BMJ Open

# Psychometric properties of self-reported financial toxicity measures in cancer survivors: a systematic review

Zheng Zhu ,[1,2] Weijie Xing,[1,2] Huan Wen,[3] Yanling Sun,[3] Winnie K W So,[4] Lucylynn Lizarondo,[5] Jian Peng,[1] Yan Hu[1,2]

**To cite:** Zhu Z, Xing W, Wen H, *et al*. Psychometric properties of self-reported financial toxicity measures in cancer survivors: a systematic review. *BMJ Open* 2022;**12**:e057215. doi:10.1136/bmjopen-2021-057215

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2021-057215).

Received 13 September 2021
Accepted 06 June 2022



[1]School of Nursing, Fudan University, Shanghai, China
[2]Fudan University Centre for Evidence-based Nursing: A Joanna Briggs Institute Centre of Excellence, Fudan University, Shanghai, China
[3]School of Public Health, Fudan University, Shanghai, China
[4]The Nethersole School of Nursing, The Chinese University of Hong Kong, Hong Kong, China
[5]The Joanna Briggs Institute, University of Adelaide, Adelaide, South Australia, Australia

**Correspondence to**
Dr Weijie Xing;
xingweijie@fudan.edu.cn

## ABSTRACT

**Objective** The aim of this systematic review was to summarise the psychometric properties of patient-reported outcome measures (PROMs) measuring financial toxicity (FT) in cancer survivors.

**Design** This systematic review was conducted according to the guidance of the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) methodology.

**Data sources** Comprehensive searches were performed in PubMed, MEDLINE, Embase, CINAHL, PsycINFO, Web of Science, ProQuest and Cochrane Library from database inception to February 2022.

**Eligibility criteria for selecting studies** We included studies that reported any PROMs for measuring FT in cancer survivors who were ≥18 years old. FT was defined as perceived subjective financial distress resulting from objective financial burden. Studies that were not validation studies and that used a PROM only as an outcome measurement were excluded.

**Data extraction and synthesis** Two reviewers independently extracted data from the included papers. We used the COSMIN criteria to summarise and evaluate the psychometric properties of each study regarding structural validity, internal consistency, reliability, measurement error, hypothesis testing for construct validity, cross-cultural validity/measurement invariance, criterion validity and responsiveness.

**Results** A total of 23 articles (21 PROMs) were eligible for inclusion in this study. The findings highlighted that the Comprehensive Score for Financial Toxicity (COST) had an adequate development process and showed better psychometric properties than other PROMs, especially in internal consistency (Cronbach's α=0.92), reliability (intraclass correlation coefficient=0.80) and hypothesis testing (r=0.42–0.20).

**Conclusions** From a psychometric property perspective, the COST could be recommended as the most suitable worldwide available measure for use in research and clinical practice across different contexts. We suggest that PROMs should be selected only after careful consideration of the local socioeconomic context. Future studies are warranted to develop various FT PROMs based on different social and cultural backgrounds and to clarify the theoretical grounds for assessing FT.

## STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This is the first systematic review that comprehensively summarised the psychometric properties of 21 patient-reported outcome measures (PROMs) evaluating financial toxicity in cancer survivors.
⇒ The results may provide quantitative evidence for researchers and healthcare professionals to choose PROMs measuring cancer survivors' financial toxicity in future scientific research and clinical practice.
⇒ This review only included studies that aimed to evaluate the measurement properties of financial toxicity PROMs.

## INTRODUCTION

The rising cost associated with advancements in cancer treatment and lengthening of cancer survivorship poses a significant challenge to survivors, caregivers and public healthcare systems.[1 2] Total global spending on cancer medications grows at a compound annual growth rate of 6.5%, growing from US$96 billion in 2013 to US$173 billion in 2020, which is nearly twice the rate of global gross domestic product growth.[3–5] The majority of cancer survivors in middle-income and low-income countries/regions depend on out-of-pocket payments, which may lead to global inequalities in healthcare expenditures and financial insecurity for vulnerable groups.[6 7]

The term 'financial toxicity (FT)' has been described as the economic effect of cancer treatment in the age of precision medicine.[2 8 9] Witte *et al* described FT as 'the patient-reported outcome (PRO) of perceived subjective financial distress resulting from objective financial burden'.[10] This concept covers both the objective financial burden and the subjective financial distress that cancer survivors face as a result of high out-of-pocket medical expenses. Regarding the terminology, 'financial toxicity', 'financial burden' and 'financial distress' are often used interchangeably in research and share a similar definition.[10 11] In this review, the authors agreed to consistently use the term 'financial toxicity'. Financial toxicity is usually measured by PRO measures

(PROMs); choosing a PROM with high validity and reliability is a prerequisite for robust results.

There are a few cancer-specific and generic FT PROMs that have been reported and used in different contexts. As one of the recent cancer-specific FT PROMs, the Comprehensive Score for Financial Toxicity (COST) is the most commonly used measure for assessing FT.[12] In addition to COST, other cancer-specific measures have been widely used, including the Breast Cancer Finances Survey Inventory,[13] Socioeconomic Well-being Scale (SWBS)[14] and InCharge Financial Distress/Financial Well-being Scale (InCharge).[15] Additionally, validated subscales, such as the Social Difficulties Inventory Cancer Care Outcomes (SDI), the Cancer Care Outcomes Research and Surveillance Consortium patient survey, and Italian version of the Edmonton Symptom Assessment System-Total Care (TC), were also used to evaluate FT.[16–18] However, existing PROMs vary significantly in their state of development and degree of validation, and many PROMs have not been psychometrically tested.

A preliminary literature search was conducted in PubMed, PsycINFO (EBSCO), Cochrane Library (Wiley) and Joanna Briggs Institute (Ovid), which revealed that there exist some reviews regarding measures of FT. Witte *et al* summarised the content of 352 items from 34 studies measuring FT in cancer survivors.[10] However, this review did not report the psychometric properties of the included PROMs, and most of the included PROMs were not validated through a scientific process, which made it difficult for readers to choose the best measure from existing PROMs to evaluate the level of FT. Salman *et al* conducted a systematic review and found eight PROMs and two caregiver-reported measures for assessing financial burden in adolescents and young adults.[19] However, this review focused only on PROMs assessing FT in adolescents and young adults with cancer. The psychological properties of FT measures in adult cancer survivors are still unknown.

The reproducibility, reliability and accuracy of PROMs are the fundamental premise for achieving robust results. Therefore, it is necessary to summarise the psychometric properties of existing PROMs for future research. However, this information is still lacking. The aim of this systematic review was to summarise the psychometric properties of PROMs for measuring FT in cancer survivors. The review was conducted according to the guidance of the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) methodology and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement.[20 21] The protocol of this review was published in *BMJ OPEN* in 2020.[22] The registration number of the protocol in PROSPERO was CRD42021254721.

## METHODS

### Search strategy

First, we conducted a limited search via PubMed to capture keywords from which to develop search strategies for each database. Subsequently, all identified search strategies across databases were performed in PubMed/MEDLINE, MEDLINE (Ovid), Embase (Ovid), CINAHL (EBSCO), PsycINFO (EBSCO), Web of Science, ProQuest Dissertations and Theses, and Cochrane Library (Wiley). The search time frame was set from database inception to February 2022. To include more studies published in 2021 and 2022, the end date of the search was updated to February 2022.[22] In PubMed/Medline, we searched papers in English using MeSH terms ([cancer OR neoplasms] AND ["cancer survivors" OR patient OR survivors] AND "cost of illness") combined with (cancer OR [patient* OR survivor*] AND [cost OR bill* OR expense OR productivity loss OR "out-of-pocket" OR "economic burden" OR "financial toxicity" OR "financial hardship" OR "financial burden"]). The COSMIN measurement properties filter and exclusion filter were also used in the search box. The search strategies for each database are presented in online supplemental appendix 1. Finally, the references of all included studies were manually reviewed to supplement the database search.

### Inclusion and exclusion criteria

The inclusion criteria were as follows: (1) studies that reported any PROMs for measuring FT in cancer survivors who were ≥18 years old. If the studies reported results in a population combined with both ≥18 and <18 years old cancer survivors and the majority of survivors were not <18 years old, the studies were also considered; (2) studies that evaluated at least one measurement property; and (3) studies published in English. The exclusion criteria were as follows: (1) studies that were not validation studies and used a PROM only as an outcome measurement; (2) studies that used a PROM as a comparator for another instrument; (3) studies that did not provide empirical data and (4) if a measure was a quality of life PROM and had a domain that assessed FT, we included only the original version of the PROM. If the measure/domain included only one item and reported the measurement property as an independent domain, the measure/domain was also considered.

### Study screening and selection

We imported all identified citations by search strategies into Endnote V.X8 (Clarivate Analytics, Pennsylvania, USA). After duplicates were removed, two reviewers (ZZ and WX) independently screened all titles, abstracts and full texts (ZZ and WX) based on the established inclusion and exclusion criteria. Any disagreements were resolved by a third reviewer (YH).

### Quality appraisal

Two reviewers (HW and YS) assessed the methodological quality of the PROM of the included studies by using the COSMIN Risk of Bias Checklist (online supplemental appendix 2).[19] The checklist consisted of 10 domains (116 items), including PROM development, content validity, structural validity, internal consistency, cross-cultural

validity, reliability, measurement error, criterion validity, hypothesis testing and responsiveness. Each measurement property was rated as 'very good', 'adequate', 'doubtful' or 'inadequate quality'. According to the COSMIN guidelines, the methodological quality of a single study is rated based on the worst score count method. For example, if the lowest rating is 'inadequate' in the PROM development domain, the overall methodological quality of that domain is 'inadequate'. The worst score counts method takes into account that inadequate quality items could affect the overall results of the measurement property of each PROM. Any discrepancies were resolved by a third reviewer (ZZ).

### Data extraction

Two reviewers (ZZ and WX) independently extracted data from the included papers, including authors, year of publication, PROM, country/language, study design, target population, sample size, domains, number of items, total score range and main findings. The main findings regarding psychometric properties, including content validity, structural validity, internal consistency, cross-cultural validity, reliability, measurement error, criterion validity, hypothesis testing and responsiveness, were also extracted. Any discrepancies were resolved through discussion between the two reviewers.

### Data synthesis

We used the COSMIN criteria to summarise and evaluate the psychometric properties of each study regarding structural validity, internal consistency, reliability, measurement error, hypothesis testing for construct validity, cross-cultural validity/measurement invariance, criterion validity and responsiveness. Each measurement property from each study was rated as sufficient (+), insufficient (−) or indeterminate (?). The criteria for the measurement property rating can be found in online supplemental appendix 2. If the ratings of one psychometric property per study were all sufficient (+) or insufficient (−), the results were pooled, and the overall rating was rated as sufficient (+) or insufficient (−). If the ratings were inconsistent, explanations of inconsistency were explored (eg, different languages). For example, in our review, different language, social, economic and cultural contexts may contribute to inconsistencies in psychometric properties. Our review team (ZZ, WJ, HW and YS) discussed the potential explanations of inconsistency. If the review team regarded the explanation as reasonable, we provided ratings ('+', '−' and '?') in subgroups (eg, language subgroup). If the explanation was not reasonable, the overall rating of this measurement property was rated as inconsistent (±).

### Assessing certainty of evidence

We used a modified Grading of Recommendations Assessment, Development and Evaluation system to assess the certainty of evidence.[19] Each piece of evidence was graded for risk of bias, inconsistency, imprecision and indirectness. The instructions for downgrading for risk of bias, inconsistency, imprecision and indirectness are shown in Appendix II. Four reviewers (ZZ, WJ, HW and YS) independently assessed the grade. Any discrepancies were resolved by discussion.

### Patient and public involvement

No patients or the public were directly involved in the development of the research question, selection of the outcome measures, design and implementation of the study, or interpretation of the results.

## RESULTS

### Literature search

Figure 1 shows the process of literature screening and selection. A total of 9399 articles were identified via databases. Six articles were found by additional supplementary searches. After duplications were removed, a total of 11 731 articles were retained, 11 669 articles were deleted after reading the title and abstract, and 39 were deleted after full-text reading. Finally, a total of 23 articles (21 PROMs) were eligible for inclusion in this study.[12 14 16 23–42]

### Study description

Table 1 shows the characteristics of the included studies. All included studies were published from 2005 to 2022. Eight studies were conducted in the USA,[12 14 23 27 30 37 39 41] four in the UK,[16 29 35 38] two in Canada[31 36] and two in China (mainland and Hong Kong),[25 39] India[26 34] and Italy.[33 42] One study was conducted in 12 countries in Europe and North America.[22 23] Other studies were conducted in Brazil[32] and Iran.[34] A total of 12 362 participants were included, ranging from 7[36] to 5901[41] per study. The majority of studies assessed FT in multiple types of cancer. Only two studies focused on a single type of cancer, namely, lung, colorectal, or head and neck cancer.[31 37]

Among the 21 PROMs, 7 were FT-related domains of quality of life PROMs and 14 were independent PROMs focusing on FT. All PROMs were validated in cancer survivors. Fifteen PROMs were in English,[12 14 16 23 25–31 35 37 38 40–42] and two were in Chinese.[24 39] Other languages included French,[36] Portuguese,[32] Italian,[33 42] Hindi[25 26] and Persian.[34] The number of items evaluating FT ranged from 3[40] to 23.[36] The French version of the Patient Self-Administered Financial Effects Questionnaire (P-SAFE) did not report the total score range of the whole PROM.[36]

### Quality assessment

#### Methodological quality assessment

Table 2 shows the methodological quality of the 23 included studies by using the COSMIN checklist. In the PROM development domain, only one study was rated as adequate,[42] three studies were rated as doubt[12 24 27 29] and the others were rated as inadequate. Two studies reported adequate information in testing the relevance, comprehensiveness and comprehensibility of PROMs.[12 27 29] One study reported adequate

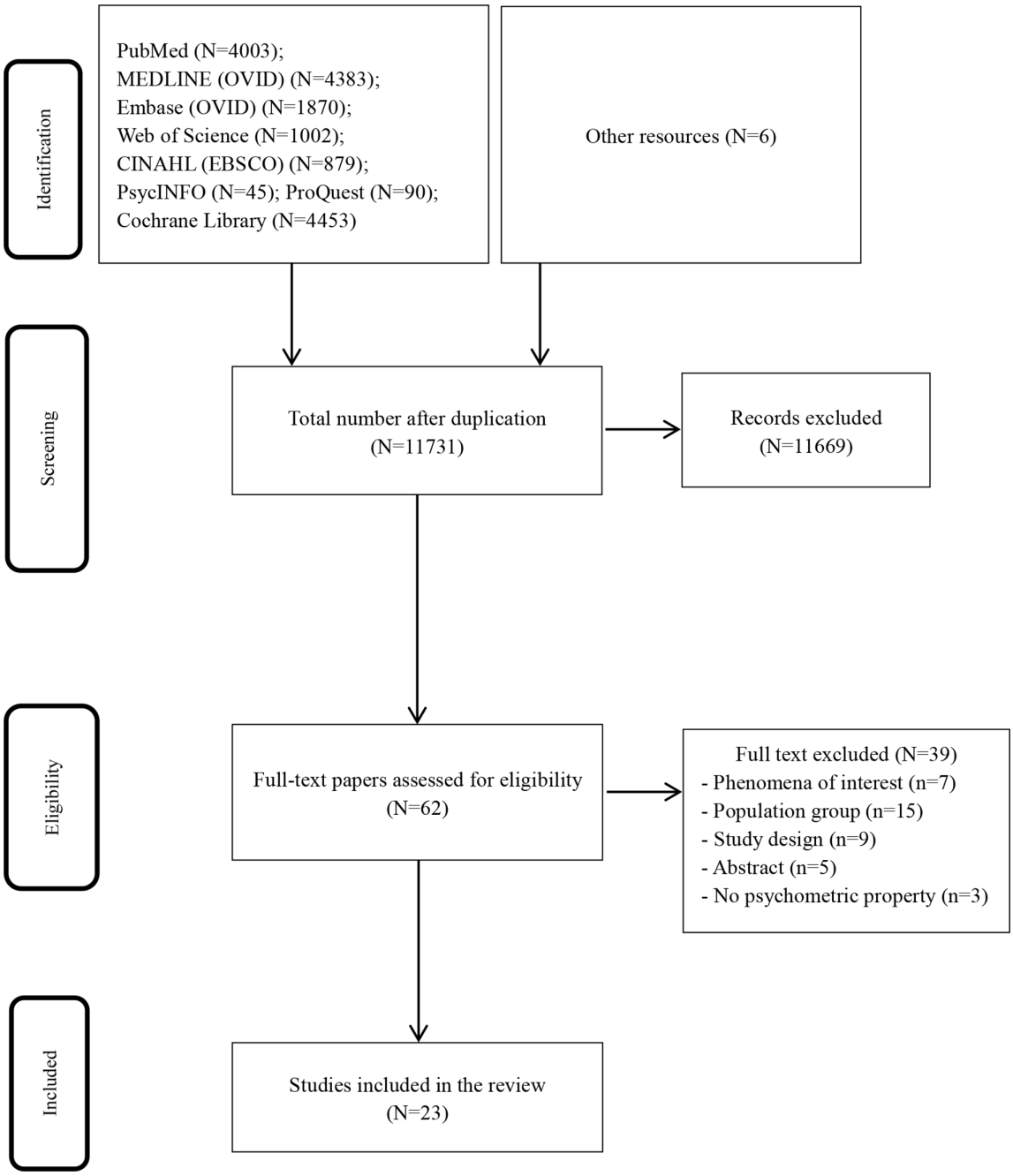


**Figure 1** PRISMA flow chat of selection process. PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

relevance and comprehensiveness.[42] Among all studies, the most reported domain was internal consistency, except one study.[36] Limited information could be retrieved on cross-cultural validity (3 studies),[31 32 36] criterion validity (6 studies),[16 23 33 35 38 40] reliability (10 studies),[12 16 24 27 28 33 35 38–40 42] and responsiveness (2 studies).[31 39] No data were identified as measurement error.

### Measurement property assessment

Table 3 shows the quality of the psychometric properties retrieved from 21 PROMs. Only the Persian version of the COST-v2 and Subjective Financial Distress Questionnaire (SFDQ) were rated as '+' in structural validity.[26 34] There were 17 PROMs rated as '+' in internal consistency.[12 14 16 23 24 26–29 31 32 34 35 37–39 41 42] Eight PROMs were rated as '+'" in reliability.[12 24 26–29 31 35] Ten PROMs were rated as '+' in hypothesis testing.[12 14 23 24 27–31 33 35 39] Limited information was retrieved on cross-cultural validity (two PROMs),[32 36] criterion validity (six PROMs),[16 24 33 35 38 40] and responsiveness (two PROMs).[31 39] No PROMs reported data on measurement error.

### Certainty of evidence

Table 4 shows the certainty of evidence for each measurement property. Among all included PROMs, the COST showed the best psychometric properties compared with other measures. The COST and its

**Table 1** Overview of the included studies

| Author (year), country | PROM | Country | Language(s) of PROM | Study design | Target population | Sample size | Measurement domain | No of items | Total score range |
|---|---|---|---|---|---|---|---|---|---|
| Avis[23] 2005 | QLACS financial problems domain | USA | English | Development and validation study | Age: 71.4±11.5 years<br>Male: 42%<br>Cancer type: Breast, bladder, head and neck, gynecologic, prostate, colorectal cancer | 242 | One domain regarding financial problems | 4 | 4–28 |
| Chan[24] 2021 | COST-v2, Traditional Chinese version | Hong Kong | Traditional Chinese | Validation study | Age: 59.9±11.1 years<br>Male: 35.3%<br>Cancer type: Breast, gynaecological, head and neck, gastric and colorectal, genitourinary, lung, haematological, skin, bone and soft tissue, brain and central nervous system cancer and others | 640 | No subdomain | 12 | 0–44 |
| Dar[25] 2021 | COST-v1, India version | India | Hindi or English | Validation study | Age: 49.5±16.8 years<br>Male: 82.8%<br>Cancer type: Tongue, gingival buccal sulcus, buccal mucosa, supraglottic larynx, hypopharynx, parotid and others | 29 | No subdomain | 11 | 0–44 |
| Dar[25] 2021 | SFDQ | India | Indian or English | Development and validation study | Age: 18%–59 68.3%; ≥60 310.7%<br>Male: 85.9%<br>Cancer type: head and neck cancer | 142 | Five domains: financial resources; financial spending; psychosocial affect; coping care and coping lifestyles; support seeking | 14 | 0–28 |
| de Alcantara Nogueira[32] 2020 | COST-v1, Brazilian version | Brazil | Brazilian Portuguese | Validation study | Mean age: 56 years<br>Male: 40.5%<br>Cancer type: Not specific | 126 | No subdomain | 11 | 0–44 |
| De Souza[12] 2014 De Souza[27] 2017 | COST-v1 | USA | English | Development study and validation study | Age: 58.4±11.5 years<br>Male: 41.6%<br>Cancer type: Not specified (diagnosis of AJCC stage IV cancer) | 233 | No subdomain | 11 | 0–44 |
| Durber[28] 2021 | COST-v1 | Australia | English | Validation study | Age: ≤50 years 23%; 51–64 years 30%; ≥65 years 48%<br>Male: 46%<br>Cancer type: Thoracic, breast, sarcoma, skin, central nervous system, gynaecological, head and neck, colorectal, upper gastrointestinal, urological and miscellaneous cancer | 257 | No subdomain | 11 | 0–44 |

Continued

**Table 1** Continued

| Author (year), country | PROM | Country | Language(s) of PROM | Study design | Target population | Sample size | Measurement domain | No of items | Total score range |
|---|---|---|---|---|---|---|---|---|---|
| Harley[29] 2019 | CCEQ financial advice domain | UK | English | Development and validation study | Pilot study:<br>Age: 65 (41-90) yrs<br>Male: 48.5%<br>Cancer type: Breast, colorectal/ gastrointestinal, gynaecological, prostate and renal cancer<br>Final study:<br>Age: 67 (41–88) yrs<br>Male: 50.0%<br>Cancer type: Breast, colorectal/ gastrointestinal, gynaecological, prostate and renal cancer | 103 for pilot study<br>313 for final study | One domain regarding financial advice | 5 | 5–25 |
| Head[14] 2008 | SWBS | USA | English | Development and validation study | Age: 59.6±12.7 years<br>Male: 35.7%<br>Cancer type: Breast, melanoma, head and neck, prostate, rectum/ anus, colon, endometrium, lung/ tracheal/bronchus and non-Hodgkin's lymphoma | 266 | Two domains: material and social capital. | 17 | 0–68 |
| Hueniken[31] 2020 | FIT | Canada | English | Development and validation study | Age: 61.6 (25.5–88.5) yrs<br>Male: 77.2%<br>Cancer type: Oropharynx, oral cavity, larynx, nasopharynx, hypopharynx cancers and others | 430 | No subdomain | 9 | 0–100 |
| Ripamonti[33] 2020 | COST-v2, Italian version | Italy | Italian | Validation study | Age: 61.5±12.7 years<br>Male: 52.5%<br>Cancer type: Breast, lung, colon, gastric, hepatocellular, endometrial, prostate, sarcoma, bladder, head and neck, Hodgkin lymphoma, non-Hodgkin's lymphoma, leukaemia, myeloma and others | 118 | No subdomain | 11 | 0–44 |
| Riva[42] 2021 | PROFFIT | Italy | Italian or English | Development and validation study | Age: 29–82 years<br>Male: 41.3%<br>Cancer type: breast, lower gastrointestinal tract, genitourinary, thoracic, upper gastrointestinal tract and others | 184 | No subdomain | Seven outcome items and eight determinant items | 0–100 |
| Sharif[34] 2020 | COST-v2, Persian version | Iran | Persian | Validation study | Age: 50.0±14.3 years<br>Male: 51.0%<br>Cancer type: Not specific | 398 | No subdomain | 11 | 0–44 |
| Shilling[35] 2018 | PRRS financial well-being domain | UK | English | Development and validation study | Age: ≤50 years 25%; 51–65 years 41%; ≥66 years 34%<br>Male: 23%<br>Cancer type: breast, gynaecological, lung and melanoma cancers | 135 | One domain regarding financial well-being | 6 | 0–24 |

Continued

**Table 1** Continued

| Author (year), country | PROM | Country | Language(s) of PROM | Study design | Target population | Sample size | Measurement domain | No of items | Total score range |
|---|---|---|---|---|---|---|---|---|---|
| Tremblay[36] 2020 | P-SAFE, French Version | Canada | French | Cross-adaption study | Age: 50–59 years 57%; 60–69 years 29%; ≥70 years 34%<br>Male: 14%<br>Cancer type: colorectal, lung, breast and prostate cancer. | 7 | NR | 23 | NR |
| Veenstra[37] 2014 | PFB | USA | English | Validation study | Age: <50 years 17%; 50–64 years 37%; 65–74 years 23%; >75 years 24%<br>Male: 53%<br>Cancer type: Stage III colorectal cancer | 956 | No subdomain | 7 | 0–7 |
| Wright[38] 2005 | SDI-21 providing for the family domain | UK | English | Development and validation study | Age: 53.8±14.1 years<br>Male: Not specific<br>Cancer type: brain, lung cancers and others | 271 | One domain regarding providing for the family | 5 | 0–20 |
| Wright[16] 2011 | SDI-16 money matters domain | UK | English | Development and validation study | Age: 56 (18-88) yrs for men; 56 (21-88) yrs for women<br>Male: 48%<br>Cancer type: breast, gastrointestinal, haematology, gynaecological, germ cell, head and neck, lung, genitourinary and others | 652 | One domain regarding money matters | 5 | 0–20 |
| Yu[39] 2021 | COST-v1, *Simplified Chinese version* | Mainland China | Chinese | Validation study | Age: 57.0±9.2 years<br>Male: 45.7%<br>Cancer type: lung, stomach, colorectal and breast cancer | 440 | No subdomain | 11 | 0–44 |
| Zebrack[40] 2010 | IOC-CS financial problems domain | USA | English | Validation study | Age: 26.7±5.3 years<br>Male: 48.0%<br>Cancer type: haematological, brain and solid tumours/soft tissue tumours | 519 | One domain regarding financial problems | 3 | 1–15 |
| Zhao[41] 2009 | CPILS employment/ financial domain | USA | English | Validation study | Age: ≤55 years 48.8%; >55 510.2%<br>Male: 41.6%<br>Cancer type: breast, prostate, colorectal, bladder, uterine, kidney, lung and ovarian cancer; melanoma of skin; non-Hodgkin's lymphoma, | 5901 | One domain regarding employment/finances | 6 | 0–12 |

AJCC, The American Joint Committee on Cancer; CCEQ, Chronic Cancer Experiences Questionnaire; COST, Comprehensive Score for Financial Toxicity; CPILS, Cancer Problems in Living Scale; FIT, Financial Index of Toxicity; IOC-CS, Impact of Cancer-Childhood Survivors; NR, not report; PFB, Personal Financial Burden; PROFFIT, Patient-Reported Outcome for Fighting Financial Toxicity; PROM, Patient-Reported Outcome Measures; PRRS, Patient Roles and Responsibilities Scale; P-SAFE, Patient Self-Administered Financial Effects questionnaire; QLACS, Quality of Life in Adult Cancer Survivors; SDI, Social Difficulties Inventory Cancer Care Outcomes; SFDQ, Subjective Financial Distress Questionnaire; SWBS, Socioeconomic Well-being Scale.

**Table 2** Methodological quality assessment of the measures

| Author (year) | PROM | Measurement property: methodological quality by study | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | PROM development | Content validity | Structural validity | Internal consistency | Cross-cultural validity | Reliability | Measurement error | Criterion validity | Hypothesis testing | Responsiveness |
| Avis[23] 2005 | QLACS financial problems domain | Inadequate | R: NR<br>C1: NR<br>C2: NR | Adequate | Very good | NR | NR | NR | Adequate | Doubtful | NR |
| Chan[24] 2021 | COST-v2, Traditional Chinese version | Doubtful | R: Doubtful<br>C1: NR<br>C2: Doubtful | Very good | Very good | NR | Doubtful | NR | NR | Very good | NR |
| Dar[25] 2021 | COST-v1, India version | Inadequate | R: NR<br>C1: NR<br>C2: NR | Inadequate | Very good | NR | NR | NR | NR | NR | NR |
| Dar[25] 2021 | SFDQ | Inadequate | R: Adequate<br>C1: NR<br>C2: NR | Very good | Very good | NR | NR | NR | NR | NR | NR |
| de Alcantara Nogueira[32] 2020 | COST-v1, Brazilian version | Inadequate | R: NR<br>C1: NR<br>C2: NR | Very good | Very good | Inadequate | NR | NR | NR | NR | NR |
| De Souza[12] 2014<br>De Souza[27] 2017 | COST-v1 | Doubtful | R: Adequate<br>C1: Adequate<br>C2: Adequate | NR | Very good | NR | Adequate | NR | NR | Very good | NR |
| Durber[28] 2021 | COST-v1, Australia version | Inadequate | R: NR<br>C1: NR<br>C2: NR | NR | Very good | NR | Adequate | NR | NR | Very good | NR |
| Harley[29] 2019 | CCEQ financial advice domain | Doubtful | R: Adequate<br>C1: Adequate<br>C2: Adequate | Adequate | Very good | NR | NR | NR | NR | Very good | NR |
| Head[14] 2008 | SWBS | Inadequate | R: NR<br>C1:Doubt<br>C2: NR | Very good | Very good | NR | NR | NR | Very good | Very good | NR |
| Hueniken[31] 2020 | FIT | Inadequate | R: NR<br>C1: NR<br>C2: NR | Adequate | Very good | Doubtful | NR | NR | NR | Very good | Very good |
| Ripamonti[33] 2020 | COST-v2, Italian version | Inadequate | R: NR<br>C1: NR<br>C2: NR | Inadequate | Very good | NR | Inadequate | NR | Very good | Very good | NR |
| Riva[42] 2021 | PROFFIT | Adequate | R: Adequate<br>C1: Adequate<br>C2: NR | Adequate | Very good | NR | Adequate | NR | NR | NR | NR |
| Sharif[34] 2020 | COST-v2, Persian version | Inadequate | R: NR<br>C1: NR<br>C2: NR | Very good | Very good | NR | NR | NR | NR | Inadequate | NR |
| Shilling[35] 2018 | PRRS financial well-being domain | Inadequate | R: NR<br>C1: NR<br>C2: NR | Adequate | Very good | NR | Inadequate | NR | Inadequate | Inadequate | NR |

Continued

**Table 2** Continued

| Author (year) | PROM | Measurement property: methodological quality by study | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | PROM development | Content validity | Structural validity | Internal consistency | Cross-cultural validity | Reliability | Measurement error | Criterion validity | Hypothesis testing | Responsiveness |
| Tremblay[36] 2020 | P-SAFE, French Version | Inadequate | R: NR<br>C1: NR<br>C2: NR | NR | NR | Inadequate | NR | NR | NR | NR | NR |
| Veenstra[37] 2014 | PFB | Inadequate | R: NR<br>C1: NR<br>C2: NR | Very good | Inadequate | NR | NR | NR | NR | NR | NR |
| Wright[38] 2005 | SDI-21 providing for the family domain | Inadequate | R: Doubt<br>C1: NR<br>C2: NR | Very good | Very good | NR | NR | NR | NR | NR | NR |
| Wright[16] 2011 | SDI-16 money matters domain | Inadequate | R: Doubt<br>C1: NR<br>C2: NR | Very good | Very good | NR | Inadequate | NR | Inadequate | NR | NR |
| Yu[39] 2021 | COST-v1, Simplified Chinese version | Inadequate | R: NR<br>C1: NR<br>C2: NR | Very good | Very good | NR | Very good | NR | NR | Doubt | Doubt |
| Zebrack[40] 2010 | IOC-CS financial problems domain | Inadequate | R: NR<br>C1: NR<br>C2: Doubt | Adequate | Very good | NR | Adequate | NR | Inadequate | Very good | NR |
| Zhao[41] 2009 | CPILS employment/ financial domain | Inadequate | R: NR<br>C1: NR<br>C2: NR | Very good | Very good | NR | NR | NR | NR | Very good | NR |

CCEQ, Chronic Cancer Experiences Questionnaire; COST, Comprehensive Score for Financial Toxicity; CPILS, Cancer Problems in Living Scale; FIT, Financial Index of Toxicity; IOC-CS, Impact of Cancer-Childhood Survivors; NA, not applicable; NR, not report; PFB, Personal Financial Burden; PROFFIT, Patient-Reported Outcome for Fighting Financial Toxicity; PROM, Patient-Reported Outcome Measures; PRRS, Patient Roles and Responsibilities Scale; P-SAFE, Patient Self-Administered Financial Effects questionnaire; QLACS, Quality of Life in Adult Cancer Survivors; SDI, Social Difficulties Inventory Cancer Care Outcomes; SFDQ, Subjective Financial Distress Questionnaire; SWBS, Socioeconomic Well-being Scale.

**Table 3** Rating of measurement properties

| PROM | Author (year) | Structural validity | Internal consistency | Reliability | Measurement error | Hypothesis testing | Cross-cultural validity | Criterion validity | Responsiveness |
|---|---|---|---|---|---|---|---|---|---|
| CCEQ financial advice domain | Harley[29] 2019 | – (no data) | + (Cronbach's α=0.71–0.88) | NR | NR | + (r=0.35–0.39) | NR | NR | NR |
| COST-v1 | De Souza[12] 2014; De Souza[27] 2017 | NR | + (Cronbach's α=0.92) | + (ICC=0.80) | NR | + (r=0.20–0.42) | NR | NR | NR |
| COST-v1, *Australia version* | Durber[28] 2021 | NR | + (Cronbach's α=0.89) | + (ICC=0.80) | NR | + (r=-0.39–0.52) | NR | NR | NR |
| COST-v1, *Brazilian version* | de Alcantara Nogueira[32] 2020 | - (RMSEA=1.20) | + (Cronbach's α=0.83) | NR | NR | NR | – (p<0.01) | NR | NR |
| COST-v1, *India version* | Dar[25] 2021 | – (EFA: $\chi^2$=60.82) | ? (Cronbach's α=0.92) | NR | NR | NR | NR | NR | NR |
| COST-v1 *Simplified Chinese version* | Yu[39] 2021 | – (CFI=0.86; SRMR=0.08) | + (Cronbach's α=0.85–0.88) | + (ICC=0.85) | NR | + (r=-0.57–0.88) | NR | NR | ? (no clear hypothesis) |
| COST-v2, *Italian version* | Ripamonti[33] 2020 | ? (CFI=1.00; RMSEA=0.04) | ? (Cronbach's α=0.83) | + (ICC=0.79) | NR | + (r=-0.66–0.15) | NR | + (β=–0.55) | NR |
| COST-v2, *Persian version* | Sharif[34] 2020 | + (CFI=0.97; RMSEA=0.07) | + (Cronbach's α=0.89) | NR | NR | ? (no clear hypothesis) | NR | NR | NR |
| COST-v2, *Traditional Chinese version* | Chan[24] 2021 | – (CFI=0.91; RMSEA=0.15) | + (Cronbach's α=0.86) | + (ICC=0.71) | NR | + (r=-0.46–0.47) | NR | NR | NR |
| CPILS employment/ financial domain | Zhao[41] 2009 | – (EFA: no model data) | + (Cronbach's α=0.78–0.97) | NR | NR | ? (no clear hypothesis) | NR | NR | NR |
| FIT | Hueniken[31] 2020 | – (EFA: no model data) | + (Cronbach's α=0.77) | + (ICC=0.70) | NR | + (r=0.26–0.61) | NR | NR | + (r=–0.25) |
| IOC-CS financial problems domain | Zebrack[40] 2010 | – (EFA: no model data) | – (Cronbach's α=0.70–0.86) | – (ICC ≥0.75) | NR | – (>75% were not significant) | NR | ? (no data) | NR |
| PFB | Veenstra[37] 2014 | – (CFA: no model data) | + (Cronbach's α=0.79) | NR | NR | NR | NR | NR | NR |
| PROFFIT | Riva[42] 2021 | – (EFA: no model data) | + (Cronbach's α=0.87) | + (ICC=0.80) | NR | NR | NR | NR | NR |
| PRRS financial well-being domain | Shilling[35] 2018 | – (EFA: no model data) | + (Cronbach's α=0.90) | + (ICC=0.86) | NR | + (r=-0.71–0.65) | NR | + (r=–0.65) | NR |
| P-SAFE, *French Version* | Tremblay[36] 2020 | NR | NR | NR | NR | NR | ? (no data) | NR | NR |
| QLACS financial problems domain | Avis[23] 2005 | – (CFA: no model data) | + (Cronbach's α=0.72) | NR | NR | + (r=-0.57–0.60) | NR | – (r=–0.72) | NR |

Continued

**Table 3** Continued

| PROM | Author (year) | Structural validity | Internal consistency | Reliability | Measurement error | Hypothesis testing | Cross-cultural validity | Criterion validity | Responsiveness |
|---|---|---|---|---|---|---|---|---|---|
| SDI-16 money matters domain | Wright[16] 2011 | – (EFA: no model data) | + (Cronbach's α=0.71–0.82) | NR | NR | NR | NR | NR | NR |
| SDI-21 providing for the family domain | Wright[38] 2005 | – (EFA: no model data) | + (Cronbach's α=0.50–0.86) | – (Weighted kappa=0.54–0.80) | NR | – (33% were not significant) | NR | ? (r=−0.72) | NR |
| SFDQ | Dar[25] 2021 | + (CFI=0.98, TLI=0.97, RMSEA=0.045, SRMR=0.014) | + (Cronbach's α=0.85–0.88) | NR | NR | NR | NR | NR | NR |
| SWBS | Head[14] 2008 | – (CFA: no model data) | + (Cronbach's α=0.92) | NR | NR | + (r=−0.57–0.60) | NR | + (r=−0.12–0.03) | NR |

CCEQ, Chronic Cancer Experiences Questionnaire; CFA, Confirmatory Factor Analysis; CFI, Comparative Fit Index; COST, Comprehensive Score for Financial Toxicity; CPILS, Cancer Problems in Living Scale; EFA, Exploratory Factor Analysis; FIT, Financial Index of Toxicity; ICC, Intraclass Correlation Coefficient; IOC-CS, Impact of Cancer-Childhood Survivors; NA, not applicable; NR, not report; PFB, Personal Financial Burden ; PROFFIT, Patient-Reported Outcome for Fighting Financial Toxicity; PROM, patient-reported outcome measures; PRRS, Patient Roles and Responsibilities Scale ; P-SAFE, Patient Self-Administered Financial Effects questionnaire; QLACS, Quality of Life in Adult Cancer Survivors; RMSEA, Root Mean Square Error of Approximation; SDI, Social Difficulties Inventory Cancer Care Outcomes; SFDQ, Subjective Financial Distress Questionnaire; SRMR, Standardized Root Mean Square Residual; SWBS, Socioeconomic Well-being Scale.

seven versions were rated as having high evidence of structural validity, internal consistency, hypothesis testing and criterion validity.[12 24 25 27 28 32–34 39] The Financial Index of Toxicity (FIT) and Impact of Cancer-Childhood Survivors (IOC-CS) financial problems domain reported data on five properties and were rated on a scale from 'very low evidence' to 'high evidence'.[31 40]

## DISCUSSION

This systematic review identified 21 PROMs and domains of PROMs evaluating FT in cancer survivors, including the COST (original, Brazilian, India, Italian, Persian, Simplified Chinese, Traditional Chinese version), FIT, Personal Financial Burden, P-SAFE, SWBS, Quality of Life in Adult Cancer Survivors (QLACS) financial problems domain, Chronic Cancer Experiences Questionnaire financial advice domain, Patient-Reported Outcome for Fighting Financial Toxicity (PROFFIT), Patient Roles and Responsibilities Scale financial well-being domain, SDI-21 providing for the family domain, SDI-16 money matters domain, SFDQ, IOC-CS financial problems domain and Cancer Problems in Living Scale (CPILS) employment/financial domain. Overall, the COST had a complete development process compared with other PROMs and showed the best psychometric properties, especially in terms of internal consistency, reliability and hypothesis testing. To the best of our knowledge, this is the first systematic review that has summarised the psychometric properties of FT PROMs in cancer survivors and reported the certainty of evidence for each property of PROMs. The results may provide quantitative evidence for researchers and healthcare professionals to choose PROMs measuring cancer survivors' FT in future scientific research and clinical practice.

The results highlighted that the COST (of which we studied both version 1 and version 2) had better psychometric properties than other specific and generic PROMs in terms of internal consistency, reliability and hypothesis testing. The COST could be recommended as the most suitable worldwide available measure for use in research and clinical practice across different contexts. Other systematic reviews have also suggested that the COST is a promising measure from a content perspective.[10 11] From a psychometric standpoint, there are a few issues that one must face when evaluating financial toxicity in cancer survivors using the COST. First, caution should be taken when using the COST in different socioeconomic conditions outside the USA. In some countries in Europe or Asia, the majority of medical expenses are covered by social health insurance, and direct out-of-pocket payments are replaced by prepayment from health insurance contributions.[43 44] In addition, social security systems can benefit cancer survivors who are not able to work.[45] These two socioeconomic factors may affect cancer survivors' understanding regarding some items related to medical spending and indirect cost. However, few COST

**Table 4** Certainty of evidence of measurement properties

| PROM | Author (year) | Structural validity | Internal consistency | Reliability | Measurement error | Hypothesis testing | Cross-cultural validity | Criterion validity | Responsiveness |
|---|---|---|---|---|---|---|---|---|---|
| CCEQ financial advice domain | Harley[29] 2019 | Moderate | High | – | – | High | – | – | – |
| COST-v1 | De Souza[12] 2014; De Souza[27] 2017 | – | High | Moderate | – | High | – | – | – |
| COST-v1, Australia version | Durber[28] 2021 | – | High | Moderate | – | High | – | – | – |
| COST-v1, Brazilian version | de Alcantara Nogueira[32] 2020 | High | High | – | – | – | Very low | – | – |
| COST-v1, India version | Dar[25] 2021 | Very Low | Low | – | – | – | – | – | – |
| COST-v1, Simplified Chinese version | Yu[39] 2021 | High | High | High | – | Low | – | – | Low |
| COST-v2, Italian version | Ripamonti[33] 2020 | Very low | High | Very low | – | High | – | High | – |
| COST-v2, Persian version | Sharif[34] 2020 | High | High | – | – | Very low | – | – | – |
| COST-v2, Traditional Chinese version | Chan[24] 2021 | High | High | Low | – | High | – | – | – |
| CPILS employment/ financial domain | Zhao[41] 2009 | High | High | – | – | High | – | – | – |
| FIT | Hueniken[31] 2020 | Low | Moderate | Very Low | – | Moderate | – | – | Moderate |
| IOC-CS financial problems domain | Zebrack[40] 2010 | Moderate | High | Moderate | – | High | – | Very low | – |
| PFB | Veenstra[37] 2014 | Moderate | Very low | – | – | – | – | – | – |
| PROFFIT | Riva[42] 2021 | Moderate | High | Moderate | – | – | – | – | – |
| PRRS financial well-being domain | Shilling[35] 2018 | Moderate | High | Very low | – | Very low | – | Very low | – |
| P-SAFE, French Version | Tremblay[36] 2020 | – | – | – | – | – | Very low | – | – |
| QLACS financial problems domain | Avis[23] 2005 | Moderate | High | – | – | Low | – | Moderate | – |
| SDI-16 money matters domain | Wright[16] 2011 | High | High | Very low | – | – | – | Very low | – |
| SDI-21 providing for the family domain | Wright[38] 2005 | High | High | – | – | – | – | – | – |
| SFDQ | Dar[26] 2022 | High | High | – | – | – | – | – | – |
| SWBS | Head[14] 2008 | High | High | – | – | High | – | – | – |

CCEQ, Chronic Cancer Experiences Questionnaire; COST, Comprehensive Score for Financial Toxicity; CPILS, Cancer Problems in Living Scale; FIT, Financial Index of Toxicity; IOC-CS, Impact of Cancer-Childhood Survivors; PFB, Personal Financial Burden; PROM, Patient-Reported Outcome Measures; PROFFIT, Patient-Reported Outcome for Fighting Financial Toxicity; PRRS, Patient Roles and Responsibilities Scale ; P-SAFE, Patient Self-Administered Financial Effects questionnaire; QLACS, Quality of Life in Adult Cancer Survivors; RMSEA, Root Mean Square Error of Approximation; SDI, Social Difficulties Inventory Cancer Care Outcomes; SFDQ, Subjective Financial Distress Questionnaire; SWBS, Socioeconomic Well-being Scale.

validation studies have considered socioeconomic issues, adapted the measure in a local context or provided data on cross-cultural validity. It is recommended that future COST validation studies recruit cancer survivors across multiple social and cultural backgrounds to assess cross-cultural measurement invariance.

Second, the original construct and item generation for the COST were based on a literature search; thus, the theoretical grounds for the measure are unclear, and the instrument may not capture detailed information related to the construct. Theoretical frameworks and conceptual models are crucial for self-reported measures to capture subtle changes in constructs.[46] Although FT is a relatively new concept, certain models can guide item generation in the development of future FT PROMs. Tucker-Seeley and colleagues developed a conceptual model of FT and emphasised three components of financial burden, namely, the material, psychosocial and behavioural domains.[47] Head developed SWBS based on James Coleman's Theory of Social Class; this scale contains 17 items across 3 domains: human capital, material capital and social capital.[14 30 48] Witte *et al*'s systematic review analysed 352 different questions regarding financial spending and found six domains (financial spending, financial resources, psychosocial affect, support seeking, coping care and coping lifestyle) that can represent reactions to subjective financial distress.[10] Other theories and models, including the Wreckers theory of financial distress, ecological theory and the functionalist tradition, have also been widely used in cancer survivors.[49–51] With the increasing number of theoretical studies related to FT, the theoretical grounds for future PROMs need to be clarified.

In addition to the COST, two other PROMs, namely, the FIT and the IOC-CS financial problems domains, also provided adequate data on psychometric properties. The FIT is relatively new and has fewer items than the other included measures. This measure was developed by Hueniken *et al* and has been validated only in survivors with head and neck cancer.[31] Head and neck cancer, especially laryngeal and hypopharyngeal cancer, has particularly large impacts on survivors' daily function (eg, speech and eating) after treatment and affects survivors' ability to return to work.[52 53] Only 32%–59% of head and neck cancer survivors return to work after treatment.[54] This form of cancer also has short-term and long-term financial consequences for caregivers and their families.[55] Therefore, future studies should be aware that the FIT may not be directly applicable to other cancer populations.

Regarding PROM development, we found that only two PROMs, PROFFIT and SFDQ, were not developed in the context of English-speaking developed countries such as the USA, the UK and Canada. The socioeconomic contexts and healthcare systems in these countries may be significantly different from those in other parts of the world and ultimately lead to a nuance in the perceived causes and consequences of FT. Previous studies have reported that FT is closely related to broad social determinants of economic circumstances. Factors including healthcare policy, healthcare system, insurance system, specific micro contexts and the level of regional economic development could not only affect the cancer survivors' perceived level of FT but also determine the origins of FT.[56 57] Additionally, cultural factors (eg, a cultural emphasis on saving and a cultural imperative to have a large family) also affect cancer survivors' perceived financial security and economic burden.[58]

PROFFIT, which was developed in 2021 in the Italian context, also reported higher quality PROM development and content validity than other PROMs. We would consider it to be a good FT PROM against the COSMIN criteria if more validation studies were conducted to report a greater effect size of the measurement properties. Therefore, we recommend that researchers use context-specific measures to assess FT in cancer survivors (eg, using PROFFIT in Italy). Further studies are warranted to develop various FT PROMs based on different social and cultural backgrounds. Worldwide measures, such as COST, should be analysed to determine the differences between social, cultural and economic contexts.

### Limitations

We acknowledge that there are some limitations to this study. First, this review included only studies that aimed to evaluate the measurement properties of FT PROMs. Many studies that aimed to explore the level of FT in cancer survivors also reported the reliability and validity of PROMs. Therefore, the PROMs we summarised in this systematic review had higher psychometric quality than other measures that we did not list in this review. Second, we included only studies published in English. Therefore, studies published in other languages were not included, which may affect the conclusion of this review. Third, we included only the original version of the FT domain from PROMs assessing quality of life in cancer survivors, such as the EORTC QLQ-C30 and the QLACS. Over 20 language versions of these PROMs do not provide sufficient details on the FT domain individually.

## CONCLUSION

This systematic review summarised the psychometric properties of 20 PROMs evaluating FT in cancer survivors. The findings highlighted that, from a psychometric property perspective, the COST had an adequate PROM development process and showed the best psychometric properties among all examined PROMs, especially in internal consistency, reliability and hypothesis testing; thus, we recommend the COST as the most suitable worldwide available measures for use in research and clinical practice across different contexts. The FIT and the IOC-CS financial problems domain also had adequate psychometric properties. We suggest that PROMs should be selected only after careful consideration of the local socioeconomic context. Future studies are warranted to develop various FT PROMs based on different social and

cultural backgrounds and a clear theoretical basis for assessing FT.

**Twitter** Lucylynn Lizarondo @lucylizarondo20

**Contributors** WX took full responsibility for the work, had access to the data, and controlled the decision to publish. ZZ and WX designed the systematic review, conducted data searching, extraction and analysis, assessing the certainty of evidence, and wrote the draft of the manuscript. HW and YS conducted quality appraisal and assessing the certainty of evidence. WKWS, LL, JP and YH provided critical comments. All authors approved the final version of the manuscript. WX is the guarantor.

**Funding** This work was supported by National Natural Science Foundation of China (Grant number: 72004034), China Medical Board Open Competition Program (Grant number: 20-371), Shanghai Pujiang Program (Grant number: 2019PJC017), and Shanghai Soft Science Key Program (Grant number: 20692104800)

**Competing interests** None declared.

**Patient and public involvement** Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

**Patient consent for publication** Not applicable.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** No data are available.



**ORCID iD**

Zheng Zhu http://orcid.org/0000-0001-9651-8311

## REFERENCES

1 Bosland MC. Global health and cancer. *Lancet* 2019;393:983.
2 de Souza JA, Wong Y-N. Financial distress in cancer patients. *J Med Person* 2013;11:73–7.
3 World cancer report [Internet]. Available: https://publications.iarc.fr/586 [Accessed 31 July 2021].
4 Global Oncology Trends 2021[Internet], 2021. Available: https://www.iqvia.com/insights/the-iqvia-institute/reports/global-oncology-trends-2021 [Accessed 31 July 2021].
5 Mariotto AB, Yabroff KR, Shao Y, *et al*. Projections of the cost of cancer care in the United States: 2010-2020. *J Natl Cancer Inst* 2011;103:117–28.
6 ACTION Study Group, Jan S, Kimman M, *et al*. Financial catastrophe, treatment discontinuation and death associated with surgically operable cancer in south-east Asia: results from the action study. *Surgery* 2015;157:971–82.
7 Mackintosh M, Channon A, Karan A, *et al*. What is the private sector? understanding private provision in the health systems of low-income and middle-income countries. *Lancet* 2016;388:596–605.
8 Zafar SY, Yousuf Zafar S. Financial toxicity of cancer care: it's time to intervene. *J Natl Cancer Inst* 2016;108:djv370.
9 Carrera PM, Kantarjian HM, Blinder VS. The financial burden and distress of patients with cancer: understanding and stepping-up action on the financial toxicity of cancer treatment. *CA Cancer J Clin* 2018;68:153–65.
10 Witte J, Mehlis K, Surmann B, *et al*. Methods for measuring financial toxicity after cancer diagnosis and treatment: a systematic review and its implications. *Ann Oncol* 2019;30:1061–70.
11 Gordon LG, Merollini KMD, Lowe A, *et al*. A Systematic review of financial toxicity among cancer survivors: we can't pay the co-pay. *Patient* 2017;10:295–309.
12 de Souza JA, Yap BJ, Hlubocky FJ, *et al*. The development of a financial toxicity patient-reported outcome in cancer: the cost measure. *Cancer* 2014;120:3245–53.
13 Given BA, Given CW, Stommel M. Family and out-of-pocket costs for women with breast cancer. *Cancer Pract* 1994;2:187–93.
14 Head BA, Faul AC. Development and validation of a scale to measure socioeconomic well-being in persons with cancer. *J Support Oncol* 2008;6:183–92.
15 Prawitz AD, Garman ET, Sorhaindo B. InCharge financial distress/financial well-being scale: development, administration, and score interpretation. *Financial Counsel Plann* 2006;17:34–50.
16 Wright P, Smith AB, Keding A, *et al*. The social difficulties inventory (SDI): development of subscales and scoring guidance for staff. *Psychooncology* 2011;20:36–43.
17 Malin JL, Ko C, Ayanian JZ, *et al*. Understanding cancer patients' experience and outcomes: development and pilot study of the cancer care outcomes research and surveillance patient survey. *Support Care Cancer* 2006;14:837–48.
18 Ripamonti C, Leporati R, De Feo G, *et al*. Italian version of the edmonton symptom assessment system (ESAS)-total care (TC): development and psychometric validation in patients undergoing cancer treatment or follow-up. *Support Care Cancer* 2022;30:1923–33.
19 Salsman JM, Danhauer SC, Moore JB, *et al*. Systematic review of financial burden assessment in cancer: evaluation of measures and utility among adolescents and young adults and caregivers. *Cancer* 2021;127:1739–48.
20 Prinsen CAC, Mokkink LB, Bouter LM, *et al*. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27:1147–57.
21 Page MJ, McKenzie JE, Bossuyt PM, *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71.
22 Zhu Z, Xing W, Lizarondo L, *et al*. Psychometric properties of self-reported financial toxicity measures in cancer survivors: a systematic review protocol using COSMIN methodology. *BMJ Open* 2020;10:e036365.
23 Avis NE, Smith KW, McGraw S, *et al*. Assessing quality of life in adult cancer survivors (QLACS). *Qual Life Res* 2005;14:1007–23.
24 Chan DNS, Choi KC, Ng MSN, *et al*. Translation and validation of the traditional Chinese version of the comprehensive score for financial toxicity-functional assessment of chronic illness therapy (version 2). *Health Qual Life Outcomes* 2021;19:17.
25 Dar MA, Chauhan R, Sharma KK, *et al*. Assessing the reliability and validity of comprehensive score for financial toxicity (cost) among radiation oncology patients in India: a cross-sectional pilot study. *Ecancermedicalscience* 2021;15:1219.
26 Dar MA, Chauhan R, Murti K, *et al*. Development and validation of subjective financial distress questionnaire (SFDQ): a patient reported outcome measure for assessment of financial toxicity among radiation oncology patients. *Front Oncol* 2022;11:819313.
27 de Souza JA, Yap BJ, Wroblewski K, *et al*. Measuring financial toxicity as a clinically relevant patient-reported outcome: the validation of the comprehensive score for financial toxicity (cost). *Cancer* 2017;123:476–84.
28 Durber K, Halkett GK, McMullen M, *et al*. Measuring financial toxicity in Australian cancer patients - validation of the comprehensive score for financial toxicity (FACT COST) measuring financial toxicity in Australian cancer patients. *Asia Pac J Clin Oncol* 2021;17:377–87.
29 Harley C, Pini S, Kenyon L, *et al*. Evaluating the experiences and support needs of people living with chronic cancer: development and initial validation of the chronic cancer experiences questionnaire (CCEQ). *BMJ Support Palliat Care* 2019;9:e15.
30 Head BA. Development and validation of a scale to measure health-related socioeconomic well-being in persons with a cancer diagnosis. electronic theses and Dissertations. paper 594
31 Hueniken K, Douglas CM, Jethwa AR, *et al*. Measuring financial toxicity incurred after treatment of head and neck cancer: development and validation of the financial index of toxicity questionnaire. *Cancer* 2020;126:4042–50.
32 de Alcantara Nogueira L, Koller FJ, Marcondes L, *et al*. Validation of the comprehensive score for financial toxicity for Brazilian culture. *Ecancermedicalscience* 2020;14:1158.
33 Ripamonti CI, Chiesi F, Di Pede P, *et al*. The validation of the Italian version of the comprehensive score for financial toxicity (cost). *Support Care Cancer* 2020;28:4477–85.

34 Sharif SP, Sim OF, Moshtagh M. Psychometric evaluation of the comprehensive score for financial toxicity scale among iranian cancer patients. *Research Square* 2020.
35 Shilling V, Starkings R, Jenkins V, *et al*. Development and validation of the patient roles and responsibilities scale in cancer patients. *Qual Life Res* 2018;27:2923–34.
36 Tremblay D, Poder TG, Vasiliadis H-M, *et al*. Translation and cultural adaptation of the patient self-administered financial effects (P-SAFE) questionnaire to assess the financial burden of cancer in French-speaking patients. *Healthcare* 2020;8:366.
37 Veenstra CM, Regenbogen SE, Hawley ST, *et al*. A composite measure of personal financial burden among patients with stage III colorectal cancer. *Med Care* 2014;52:957–62.
38 Wright EP, Kiely M, Johnston C, *et al*. Development and evaluation of an instrument to assess social difficulties in routine oncology practice. *Qual Life Res* 2005;14:373–86.
39 Yu H-H, Yu Z-F, Li H, *et al*. The comprehensive score for financial toxicity in China: validation and responsiveness. *J Pain Symptom Manage* 2021;61:1297–304.
40 Zebrack BJ, Donohue JE, Gurney JG, *et al*. Psychometric evaluation of the impact of cancer (IOC-CS) scale for young adult survivors of childhood cancer. *Qual Life Res* 2010;19:207–18.
41 Zhao L, Portier K, Stein K, *et al*. Exploratory factor analysis of the cancer problems in living scale: a report from the American cancer society's studies of cancer survivors. *J Pain Symptom Manage* 2009;37:676–86.
42 Riva S, Arenare L, Di Maio M, *et al*. Cross-Sectional study to develop and describe psychometric characteristics of a patient-reported instrument (PROFFIT) for measuring financial toxicity of cancer within a public healthcare system. *BMJ Open* 2021;11:e049128.
43 De Lorenzo F, Apostolidis K, Florindi F, *et al*. Improving European policy to support cancer survivors. *J Cancer Policy* 2018;15:72–5.
44 Bhoo-Pathy N, Ng C-W, Lim GC-C, *et al*. Financial toxicity after cancer in a setting with universal health coverage: a call for urgent action. *J Oncol Pract* 2019;15:e537–46.
45 Philip CC, Mathew A, John MJ. Cancer care: challenges in the developing world. *Cancer Res Stat Treat* 2018;1:58.
46 Mokkink LB, Boers M, van der Vleuten CPM, *et al*. COSMIN risk of bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a delphi study. *BMC Med Res Methodol* 2020;20:293.
47 Tucker-Seeley RD, Thorpe RJ. Material-psychosocial-behavioral aspects of financial hardship: a conceptual model for cancer prevention. *Gerontologist* 2019;59:S88–93.
48 Coleman JS. *Foundations of social theory*. Harvard university press, 1994.
49 A 'Wreckers Theory'of Financial Distress. [Internet], 2005. Available: https://www.econstor.eu/bitstream/10419/19625/1/200540dkp.pdf [Accessed 31 July 2021].
50 Bergeron K, Abdi S, DeCorby K, *et al*. Theories, models and frameworks used in capacity building interventions relevant to public health: a systematic review. *BMC Public Health* 2017;17:914.
51 Acker J, Baca-Zinn M, Collins PH. *Social class and stratification: classic statements and theoretical debates*. Rowman & Littlefield Publishers, 2006.
52 Miller A. Returning to work after head and neck cancer. *Curr Opin Otolaryngol Head Neck Surg* 2020;28:155–60.
53 Koch R, Wittekindt C, Altendorf-Hofmann A, *et al*. Employment pathways and work-related issues in head and neck cancer survivors. *Head Neck* 2015;37:585–93.
54 Giuliani M, Papadakos J, Broadhurst M, *et al*. The prevalence and determinants of return to work in head and neck cancer survivors. *Support Care Cancer* 2019;27:539–46.
55 Dewa CS, Trojanowski L, Tamminga SJ, *et al*. Work-Related experiences of head and neck cancer survivors: an exploratory and descriptive qualitative study. *Disabil Rehabil* 2018;40:1252–8.
56 Zhu Z, Xing W, Zhang X, *et al*. Cancer survivors' experiences with financial toxicity: a systematic review and meta-synthesis of qualitative studies. *Psychooncology* 2020;29:945–59.
57 Chebli P, Lemus J, Avila C, *et al*. Multilevel determinants of financial toxicity in breast cancer care: perspectives of healthcare professionals and latina survivors. *Support Care Cancer* 2020;28:3179–88.
58 Kong Y-C, Wong L-P, Ng C-W, *et al*. Understanding the financial needs following diagnosis of breast cancer in a setting with universal health coverage. *Oncologist* 2020;25:497–504.

**Appendix I Search strategy and results**

**Search strategy for PubMed**

Search time: 2022-3-13 15:07 (UTC+8)

| Search | Query | Items found |
|---|---|---|
| #1 | Cancer[Title/Abstract] OR neoplasms[MeSH] | 4176348 |
| #2 | Patient?[Title/Abstract] OR survivor?[Title/Abstract] OR patients[MeSH] OR "cancer survivors"[MeSH] OR survivors[MeSH] | 2721864 |
| #3 | Cost[Title/Abstract] OR bill?[Title/Abstract] OR expense[Title/Abstract] OR "productivity loss"[Title/Abstract] OR "out-of-pocket"[Title/Abstract] OR "economic burden"[Title/Abstract] OR "financial toxicity"[Title/Abstract] OR "financial hardship"[Title/Abstract] OR "financial burden"[Title/Abstract] OR "financial effect"[Title/Abstract] OR "financial stress"[Title/Abstract] OR "economic burden"[Title/Abstract] OR "economic hardship"[Title/Abstract] OR "co-payment"[Title/Abstract] OR "cost of illness"[MeSH] | 573062 |
| #4 | Scale?[Title/Abstract] OR "patient reported outcome measur*"[Title/Abstract] OR PROM? [Title/Abstract] OR measure* [Title/Abstract] OR "Patient Reported Outcome Measures*"[MeSH] OR "Surveys and Questionnaires"[MeSH] | 5065964 |
| #5 | (instrumentation[sh] OR methods[sh] OR "Validation Studies"[pt] OR "Comparative Study"[pt] OR "psychometrics"[MeSH] OR psychometr*[tiab] OR clinimetr*[tw] OR clinometr*[tw] OR "outcome assessment (health care)"[MeSH] OR "outcome assessment"[tiab] OR "outcome measure*"[tw] OR "observer variation"[MeSH] OR "observer variation"[tiab] OR "Health Status Indicators"[Mesh] OR "reproducibility of results"[MeSH] OR reproducib*[tiab] OR "discriminant analysis"[MeSH] OR reliab*[tiab] OR unreliab*[tiab] OR valid*[tiab] OR "coefficient of variation"[tiab] OR coefficient[tiab] OR homogeneity[tiab] OR homogeneous[tiab] OR "internal consistency"[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR (item[tiab] AND (correlation*[tiab] OR selection*[tiab] OR reduction*[tiab])) OR agreement[tw] OR precision[tw] OR imprecision[tw] OR "precise values"[tw] OR test-retest[tiab] OR (test[tiab] AND retest[tiab]) OR (reliab*[tiab] AND (test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] OR intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-tester[tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intra-observer[tiab] OR intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intraassay[tiab] OR intra-assay[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[tiab] OR kappa's[tiab] OR kappas[tiab] OR repeatab*[tw] OR ((replicab*[tw] OR repeated[tw]) AND (measure[tw] OR measures[tw] OR findings[tw] OR result[tw] OR results[tw] OR test[tw] OR tests[tw])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR "known group"[tiab] OR "factor analysis"[tiab] OR "factor analyses"[tiab] OR "factor structure"[tiab] OR "factor structures"[tiab] OR dimension*[tiab] OR subscale*[tiab] OR (multitrait[tiab] AND scaling[tiab] AND (analysis[tiab] OR analyses[tiab])) OR "item discriminant"[tiab] OR "interscale correlation*"[tiab] OR error[tiab] OR errors[tiab] OR "individual variability"[tiab] | 10114928 |

| | | |
|---|---|---|
| | OR “interval variability”[tiab] OR “rate variability”[tiab] OR (variability[tiab] AND (analysis[tiab] OR values[tiab])) OR (uncertainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR “standard error of measurement”[tiab] OR sensitiv*[tiab] OR responsive*[tiab] OR (limit[tiab] AND detection[tiab]) OR “minimal detectable concentration”[tiab] OR interpretab*[tiab] OR ((minimal[tiab] OR minimally[tiab] OR clinical[tiab] OR clinically[tiab]) AND (important[tiab] OR significant[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR (small*[tiab] AND (real[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR “meaningful change”[tiab] OR “ceiling effect”[tiab] OR “floor effect”[tiab] OR “Item response model”[tiab] OR IRT[tiab] OR Rasch[tiab] OR “Differential item functioning”[tiab] OR DIF[tiab] OR “computer adaptive testing”[tiab] OR “item bank”[tiab] OR “cross-cultural equivalence”[tiab]) | |
| #6 | (“addresses”[Publication Type] OR “biography”[Publication Type] OR “case reports”[Publication Type] OR “comment”[Publication Type] OR “directory”[Publication Type] OR “editorial”[Publication Type] OR “festschrift”[Publication Type] OR “interview”[Publication Type] OR “lectures”[Publication Type] OR “legal cases”[Publication Type] OR “legislation”[Publication Type] OR “letter”[Publication Type] OR “news”[Publication Type] OR “newspaper article”[Publication Type] OR “patient education handout”[Publication Type] OR “popular works”[Publication Type] OR “congresses”[Publication Type] OR “consensus development conference”[Publication Type] OR “consensus development conference, nih”[Publication Type] OR “practice guideline”[Publication Type]) NOT (“animals”[MeSH Terms] NOT “humans”[MeSH Terms]) | 4415236 |
| #7 | #1 AND #2 AND #3 AND #4 AND #5 | 4097 |
| #8 | #7 NOT #6 | 4003 |

**Search strategy for MEDLINE (Ovid)**

Search time: 2022-3-13 15:12 (UTC+8)

| **Search** | **Query** | **Items found** |
|---|---|---|
| #1 | cancer.ab. or cancer.ti. or neoplasms.hw. | 3208957 |
| #2 | patient?.ab. or patient?.ti. or survivor?.ab. or survivor?.ti. or patients.hw. or cancer survivors.hw. or survivors.hw. | 6631761 |
| #3 | Cost.ab. or Cost.ti. or bill?.ab. or bill?.ti. or expense.ab. or expense.ti. or productivity loss.ab. or productivity loss.ti. or out-of-pocket.ab. or out-of-pocket.ti. or economic burden.ab. or economic burden.ti. or financial toxicity.ab. or financial toxicity.ti. or financial hardship.ab. or financial hardship.ti. or financial burden.ab. or financial burden.ti. or financial effect.ab. or financial effect.ti. or financial stress.ab. or financial stress.ti. or economic burden.ab. or economic burden.ti. or economic hardship.ab. or economic hardship.ti. or co-payment.ab. or co-payment.ti. or (cost of illness).hw. | 459262 |
| #4 | Scale?.ab. or Scale?.ti. or patient reported outcome measur*.ab. or patient reported outcome measur*.ti. or PROM?.ab. or PROM?.ti. or measure*.ab. or measure*.ti. or Patient Reported Outcome Measures*.ab. or Patient Reported Outcome Measures*.ti. or (Surveys and Questionnaires).hw. | 3951480 |
| #5 | (instrumentation or methods).fs. | 4440288 |
| #6 | (Validation Studies or Comparative Study).pt. | 1910291 |

| | | |
|---|---|---|
| 7 | exp Psychometrics/ | 83581 |
| 8 | psychometr*.ti,ab. | 46299 |
| 9 | (clinimetr* or clinometr*).tw. | 1135 |
| 10 | exp Outcome Assessment Health Care/ | 1270817 |
| 11 | outcome assessment.ti,ab. | 3870 |
| 12 | outcome measure*.tw. | 230270 |
| 13 | exp Observer Variation/ | 44491 |
| 14 | observer variation.ti,ab. | 1055 |
| 15 | exp Health Status Indicators/ | 335480 |
| 16 | exp Reproducibility of Results/ | 442637 |
| 17 | reproducib*.ti,ab. | 152648 |
| 18 | exp Discriminant Analysis/ | 11624 |
| 19 | (reliab* or unreliab* or valid* or coefficient or homogeneity or homogeneous or internal consistency).ti,ab. | 1334500 |
| 20 | (cronbach* and (alpha or alphas)).ti,ab. | 22858 |
| 21 | (item and (correlation* or selection* or reduction*)).ti,ab. | 22439 |
| 22 | (agreement or precision or imprecision or precise values or test-retest).ti,ab. | 369056 |
| 23 | (test and retest).ti,ab. | 26981 |
| 24 | (reliab* and (test or retest)).ti,ab. | 87641 |
| 25 | (stability or interrater or inter-rater or intrarater or intra-rater or intertester or inter-tester or intratester or intra-tester or interobserver or inter-observer or intraobserver or intraobserver or intertechnician or inter-technician or intratechnician or intra-technician or interexaminer or inter-examiner or intraexaminer or intra-examiner or interassay or interassay or intraassay or intra-assay or interindividual or inter-individual or intraindividual or intra-individual or interparticipant or inter-participant or intraparticipant or intra-participant or kappa or kappa's or kappas or repeatab*).ti,ab. | 551861 |
| 26 | ((replicab* or repeated) and (measure or measures or findings or result or results or test or tests)).ti,ab. | 192657 |
| 27 | (generaliza* or generalisa* or concordance).ti,ab. | 87243 |
| 28 | (intraclass and correlation*).ti,ab. | 26027 |
| 29 | (discriminative or known group or factor analysis or factor analyses or dimension* or subscale*).ti,ab. | 547916 |
| 30 | (multitrait and scaling and (analysis or analyses)).ti,ab. | 134 |
| 31 | (item discriminant or interscale correlation* or error or errors or individual variability).ti,ab. | 273771 |
| 32 | (variability and (analysis or values)).ti,ab. | 97699 |
| 33 | (uncertainty and (measurement or measuring)).ti,ab. | 5737 |
| 34 | (standard error of measurement or sensitiv* or responsive*).ti,ab. | 1522843 |
| 35 | ((minimal or minimally or clinical or clinically) and (important or significant or detectable) and (change or difference)).ti,ab. | 236237 |
| 36 | (small* and (real or detectable) and (change or difference)).ti,ab. | 7151 |
| 37 | (meaningful change or ceiling effect or floor effect or Item response model or IRT or Rasch or Differential item functioning or DIF or computer adaptive testing or item bank or cross-cultural equivalence).ti,ab. | 13325 |

| | | |
|---|---|---|
| 38 | 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 | 9593937 |
| 39 | (child* or pediatric* or infan* or neonat* or newborn* or teen* or youth*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept, rare disease supplementary concept, unique identifier] | 3289302 |
| 40 | 1 and 2 and 3 and 4 and 38 | 4635 |
| 41 | 40 not 39 | 4383 |

**Search strategy for EMBASE (Ovid)**

Search time: 2022-3-13 16:17 (UTC+8)

| **Search** | **Query** | **Items found** |
|---|---|---|
| #1 | cancer.ab. or cancer.ti. or neoplasms.hw. | 2728162 |
| #2 | patient?.ab. or patient?.ti. or survivor?.ab. or survivor?.ti. or patients.hw. or cancer survivors.hw. or survivors.hw. | 10910535 |
| #3 | Cost.ab. or Cost.ti. or bill?.ab. or bill?.ti. or expense.ab. or expense.ti. or productivity loss.ab. or productivity loss.ti. or out-of-pocket.ab. or out-of-pocket.ti. or economic burden.ab. or economic burden.ti. or financial toxicity.ab. or financial toxicity.ti. or financial hardship.ab. or financial hardship.ti. or financial burden.ab. or financial burden.ti. or financial effect.ab. or financial effect.ti. or financial stress.ab. or financial stress.ti. or economic burden.ab. or economic burden.ti. or economic hardship.ab. or economic hardship.ti. or co-payment.ab. or co-payment.ti. or (cost of illness).hw. | 743789 |
| #4 | Scale?.ab. or Scale?.ti. or patient reported outcome measur*.ab. or patient reported outcome measur*.ti. or PROM?.ab. or PROM?.ti. or measure*.ab. or measure*.ti. or Patient Reported Outcome Measures*.ab. or Patient Reported Outcome Measures*.ti. or (Surveys and Questionnaires).hw. | 5577750 |
| #5 | exp Intermethod comparison/ OR exp data collection method/ OR exp validation study/ OR exp feasibility study/ OR exp pilot study/ OR exp psychometry/ OR exp reproducibility/ OR reproducib*:ab,ti OR audit:ab,ti OR psychometr*:ab,ti OR clinimetr*:ab,ti OR clinometr*:ab,ti OR exp observer variation/ OR observer variation:ab,ti OR exp discriminant analysis/ OR exp validity/ OR reliab*:ab,ti OR valid*:ab,ti OR coefficient:ab,ti OR internal consistency:ab,ti OR (cronbach*:ab,ti AND (alpha:ab,ti OR alphas:ab,ti)) OR item correlation:ab,ti OR item correlations:ab,ti OR item selection:ab,ti OR item selections:ab,ti OR item reduction:ab,ti OR item reductions:ab,ti OR agreement:ab,ti OR precision:ab,ti OR imprecision:ab,ti OR precise values:ab,ti OR test-retest:ab,ti OR (test:ab,ti AND retest:ab,ti) OR (reliab*:ab,ti AND (test:ab,ti OR retest:ab,ti)) OR stability:ab,ti OR interrater:ab,ti OR inter-rater:ab,ti OR intrarater:ab,ti OR intra-rater:ab,ti OR intertester:ab,ti OR inter-tester:ab,ti OR intratester:ab,ti OR intratester:ab,ti OR interobeserver:ab,ti OR inter-observer:ab,ti OR intraobserver:ab,ti OR intraobserver:ab,ti OR intertechnician:ab,ti OR inter-technician:ab,ti OR intratechnician:ab,ti OR intratechnician:ab,ti OR interexaminer:ab,ti OR inter-examiner:ab,ti OR intraexaminer:ab,ti OR intraexaminer:ab,ti OR interassay:ab,ti OR inter-assay:ab,ti OR intraassay:ab,ti OR intra-assay:ab,ti OR interindividual:ab,ti OR inter-individual:ab,ti OR intraindividual:ab,ti OR intra-individual:ab,ti OR interparticipant:ab,ti OR inter-participant:ab,ti OR intraparticipant:ab,ti OR intraparticipant:ab,ti OR kappa:ab,ti OR kappas:ab,ti OR coefficient of variation:ab,ti OR | 2229098 |

| | | |
|---|---|---|
| | repeatab*:ab,ti OR (replicab*:ab,ti OR repeated:ab,ti AND (measure:ab,ti OR measures:ab,ti OR findings:ab,ti OR result:ab,ti OR results:ab,ti OR test:ab,ti OR tests:ab,ti)) OR generaliza*:ab,ti OR generalisa*:ab,ti OR concordance:ab,ti OR (intraclass:ab,ti AND correlation*:ab,ti) OR discriminative:ab,ti OR known group:ab,ti OR factor analysis:ab,ti OR factor analyses:ab,ti OR factor structure:ab,ti OR factor structures:ab,ti OR dimensionality:ab,ti OR subscale*:ab,ti OR multitrait scaling analysis:ab,ti OR multitrait scaling analyses:ab,ti OR item discriminant:ab,ti OR interscale correlation:ab,ti OR interscale correlations:ab,ti OR (error:ab,ti OR errors:ab,ti AND (measure*:ab,ti OR correlat*:ab,ti OR evaluat*:ab,ti OR accuracy:ab,ti OR accurate:ab,ti OR precision:ab,ti OR mean:ab,ti)) OR individual variability:ab,ti OR interval variability:ab,ti OR rate variability:ab,ti OR variability analysis:ab,ti OR (uncertainty:ab,ti AND (measurement:ab,ti OR measuring:ab,ti)) OR standard error of measurement:ab,ti OR sensitiv*:ab,ti OR responsive*:ab,ti OR (limit:ab,ti AND detection:ab,ti) OR minimal detectable concentration:ab,ti OR interpretab*:ab,ti OR (small*:ab,ti AND (real:ab,ti OR detectable:ab,ti) AND (change:ab,ti OR difference:ab,ti)) OR meaningful change:ab,ti OR minimal important change:ab,ti OR minimal important difference:ab,ti OR minimally important change:ab,ti OR minimally important difference:ab,ti OR minimal detectable change:ab,ti OR minimal detectable difference:ab,ti OR minimally detectable change:ab,ti OR minimally detectable difference:ab,ti OR minimal real change:ab,ti OR minimal real difference:ab,ti OR minimally real change:ab,ti OR minimally real difference:ab,ti OR ceiling effect:ab,ti OR floor effect:ab,ti OR item response model:ab,ti OR irt:ab,ti OR rasch:ab,ti OR differential item functioning:ab,ti OR dif:ab,ti OR computer adaptive testing:ab,ti OR item bank:ab,ti OR cross-cultural equivalence:ab,ti | |
| #6 | 1 and 2 and 3 and 4 and 5 | 1870 |

**Search strategy for CINAHL (EBSCO)**

Search time: 2022-3-13 17:10 (UTC+8)

| Search | Query | Items found |
|---|---|---|
| #1 | TI cancer OR AB cancer OR MH neoplasms | 478227 |
| #2 | AB patient? OR TI patient? OR AB survivor? OR TI survivor? OR MH patients OR MH "cancer survivors" OR MH survivors | 2004224 |
| #3 | AB Cost OR TI Cost OR AB bill? OR TI bill? OR AB expense OR TI expense OR AB "productivity loss" OR TI "productivity loss" OR AB out-of-pocket OR TI out-of-pocket OR AB "economic burden" OR TI "economic burden" OR AB "financial toxicity" OR TI "financial toxicity" OR AB "financial hardship" OR TI "financial hardship" OR AB "financial burden" OR TI "financial burden" OR AB "financial effect" OR TI "financial effect" OR AB "financial stress" OR TI "financial stress" OR AB "economic burden" OR TI "economic burden" OR AB "economic hardship" OR TI "economic hardship" OR AB co-payment OR TI co-payment OR MH "cost of illness" | 211023 |
| #4 | AB Scale? OR TI Scale? OR AB "patient reported outcome measur*" OR TI "patient reported outcome measur*" OR AB PROM? OR TI PROM? OR AB measure* OR TI measure* OR AB "Patient Reported Outcome Measures*" OR MH "Patient Reported Outcome Measures*" OR MH "Surveys and Questionnaires" | 942889 |
| #5 | (MH "Psychometrics") or ( TI psychometr* or AB psychometr* ) or ( TI clinimetr* or AB | 702880 |

| | | |
|---|---|---|
| | clinimetr* ) or ( TI clinometr* OR AB clinometr* ) or (MH "Outcome Assessment") or ( TI outcome assessment or AB outcome assessment ) or ( TI outcome measure* or AB outcome measure* ) or (MH "Health Status Indicators") or (MH "Reproducibility of Results") or (MH"Discriminant Analysis") or ( ( TI reproducib* or AB reproducib* ) or ( TI reliab* or AB reliab* ) or ( TI unreliab* or AB unreliab* ) ) or ( ( TI valid* or AB valid* ) or ( TI coefficient or AB coefficient ) or ( TI homogeneity or AB homogeneity ) ) or ( TI homogeneous or AB homogeneous ) or ( TI "coefficient of variation" or AB "coefficient of variation" ) or ( TI "internal consistency" or AB "internal consistency" ) or (MH "Internal Consistency+") or (MH "Reliability+") or (MH "Measurement Error+") or (MH "Content Validity+") or "hypothesis testing" or "structural validity" or "cross-cultural validity" or (MH "Criterion-Related Validity+") or "responsiveness" or "interpretability" or ( TI reliab* or AB reliab* ) and ( (TI test or AB test) OR (TI retest or AB retest) ) or ( TI stability or AB stability ) or ( TI interrater or AB interrater ) or ( TI inter-rater or AB inter-rater ) or ( TI intrarater or AB intrarater ) or ( TI intra-rater or AB intrarater) or ( TI intertester or AB intertester) or (TI inter-tester or AB inter-tester) or ( TI intratester or AB intratester) or ( TI intra-tester or AB intra-tester) or ( TI interobserver or AB interobserver) or (TI inter-observer or AB inter-observer ) or ( TI intraobserver or AB intraobserver) or ( TI intra-observer or AB intra-observer) or ( TI intertechnician or AB intertechnician) or (TI inter-technician or AB inter-technician) or ( TI intratechnician or AB intratechnician ) or ( TI intra-technician or AB intra-technician ) or ( TI interexaminer or AB interexaminer ) or (TI inter-examiner or AB inter-examiner) or (TI intraexaminer or AB intraexaminer ) OR (TI intra-examiner or AB intra-examiner ) or (TI intra-examiner or AB intraexaminer) or (TI interassay or AB interassay ) or ( TI inter-assay or AB inter-assay ) or ( TI intraassay or AB intraassay) or ( TI intra-assay or AB intra-assay ) or (TI interindividual or AB interindividual) or (TI inter-individual or AB inter-individual) OR (TI intraindividual or AB intraindividual) or (TI intra-individual or AB intra-individual) or (TI interparticipant or AB interparticipant) or (TI inter-participant or AB inter-participant ) or (TI intraparticipant or AB intraparticipant) or (TI intra-participant or AB intra-participant ) or (TI kappa or AB kappa) or (TI kappa's or AB kappa's ) or (TI kappas or AB kappas) or (TI repeatab* or AB repeatab*) or ( TI responsive* or AB responsive* ) or ( TI interpretab* or AB interpretab* ) | |
| #6 | 1 and 2 and 3 and 4 and 5 | 879 |

**Search strategy for PsycINFO (EBSCO)**

Search time: 2022-3-13 17:32 (UTC+8)

| **Search** | **Query** | **Items found** |
|---|---|---|
| #1 | TI cancer OR AB cancer OR MH neoplasms | 1240 |
| #2 | AB patient? OR TI patient? OR AB survivor? OR TI survivor? OR MH patients OR MH "cancer survivors" OR MH survivors | 17860 |
| #3 | AB Cost OR TI Cost OR AB bill? OR TI bill? OR AB expense OR TI expense OR AB "productivity loss" OR TI "productivity loss" OR AB out-of-pocket OR TI out-of-pocket OR AB "economic burden" OR TI "economic burden" OR AB "financial toxicity" OR TI "financial toxicity" OR AB "financial hardship" OR TI "financial hardship" OR AB "financial burden" OR TI "financial burden" OR AB "financial effect" OR TI "financial effect" OR AB "financial | 4217 |

| | | |
|---|---|---|
| | stress" OR TI "financial stress" OR AB "economic burden" OR TI "economic burden" OR AB "economic hardship" OR TI "economic hardship" OR AB co-payment OR TI co-payment OR MH "cost of illness" | |
| #4 | AB Scale? OR TI Scale? OR AB "patient reported outcome measur*" OR TI "patient reported outcome measur*" OR AB PROM? OR TI PROM? OR AB measure* OR TI measure* OR AB "Patient Reported Outcome Measures*" OR MH "Patient Reported Outcome Measures*" OR MH "Surveys and Questionnaires" | 50879 |
| #6 | 1 and 2 and 3 and 4 | 45 |

**Search strategy for Web of Science**

Search time: 2022-3-13 18:17 (UTC+8)

| Search | Query | Items found |
|---|---|---|
| #1 | TI=cancer OR TS= (cancer OR neoplasms) | 6889664 |
| #2 | TI=(patient? OR survivor?) OR TS=(patients OR "cancer survivors" OR survivors) | 12513517 |
| #3 | TI=(Cost OR bill? OR expense OR "productivity loss" OR out-of-pocket OR "economic burden" OR "financial toxicity" OR "financial hardship" OR "financial burden" OR "financial effect" OR "financial stress" OR "economic burden" OR "economic hardship") OR TS="cost of illness" | 463395 |
| #4 | TI=(Scale? OR "patient reported outcome measur*" OR PROM? OR measure*) OR TS=("Patient Reported Outcome Measures*" OR "Surveys and Questionnaires") | 2955804 |
| #6 | #4 AND #3 AND #2 AND #1 | 1002 |

**Search strategy for ProQuest Dissertations and Theses**

Search time: 2022-3-13 18:41 (UTC+8)

| Search | Query | Items found |
|---|---|---|
| #1 | ti(cancer OR neoplasms) OR su(cancer OR neoplasms) | 48268 |
| #2 | ti(patient? OR survivor? OR "cancer survivors") OR su(patient? OR survivor? OR cancer survivors) | 51934 |
| #3 | ti(Cost OR bill? OR expense OR "productivity loss" OR out-of-pocket OR "economic burden" OR "financial toxicity" OR "financial hardship" OR "financial burden" OR "financial effect" OR "financial stress") OR su(Cost OR bill? OR expense OR "productivity loss" OR out-of-pocket OR "economic burden" OR "financial toxicity" OR "financial hardship" OR "financial burden" OR "financial effect" OR "financial stress") OR ti("economic burden" OR "economic hardship" OR "cost of illness") OR su("economic burden" OR "economic hardship" OR "cost of illness") | 45357 |
| #4 | All : Scale? OR "patient reported outcome measur*" OR PROM? OR measure* | 2854129 |
| #6 | #4 AND #3 AND #2 AND #1 | 90 |

**Search strategy for Cochrane Library (Wiley)**

Search time: 2022-3-13 19:05 (UTC+8)

| Search | Query | Items found |
|---|---|---|

| | | |
|---|---|---|
| #1 | (Cancer):ti,ab,kw | 177850 |
| #2 | MeSH descriptor: [Neoplasms] explode all trees | 86823 |
| #3 | (patient? OR survivor? OR "cancer survivors"):ti,ab,kw | 1070252 |
| #4 | MeSH descriptor: [Patients] explode all trees | 2946 |
| #5 | MeSH descriptor: [Cancer Survivors] explode all trees | 476 |
| #6 | MeSH descriptor: [Survivors] explode all trees | 1760 |
| #7 | (Cost OR bill? OR expense OR "productivity loss" OR out-of-pocket OR "economic burden" OR "financial toxicity" OR "financial hardship" OR "financial burden" OR "financial effect" OR "financial stress" OR "economic burden" OR "economic hardship" OR "cost of illness"):ti,ab,kw | 67224 |
| #8 | MeSH descriptor: [Cost of Illness] explode all trees | 864 |
| #9 | (Scale? OR “patient reported outcome measur*” OR PROM? OR measure*):ti,ab,kw | 602173 |
| #10 | MeSH descriptor: [Patient Reported Outcome Measures] explode all trees | 929 |
| #11 | MeSH descriptor: [Surveys and Questionnaires] explode all trees | 58363 |
| #12 | (“Validation Studies” OR “Comparative Study”):pt | 169332 |
| #13 | MeSH descriptor: [Psychometrics] explode all trees | 2951 |
| #14 | (psychometr*):ti,ab,kw | 6755 |
| #15 | #1 OR #2 | 208929 |
| #16 | #3 OR #4 OR #5 OR #6 | 1070725 |
| #17 | #7 OR #8 OR #9 OR #10 OR #11 | 666816 |
| #18 | #12 OR #13 OR #14 | 175091 |
| #19 | #15 AND #16 AND #17 AND #18 | 4453 |

**Appendix II COSMIN Checklists**

**eTable 1** COSMIN risk of bias checklist

| PROM Development | Results |
|---|---|
| 1. Is a clear description provided of the construct to be measured? | |
| 2. Is the origin of the construct clear: was a theory, conceptual framework or disease model used or clear rationale provided to define the construct to be measured? | |
| 3. Is a clear description provided of the target population for which the PROM was developed? | |
| 4. Is a clear description provided of the context of use | |
| 5. Was the PROM development study performed in a sample representing the target population for which the PROM was developed? | |
| 6. Was an appropriate qualitative data collection method used to identify relevant items for a new PROM? | |
| 7. Were skilled group moderators/interviewers used? | |
| 8. Were the group meetings or interviews based on an appropriate topic or interview guide? | |
| 9. Were the group meetings or interviews recorded and transcribed verbatim? | |
| 10. Was an appropriate approach used to analyse the data? | |
| 11. Was at least part of the data coded independently? | |
| 12. Was data collection continued until saturation was reached? | |
| 13. For quantitative studies (surveys): was the sample size appropriate? | |
| 14. Was a cognitive interview study or other pilot test conducted? | |
| 15. Was the cognitive interview study or other pilot test performed in a sample representing the target population? | |
| 16. Were patients asked about the comprehensibility of the PROM? | |
| 17. Were all items tested in their final form? | |
| 18. Was an appropriate qualitative method used to assess the comprehensibility of the PROM instructions, items, response options, and recall period? | |
| 19. Was each item tested in an appropriate number of patients? | |
| 20. Were skilled interviewers used? | |
| 21. Were the interviews based on an appropriate interview guide? | |
| 22. Were the interviews recorded and transcribed verbatim? | |
| 23. Was an appropriate approach used to analyse the data? | |
| 24. Were at least two researchers involved in the analysis? | |
| 25. Were problems regarding the comprehensibility of the PROM instructions, items, response options, and recall period appropriately addressed by adapting the PROM? | |
| 26. Were patients asked about the comprehensiveness of the PROM? | |
| 27. Was the final set of items tested? | |
| 28. Was an appropriate method used for assessing the comprehensiveness of the PROM? | |
| 29. Was each item tested in an appropriate number of patients? | |
| 30. Were skilled interviewers used? | |
| 31. Were the interviews based on an appropriate interview guide? | |
| 32. Were the interviews recorded and transcribed verbatim? | |

| | |
|---|---|
| 33. Was an appropriate approach used to analyse the data? | |
| 34. Were at least two researchers involved in the analysis? | |
| 35. Were problems regarding the comprehensiveness of the PROM appropriately addressed by adapting the PROM? | |
| Content validity | |
| 1. Was an appropriate method used to ask patients whether each item is relevant for their experience with the condition? | |
| 2. Was each item tested in an appropriate number of patients? | |
| 3. Were skilled group moderators/interviewers used? | |
| 4. Were the group meetings or interviews based on an appropriate topic or interview guide? | |
| 5. Were the group meetings or interviews recorded and transcribed verbatim? | |
| 6. Was an appropriate approach used to analyse the data? | |
| 7. Were at least two researchers involved in the analysis? | |
| 8. Was an appropriate method used for assessing the comprehensiveness of the PROM? | |
| 9. Was each item tested in an appropriate number of patients? | |
| 10. Were skilled group moderators/interviewers used? | |
| 11. Were the group meetings or interviews based on an appropriate topic or interview guide? | |
| 12. Were the group meetings or interviews recorded and transcribed verbatim? | |
| 13. Was an appropriate approach used to analyse the data? | |
| 14. Were at least two researchers involved in the analysis? | |
| 15. Was an appropriate qualitative method used for assessing the comprehensibility of the PROM instructions, items, response options, and recall period? | |
| 16. Was each item tested in an appropriate number of patients? | |
| 17. Were skilled group moderators/interviewers used? | |
| 18. Were the group meetings or interviews based on an appropriate topic or interview guide? | |
| 19. Were the group meetings or interviews recorded and transcribed verbatim? | |
| 20. Was an appropriate approach used to analyse the data? | |
| 21. Were at least two researchers involved in the analysis? | |
| 22. Was an appropriate method used to ask professionals whether each item is relevant for the construct of interest? | |
| 23. Were professionals from all relevant disciplines included? | |
| 24. Was each item tested in an appropriate number of professionals? | |
| 25. Was an appropriate approach used to analyse the data? | |
| 26. Were at least two researchers involved in the analysis? | |
| 27. Was an appropriate method used for assessing the comprehensiveness of the PROM? | |
| 28. Were professionals from all relevant disciplines included? | |
| 29. Was each item tested in an appropriate number of professionals? | |
| 30. Was an appropriate approach used to analyse the data? | |
| 31. Were at least two researchers involved in the analysis? | |
| Structural validity | |

| | |
|---|---|
| 1. For CTT: Was exploratory or confirmatory factor analysis performed? | |
| 2. For IRT/Rasch: does the chosen model fit to the research question? | |
| 3. Was the sample size included in the analysis adequate? | |
| 4. Were there any other important flaws in the design or statistical methods of the study? | |
| Internal consistency | |
| 1. Was an internal consistency statistic calculated for each unidimensional scale or subscale separately? | |
| 2. For continuous scores: Was Cronbach's alpha or omega calculated? | |
| 3. For dichotomous scores: Was Cronbach's alpha or KR-20 calculated? | |
| 4. For IRT-based scores: Was standard error of the theta (SE (θ)) or reliability coefficient of estimated latent trait value (index of (subject or item) separation) calculated? | |
| 5. Were there any other important flaws in the design or statistical methods of the study? | |
| Cross-cultural validity | |
| 1. Were the samples similar for relevant characteristics except for the group variable? | |
| 2. Was an appropriate approach used to analyse the data? | |
| 3. Was the sample size included in the analysis adequate? | |
| 4. Were there any other important flaws in the design or statistical methods of the study? | |
| Reliability | |
| 1. Were patients stable in the interim period on the construct to be measured? | |
| 2. Was the time interval appropriate? | |
| 3. Were the test conditions similar for the measurements? e.g. type of administration, environment, instructions | |
| 4. For continuous scores: Was an intraclass correlation coefficient (ICC) calculated? | |
| 5. For dichotomous/nominal/ordinal scores: Was kappa calculated? | |
| 6. For ordinal scores: Was a weighted kappa calculated? | |
| 7. For ordinal scores: Was the weighting scheme described? e.g. linear, quadratic | |
| 8. Were there any other important flaws in the design or statistical methods of the study? | |
| Measurement error | |
| 1. Were patients stable in the interim period on the construct to be measured? | |
| 2. Was the time interval appropriate? | |
| 3. Were the test conditions similar for the measurements? (e.g. type of administration, environment, instructions) | |
| 4. For continuous scores: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated? | |
| 5. For dichotomous/nominal/ordinal scores: Was the percentage (positive and negative) agreement calculated? | |
| 6. Were there any other important flaws in the design or statistical methods of the study? | |
| Criterion validity | |
| 1. For continuous scores: Were correlations, or the area under the receiver operating curve calculated? | |
| 2. For dichotomous scores: Were sensitivity and specificity determined? | |
| 3. Were there any other important flaws in the design or statistical methods of the study? | |
| Hypotheses testing for construct validity | |

| | |
|---|---|
| 1. Is it clear what the comparator instrument(s) measure(s)? | |
| 2. Were the measurement properties of the comparator instrument(s) sufficient? | |
| 3. Was the statistical method appropriate for the hypotheses to be tested? | |
| 4. Were there any other important flaws in the design or statistical methods of the study? | |
| 5. Was an adequate description provided of important characteristics of the subgroups? | |
| 6. Was the statistical method appropriate for the hypotheses to be tested? | |
| 7. Were there any other important flaws in the design or statistical methods of the study? | |
| Responsiveness | |
| 1. For continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated? | |
| 2. For dichotomous scales: Were sensitivity and specificity (changed versus not changed) determined? | |
| 3. Were there any other important flaws in the design or statistical methods of the study? | |
| 4. Is it clear what the comparator instrument(s) measure(s)? | |
| 5. Were the measurement properties of the comparator instrument(s) sufficient? | |
| 6. Was the statistical method appropriate for the hypotheses to be tested? | |
| 7. Were there any other important flaws in the design or statistical methods of the study? | |
| 8. Was an adequate description provided of important characteristics of the subgroups? | |
| 9. Was the statistical method appropriate for the hypotheses to be tested? | |
| 10. Were there any other important flaws in the design or statistical methods of the study? | |
| 11. Was an adequate description provided of the intervention given? | |
| 12. Was the statistical method appropriate for the hypotheses to be tested? | |
| 13. Were there any other important flaws in the design or statistical methods of the study? | |

**eTable 2** Criteria for good measurement properties

| Measurement property | Rating | Criteria |
|---|---|---|
| Structural validity | + | **CTT:**<br>CFA: CFI or TLI or comparable measure >0.95 OR RMSEA <0.06 OR SRMR <0.082<br><br>**IRT/Rasch:**<br>No violation of unidimensionality3: CFI or TLI or comparable<br>measure >0.95 OR RMSEA <0.06 OR SRMR <0.08<br>AND<br>no violation of local independence: residual correlations among the items after controlling for the dominant factor <0.20 OR Q3's < 0.37<br>AND<br>no violation of monotonicity: adequate looking graphs OR item scalability >0.30<br>AND<br>adequate model fit:<br>IRT: $\chi 2$ >0.01<br>Rasch: infit and outfit mean squares ≥ 0.5 and ≤ 1.5 OR Z standardized values >-2 and <2 |
| | ? | CTT: Not all information for '+' reported<br>IRT/Rasch: Model fit not reported |
| | - | Criteria for '+' not met |
| Internal consistency | + | At least low evidence for sufficient structural validity AND Cronbach's alpha(s) ≥ 0.70 for each unidimensional scale or subscale |
| | ? | Criteria for "At least low evidence4 for sufficient structural validity" not met |
| | - | At least low evidence for sufficient structural validity AND Cronbach's alpha(s) < 0.70 for each unidimensional scale or subscale |
| Reliability | + | ICC or weighted Kappa ≥ 0.70 |
| | ? | ICC or weighted Kappa not reported |
| | - | ICC or weighted Kappa < 0.70 |
| Measurement error | + | SDC or LoA < MIC |
| | ? | MIC not defined |
| | - | SDC or LoA > MIC |
| Hypotheses testing for construct validity | + | The result is in accordance with the hypothesis |
| | ? | No hypothesis defined (by the review team) |
| | - | The result is not in accordance with the hypothesis |
| Cross-cultural | + | No important differences found between group factors (such |

| | | |
|---|---|---|
| validity\measurement invariance | | as age, gender, language) in multiple group factor analysis OR no important DIF for group factors (McFadden's R2 < 0.02) |
| | ? | No multiple group factor analysis OR DIF analysis performed |
| | - | Important differences between group factors OR DIF was found |
| Criterion validity | + | Correlation with gold standard ≥ 0.70 OR AUC ≥ 0.70 |
| | ? | Not all information for '+' reported |
| | - | Correlation with gold standard < 0.70 OR AUC < 0.70 |
| Responsiveness | + | The result is in accordance with the hypothesis7 OR AUC ≥ 0.70 |
| | ? | No hypothesis defined (by the review team) |
| | - | The result is not in accordance with the hypothesis7 OR AUC < 0.70 |

AUC: area under the curve; CFA: confirmatory factor analysis; CFI: comparative fit index; CTT: classical test theory; DIF: differential item functioning; ICC: intraclass correlation coefficient; IRT: item response theory; LoA: limits of agreement; MIC: minimal important change; RMSEA: Root Mean Square Error of Approximation; SEM: Standard Error of Measurement; SDC: smallest detectable change; SRMR: Standardized Root Mean Residuals; TLI: Tucker-Lewis index; "+": sufficient; "-": insufficient; "?": indeterminate.

**eTable 3** Modified GRADE approach for assessing certainty of evidence*

| Domain | Grade | Reason |
|---|---|---|
| Risk of bias | -0 level: No | There are multiple studies of at least adequate quality, or there is one study of very good quality available |
| | -1 level: Serious | There are multiple studies of doubtful quality available, or there is only one study of adequate quality |
| | -2 level: Very serious | There are multiple studies of inadequate quality, or there is only one study of doubtful quality available |
| | -3 level: Extremely serious | There is only one study of inadequate quality available |
| Inconsistency | -0 level: No | There is no inconsistency among pooled studies or there is only one study in subgroups |
| | -1 level: Serious | There are severe inconsistencies among pooled studies |
| | -2 level: Very serious | There are very severe inconsistencies among pooled studies. |
| Imprecision | -0 level: No | Total sample size>50-100 |
| | -1 level: Serious | Total sample size=50-100 |
| | -2 level: Very serious | Total sample size <50 |
| Indirectness | -0 level: No | There is no indirectness between results and conclusion |
| | -1 level: Serious | There is severe indirectness between results and conclusion |
| | -2 level: Very serious | There is very severe indirectness between results and conclusion |

*The starting point of quality level is high evidence. The quality of evidence is subsequently downgraded to moderate, low, or very low evidence.