text Word] OR "life satisfaction"[Text Word] OR "well-being"[Text Word] OR "wellbeing"[Text Word] OR "patient reported outcome measures"[MeSH Terms])) AND (("patient*"[MeSH Terms] OR "Survivors"[MeSH Terms] OR "Palliative Care"[MeSH Terms]) AND ("Neoplasms"[MeSH Terms] OR "Carcinoma"[MeSH Terms] OR "post-cancer"[Title/Abstract] OR "postcancer"[Title/Abstract])) AND (english[Filter])) NOT ((‘‘animals’’[MeSH Terms] NOT ‘‘humans’’[MeSH Terms]))) NOT ((((("instrument*"[Title/Abstract] OR "questionnaire*"[Title/Abstract] OR "measur*"[Title/Abstract] OR "rating*"[Title/Abstract] OR "computer*"[Title/Abstract] OR "digital*"[Title/Abstract] OR "computer-adaptive test*"[Title/Abstract] OR "computer adaptive test*"[Title/Abstract] OR "computer adaptive"[Title/Abstract] OR "computer-adaptive"[Title/Abstract] OR "computerized adaptive test*"[Title/Abstract] OR "computerised adaptive test*"[Title/Abstract] OR "CAT"[Title/Abstract]) AND ("chronbach*"[Title/Abstract] OR "cronbach*"[Title/Abstract] OR "psychometric properties"[Title/Abstract] OR "psychometr*"[Title/Abstract] OR "factor analysis"[Title/Abstract] OR "develop*"[Title/Abstract] OR "reliab*"[Title/Abstract] OR "valid*"[Title/Abstract] OR "translat*"[Title/Abstract] OR "cross-cultural"[Title/Abstract] OR "minimal clinically important difference*"[Title/Abstract] OR "minimal important change*"[Title/Abstract] OR "minimal important difference*"[Title/Abstract] OR "clinically meaningful change*"[Title/Abstract] OR "clinically meaningful difference*"[Title/Abstract] OR "responsiveness"[Title/Abstract])) AND ("quality of life"[MeSH Terms] OR "perceived health"[Text Word] OR "health status"[Text Word] OR "life satisfaction"[Text Word] OR "well-being"[Text Word] OR "wellbeing"[Text Word] OR "patient reported outcome measures"[MeSH Terms])) AND (("patient*"[MeSH Terms] OR

"Survivors"[MeSH Terms] OR "Palliative Care"[MeSH Terms]) AND ("Neoplasms"[MeSH Terms] OR "Carcinoma"[MeSH Terms] OR "post-cancer"[Title/Abstract] OR "postcancer"[Title/Abstract])) AND ((address[Filter] OR biography[Filter] OR casereports[Filter] OR comment[Filter] OR congress[Filter] OR consensusdevelopmentconference[Filter] OR consensusdevelopmentconferencenih[Filter] OR directory[Filter] OR editorial[Filter] OR festschrift[Filter] OR interview[Filter] OR lecture[Filter] OR legalcase[Filter] OR legislation[Filter] OR letter[Filter] OR news[Filter] OR newspaperarticle[Filter] OR patienteducationhandout[Filter] OR practiceguideline[Filter]))

5. ***STEP 5:***

- Combine search string of STEP 3 (POPULATION AND PSYCHOMETRIC PROPERTIES AND SPECIFIC QUESTIONNAIRE) and STEP 4 (ENTIRE search string) with NOT-function:
    - (((((("European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30") OR (eortc-qlq-c30)) OR (eortc qlq-c30)) OR (eortc qlq c30)) OR (qlq c30)) AND (("instrument*"[Title/Abstract] OR "questionnaire*"[Title/Abstract] OR "measur*"[Title/Abstract] OR "rating*"[Title/Abstract] OR "computer*"[Title/Abstract] OR "digital*"[Title/Abstract] OR "computer-adaptive test*"[Title/Abstract] OR "computer adaptive test*"[Title/Abstract] OR "computer adaptive"[Title/Abstract] OR "computer-adaptive"[Title/Abstract] OR "computerized adaptive test*"[Title/Abstract] OR "computerised adaptive test*"[Title/Abstract] OR "CAT"[Title/Abstract]) AND ("chronbach*"[Title/Abstract] OR "cronbach*"[Title/Abstract] OR "psychometric properties"[Title/Abstract] OR "psychometr*"[Title/Abstract] OR "factor analysis"[Title/Abstract] OR "develop*"[Title/Abstract] OR "reliab*"[Title/Abstract] OR "valid*"[Title/Abstract] OR "translat*"[Title/Abstract] OR "cross-cultural"[Title/Abstract] OR "minimal clinically important difference*"[Title/Abstract] OR "minimal important change*"[Title/Abstract] OR "minimal important difference*"[Title/Abstract] OR "clinically meaningful change*"[Title/Abstract] OR "clinically meaningful difference*"[Title/Abstract] OR "responsiveness"[Title/Abstract]))) AND (("patient*"[MeSH Terms] OR "Survivors"[MeSH Terms] OR "Palliative Care"[MeSH Terms]) AND ("Neoplasms"[MeSH Terms] OR "Carcinoma"[MeSH Terms] OR "post-cancer"[Title/Abstract] OR "postcancer"[Title/Abstract]))) NOT ((((((("instrument*"[Title/Abstract] OR "questionnaire*"[Title/Abstract] OR "measur*"[Title/Abstract] OR "rating*"[Title/Abstract] OR "computer*"[Title/Abstract] OR "digital*"[Title/Abstract] OR "computer-adaptive test*"[Title/Abstract] OR "computer adaptive test*"[Title/Abstract] OR "computer adaptive"[Title/Abstract] OR "computer-adaptive"[Title/Abstract] OR "computerized adaptive test*"[Title/Abstract] OR "computerised adaptive test*"[Title/Abstract] OR "CAT"[Title/Abstract]) AND ("chronbach*"[Title/Abstract] OR "cronbach*"[Title/Abstract] OR "psychometric properties"[Title/Abstract] OR "psychometr*"[Title/Abstract] OR "factor analysis"[Title/Abstract] OR "develop*"[Title/Abstract] OR "reliab*"[Title/Abstract] OR "valid*"[Title/Abstract] OR "translat*"[Title/Abstract] OR "cross-cultural"[Title/Abstract] OR "minimal clinically important difference*"[Title/Abstract] OR "minimal important change*"[Title/Abstract] OR "minimal important difference*"[Title/Abstract] OR "clinically meaningful change*"[Title/Abstract] OR "clinically meaningful difference*"[Title/Abstract] OR "responsiveness"[Title/Abstract])) AND ("quality of life"[MeSH Terms] OR "perceived health"[Text Word] OR "health status"[Text Word] OR "life satisfaction"[Text Word] OR "well-being"[Text Word] OR "wellbeing"[Text Word] OR "patient reported outcome measures"[MeSH Terms])) AND (("patient*"[MeSH Terms] OR "Survivors"[MeSH Terms] OR "Palliative Care"[MeSH Terms]) AND ("Neoplasms"[MeSH Terms] OR "Carcinoma"[MeSH Terms] OR "post-cancer"[Title/Abstract] OR "postcancer"[Title/Abstract])) AND (english[Filter])) NOT ((‘‘animals’’[MeSH Terms] NOT ‘‘humans’’[MeSH Terms]))) NOT ((((("instrument*"[Title/Abstract] OR "questionnaire*"[Title/Abstract] OR "measur*"[Title/Abstract] OR "rating*"[Title/Abstract] OR "computer*"[Title/Abstract] OR "digital*"[Title/Abstract] OR "computer-adaptive

test*"[Title/Abstract] OR "computer adaptive test*"[Title/Abstract] OR "computer adaptive"[Title/Abstract] OR "computer-adaptive"[Title/Abstract] OR "computerized adaptive test*"[Title/Abstract] OR "computerised adaptive test*"[Title/Abstract] OR "CAT"[Title/Abstract]) AND ("chronbach*"[Title/Abstract] OR "cronbach*"[Title/Abstract] OR "psychometric properties"[Title/Abstract] OR "psychometr*"[Title/Abstract] OR "factor analysis"[Title/Abstract] OR "develop*"[Title/Abstract] OR "reliab*"[Title/Abstract] OR "valid*"[Title/Abstract] OR "translat*"[Title/Abstract] OR "cross-cultural"[Title/Abstract] OR "minimal clinically important difference*"[Title/Abstract] OR "minimal important change*"[Title/Abstract] OR "minimal important difference*"[Title/Abstract] OR "clinically meaningful change*"[Title/Abstract] OR "clinically meaningful difference*"[Title/Abstract] OR "responsiveness"[Title/Abstract])) AND ("quality of life"[MeSH Terms] OR "perceived health"[Text Word] OR "health status"[Text Word] OR "life satisfaction"[Text Word] OR "well-being"[Text Word] OR "wellbeing"[Text Word] OR "patient reported outcome measures"[MeSH Terms])) AND (("patient*"[MeSH Terms] OR "Survivors"[MeSH Terms] OR "Palliative Care"[MeSH Terms]) AND ("Neoplasms"[MeSH Terms] OR "Carcinoma"[MeSH Terms] OR "post-cancer"[Title/Abstract] OR "postcancer"[Title/Abstract])) AND ((address[Filter] OR biography[Filter] OR casereports[Filter] OR comment[Filter] OR congress[Filter] OR consensusdevelopmentconference[Filter] OR consensusdevelopmentconferencenih[Filter] OR directory[Filter] OR editorial[Filter] OR festschrift[Filter] OR interview[Filter] OR lecture[Filter] OR legalcase[Filter] OR legislation[Filter] OR letter[Filter] OR news[Filter] OR newspaperarticle[Filter] OR patienteducationhandout[Filter] OR practiceguideline[Filter])))

6. ***STEP 6: Apply "English" filter***

7. ***STEP 7: Assess and screen articles for the predefined in- and exclusion criteria***

Appendix 4. Overview of the data extraction for the PROMs measurement properties

Measurement property	Data extracted
Development/ Content validity	- Level of analysis: scale/subscale - Methodological approach for concept elicitation, PROM design, relevance, comprehensiveness and comprehensibility
Structural validity/ Unidimensionality	- Level of analysis: scale/subscale - Statistical approach and related sample size: EFA, CFA or IRT - Final model and fit indexes: CFI, TLI, RMSEA (90%CI) SRMR or WRMR
Internal consistency	- Level of analysis: scale/subscale - Statistical approach and related sample size - Internal consistency reliability coefficients: Cronbach alpha, McDonald Omega, KR-20, SE(θ)
Cross-cultural validity/ Measurement invariance	- Level of analysis: scale/subscale - Statistical approach and related sample size - Group variable under investigation (e.g. country, age, gender,...) with its observed differences
Reliability	- Level of analysis: scale/subscale - Statistical approach and related sample size - Type of reliability: test-retest, inter-rater, intra-rater, parallel forms - Correlation coefficients: ICC, Spearman, Pearson, Kappa or weighted Kappa
Measurement error	- Level of analysis: scale/subscale - Statistical approach and related sample size - Standard Error of Measurement, Limits of Agreement, Smallest Detectable Change, Minimal Important Change
Construct validity with other PROM	- Level of analysis: scale/subscale - Statistical approach and related sample size - Comparator + formulated hypotheses - Correlation coefficients or effect sizes
Convergent/ divergent validity within PROM	- Level of analysis: scale/subscale - Statistical approach and related sample size - Formulated hypotheses - Correlation coefficients
Known-group comparison	- Level of analysis: scale/subscale - Statistical approach and related sample size - Formulated hypotheses - Group variable + defined subgroups with observed differences

Abbreviations: CFA = Confirmatory Factor Analysis; CFI = Comparative Fit Index; IRT = Item Response Theory; RMSEA = Root Mean Square Error of Approximation; SDC = Smallest Detectable Change; SRMR: Standardized Root Mean Residuals; TLI: Tucker-Lewis Index; WRMR: Weighted Root Mean Residuals

Appendix 5: Overview of adjustments made to the Risk of Bias rating of COSMIN Guidelines

Psychometric property	Criteria	Adjustment made
PROM development (Box 1)	23	Inadequate rating was removed from the response options.
	25	Adequate and doubtful rating were removed from the response options.
	26	Doubtful rating was removed and inadequate was defined as "NO or not clear (SKIP items 27-35)".
	35	Adequate and doubtful rating were removed from the response options.
Content validity (Box 2)	6	Inadequate rating was removed from the response options.
	13	Inadequate rating was removed from the response options.
	20	Inadequate rating was removed from the response options.
	25	Inadequate rating was removed from the response options.
	30	Inadequate rating was removed from the response options.
Structural validity (Box 3)	2	Adequate rating was removed from the response options.
Internal consistency (Box 4)	5	Criteria 5 was removed from the Risk of bias assessment.
Cross-cultural validity & Measurement invariance (Box 5)	4	Criteria 4 was removed from the Risk of bias assessment.
Reliability (Box 6)	1-3	Not applicable rating was added to the response options.
Measurement error (Box 7)	6	Adequate rating was removed from the response options.
Construct validity (with other PROM) (hypothesis testing) (Box 9.a)	4	Inadequate rating was removed from the response options.
	1-4	Not applicable rating was added to the response options.
Construct validity (Known-group comparison) (Box 9.b)	7	Inadequate rating was removed from the response options.
	5-7	Not applicable rating was added to the response options.
Construct validity (convergent & divergent validity)	1	Criteria 3 of Box 9.a was introduced.

Appendix 6: The 10 criteria for good content validity

		PROM development study		Content validity study		Rating of reviewers
1	**+**	Construct of interest is clearly described **(criterion 1 of box 1A = very good) AND** origin of construct is clear **(criterion 2 of box 1A = very good) AND** the is evidence from concept elicitation, literature or professionals that ≥85% of the items refer to construct of interest	**+**	Professionals rated the relevance of items for the construct of interest as sufficient **(criteria 22-26 of box 2D = very good, adequate or doubtful)** and found ≥85% of the items relevant for the construct		Reviewers consider ≥85% of the items relevant for the construct of interest
	-	Quality is inadequate (≥ 1 of the 3 (+)-criteria is not fulfilled)	**-**	Professionals were not involved in the content validity study OR rated <85% of the items of the PROM relevant for the construct		Reviewers consider <85% of the items relevant for the construct of interest
	?	No(t enough) information available to score a (+) or (-)	**?**	No(t enough) information available to score a (+) or (-)		
2	**+**	Target population of interest is clearly described **(criterion 3 of box 1A = very good) AND** representative patients were involved in the elicitation of relevant items **(criterion 5 of box 1A = very good or adequate) AND** concept elicitation was not inadequate **(criteria 6-13 of box 1A = very good, adequate or doubtful)**	**+**	Patients rated the relevance of items for the construct of interest as sufficient **(criteria 1-7 of box 2A = very good, adequate or doubtful)** and found ≥85% of the items relevant for them		Reviewers consider ≥85% of the items relevant for the population of interest
	-	Quality is inadequate (≥ 1 of the 3 (+)-criteria is not fulfilled)	**-**	Patients were not involved in the content validity study OR rated <85% of the items of the PROM relevant for them		Reviewers consider <85% of the items relevant for the population of interest
	?	No(t enough) information available to score a (+) or (-) OR doubtful whether study was performed in a sample representing the target population	**?**	No(t enough) information available to score a (+) or (-)		
3	**+**	The context of use of interest is clearly described **(criterion 4 of box 1A = very good)**	**+**	Professionals rated the relevance of items for the context of use as sufficient **(criteria 22-26 of box 2D = very good, adequate or doubtful)** and found ≥85% of the items relevant for the context of use		Reviewers consider ≥85% of the items relevant for the context of use of interest
	-	The context of use of interest is not clearly described **(criterion 4 of box 1A = doubtful)**	**-**	Professionals were not involved in the content validity study OR rated <85% of the items of the PROM relevant for the context of use		Reviewers consider <85% of the items relevant for the context of use of interest
	?	No(t enough) information available to score a (+) or (-)	**?**	No(t enough) information available to score a (+) or (-)		
4	**+**	A justification is provided for the response options	**+**	Patients or professionals rated the appropriateness of the response options as sufficient **(criteria 1-7 of box 2A or criteria 22-26 of box 2D = very good, adequate or doubtful)** and found ≥85% of the response options relevant		Reviewers consider ≥85% of the response options appropriate for the construct, population, and context of use of interest
	-	No justification was provided for the response options	**-**	Patients or professionals were not involved in the content validity study OR rated <85% of the response options of the PROM relevant		Reviewers consider <85% of the response options appropriate for the construct, population, and context of use of interest
	?	No(t enough) information available to score a (+) or (-)	**?**	No(t enough) information available to score a (+) or (-)		
5	**+**	A justification is provided for the recall period	**+**	Patients or professionals rated the appropriateness of the recall period as sufficient **(criteria 1-7 of box 2A or criteria 22-26 of box 2D = very good, adequate or doubtful)** and found the recall period relevant	**+**	Reviewers consider the recall period appropriate for the construct, population and context of use of interest for **≥85% of the items**.
	-	No justification is provided for the recall period	**-**	Patients or professionals were not involved in the content validity study OR rated the recall period for <85% of the items of the PROM relevant	**-**	Reviewers consider the recall period **only partially** (<85% of the items) **OR not** appropriate for the construct, population and context of use of interest.
	?	No(t enough) information available to score a (+) or (-)	**?**	No(t enough) information available to score a (+) or (-)		

6	**+**	Patients were asked about the comprehensiveness of the PROM in concept elicitation phase or cognitive interview **(criteria 6-13 of box 1A or criteria 26-35 of box 1B = very good, adequate or doubtful) AND** no key concepts were missing	**+**	Patients or professionals were asked about the comprehensiveness of the PROM **(criteria 8-14 of box 2B or criteria 27-31 of box 2E = very good, adequate or doubtful) AND** no key concepts were missing	Reviewers consider the PROM **comprehensive** for the construct, population and context of use of interest for **≥85% of the items**.
	-	Quality is inadequate (≥ 1 of the 2 (+)-criteria is not fulfilled)	**-**	Patients or professionals were not involved in the content validity study OR quality is inadequate (≥ 1 of the 2 (+)-criteria is not fulfilled)	Reviewers consider the PROM **only partially** (<85% of the items) **OR not comprehensive** for the construct, population and context of use of interest comprehensive (<85% of the items)
	?	No(t enough) information available to score a (+) or (-)	**?**	No(t enough) information available to score a (+) or (-)	
7	**+**	Patients were asked about the comprehensibility of the instructions (including recall period) in cognitive interview **(criteria 16-25 of box 1B = very good, adequate or doubtful) AND** problems were adequately addressed	**+**	Patients were asked about the comprehensibility of the instructions (including recall period) **(criteria 15-21 of box 2C = very good, adequate or doubtful) AND** no important problems were found	
	-	Quality is inadequate (≥ 1 of the 2 (+)-criteria is not fulfilled)	**-**	Patients were not involved in the content validity study OR quality is inadequate (≥ 1 of the 2 (+)-criteria is not fulfilled)	
	?	No(t enough) information available to score a (+) or (-)	**?**	No(t enough) information available to score a (+) or (-)	
8	**+**	Patients were asked about the comprehensibility of the items and response options (including wording of items and response options) in cognitive interview **(criteria 16-25 of box 1B = very good, adequate or doubtful) AND** problems were adequately addressed	**+**	Patients were asked about the comprehensibility of the items and response options (including wording of items and response options) **(criteria 15-21 of box 2C = very good, adequate or doubtful) AND** no important problems were found for ≥85% of the items and response options	
	-	Quality is inadequate (≥ 1 of the 2 (+)-criteria is not fulfilled)	**-**	Patients were not involved in the content validity study OR quality is inadequate (≥ 1 of the 2 (+)-criteria is not fulfilled)	
	?	No(t enough) information available to score a (+) or (-)	**?**	No(t enough) information available to score a (+) or (-)	
9	**+**		**+**		Reviewers consider ≥85% of the items and response options appropriately worded
	-		**-**		Reviewers consider <85% of the items and response options appropriately worded
	?		**?**		
10	**+**		**+**		Reviewers consider ≥85% of the response options matching the questions
	-		**-**		Reviewers consider <85% of the response options matching the questions
	?		**?**		

Appendix 7: Calculation of the overall relevance, comprehensiveness and comprehensibility rating per study.

		PROM development		Content validity	Reviewer rating
Relevance rating	**+**	Criteria 1 and 2 are rated sufficient (+) AND ≥2 of remaining 3 items are rated sufficient (+)	**+**	Criteria 1 and 2 are rated sufficient (+) AND ≥2 of remaining 3 items are rated sufficient (+)	Criteria 1 and 2 are rated sufficient (+) AND ≥2 of remaining 3 items are rated sufficient (+)
	-	Criteria 1 and 2 are rated insufficient (-) AND ≥2 of remaining 3 items are rated insufficient (-)	**-**	Criteria 1 and 2 are rated insufficient (-) AND ≥2 of remaining 3 items are rated insufficient (-)	Criteria 1 and 2 are rated insufficient (-) AND ≥2 of remaining 3 items are rated insufficient (-)
	?	≥2 criteria are rated indeterminate (?)	**?**	≥2 criteria are rated indeterminate (?)	
	±	All other situations	**±**	All other situations	All other situations
Comprehensiveness rating		Rating of criterion 6		Rating of criterion 6	Rating of criterion 6
Comprehensibility rating	**+**	Criterion 8 = sufficient (+) AND criterion 7 = sufficient (+) or indeterminate (?)	**+**	Criterion 8 = sufficient (+) AND criterion 7 = sufficient (+) or indeterminate (?)	Criteria 9 and 10 are rated sufficient (+)
	-	Criterion 8 = insufficient (-)	**-**	Criterion 8 = insufficient (-)	Criteria 9 and 10 are rated insufficient (-)
	?	Criterion 8 = indeterminate (?)	**?**	Criterion 8 = indeterminate (?)	
	±	Criterion 8 = sufficient (+) AND criterion 7 = insufficient (-)	**±**	Criterion 8 = sufficient (+) AND criterion 7 = insufficient (-)	One criterion = sufficient (+) AND one criterion = insufficient (-)

Appendix 8: Calculation of the overall relevance, comprehensiveness and comprehensibility rating per PROM

PROM development	Content validity	Rating reviewer	Overall RELEVANCE COMPREHENSIVENES COMPREHENSIBILITY rating
+	+	+	+
+	+	±	+
+	+	-	+
+	-	+	±
+	-	±	±
+	-	-	-
+	?	+	+
+	?	±	±
+	?	-	±
+	±	+	±
+	±	±	±
+	±	-	±
-	+	+	+
-	+	±	±
-	+	-	±
-	-	+	-
-	-	±	-
-	-	-	-
-	?	+	±
-	?	±	±
-	?	-	-
-	±	+	±
-	±	±	±
-	±	-	±
?	+	+	+
?	+	±	±
?	+	-	±
?	-	+	±
?	-	±	±
?	-	-	-
?	?	+	+
?	?	±	±
?	?	-	-
?	±	+	±
?	±	±	±
?	±	-	±
±	+	+	+
±	+	±	+
±	+	-	±
±	-	+	±
±	-	±	-
±	-	-	-
±	?	+	±
±	?	±	±
±	?	-	±
±	±	+	±
±	±	±	±
±	±	-	-

Appendix 9: Calculation of the overall content validity rating

Overall RELEVANCE rating	Overall COMPREHENSIVENESS rating	Overall COMPREHENSIBILITY rating	Overall CONTENT VALIDITY rating
+	+	+	+
+	+	±	+
+	+	-	±
+	-	+	±
+	-	±	±
+	-	-	±
+	±	+	+
+	±	±	±
+	±	-	±
-	+	+	±
-	+	±	±
-	+	-	±
-	-	+	±
-	-	±	-
-	-	-	-
-	±	+	±
-	±	±	±
-	±	-	-
±	+	+	+
±	+	±	±
±	+	-	±
±	-	+	±
±	-	±	±
±	-	-	-
±	±	+	±
±	±	±	±
±	±	-	±