BMJ Open

Psychometric properties of disease-specific health-related quality of life (HRQoL) instruments for food allergy and food intolerance: protocol for a COSMIN-based systematic review

Qirong Chen ,[1,2] Zeen Li,[1] Aimee Castro ,[3] Siyuan Tang ,[1,2] Jiarui Chen,[1,2] Chongmei Huang,[1] Jinnan Xiao ,[1] Haiyan Liu,[4] Jinfeng Ding [1,2]

To cite: Chen Q, Li Z, Castro A, *et al*. Psychometric properties of disease-specific health-related quality of life (HRQoL) instruments for food allergy and food intolerance: protocol for a COSMIN-based systematic review. *BMJ Open* 2022;**12**:e053534. doi:10.1136/bmjopen-2021-053534

► 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-2021-053534).

Received 15 May 2021
Accepted 19 November 2021

For numbered affiliations see end of article.

Correspondence to
Dr Jinfeng Ding;
jinfeng.ding@csu.edu.cn

ABSTRACT

Introduction Food allergies and food intolerances can bring burdens on patients and their caregivers and reduce health-related quality of life (HRQoL). An increasing number of disease-specific HRQoL instruments for food allergies and food intolerances has been developed, and some of them have been adapted for different cultures and languages. This report describes a protocol for a systematic review of the psychometric properties of these instruments. The aims of this systematic review are to: (1) formulate recommendations for the usage of existing validated disease-specific HRQoL instruments for patients with food allergies and/or food intolerances and their caregivers; and (2) identify knowledge gaps to inform future research relating to these instruments.

Methods and analysis This protocol adheres to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocol (PRISMA-P) 2015 checklist. The future review will follow the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) guideline for systematic reviews of patient-reported outcome measures (PROMs) and PRISMA 2020 statement guideline. Six databases (PubMed, EMBASE, Web of Science, Scopus, CINAHL and ProQuest -Health & Medical Collection) will be searched to retrieve studies focusing on the development and psychometric properties of disease-specific HRQoL instruments for patients with food allergies and/or food intolerances and their caregivers between 1 December 2021 and 31 December 2021. Two researchers will be responsible for literature screening, data extraction and literature evaluation, independently. Disagreements will be addressed by discussion or the involvement of a third researcher. The methodological quality of the included studies and the quality of the identified instruments will be assessed based on the COSMIN guideline for systematic reviews of PROMs.

Ethics and dissemination Ethical approval is not applicable for this study. We will disseminate the findings through publication in peer-reviewed journals and/or academic conferences.

PROSPERO registration number CRD42021252203.

Strengths and limitations of this study

- ► To our knowledge, this is the first Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN)-based systematic review of disease-specific health-related quality of life (HRQoL) instruments for food allergies and food intolerances.
- ► All up-to-date specific guidelines (Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Protocol 2015 checklist, updated COSMIN guideline for systematic reviews of patient-reported outcome measures (PROMs) and PRISMA 2020 checklist) will be used to guide the implementation and report of the protocol and systematic review.
- ► The COSMIN guideline for systematic reviews of PROMs will be used to evaluate the methodological quality of included studies on measurement properties of the instruments and the quality of included instruments.
- ► The systematic review may fail to include relevant literature that were not included in the searched databases.

INTRODUCTION

Adverse reactions to food encompass food allergies and food intolerances.[1] Food allergies and food intolerances are common, especially in children, with 6.25%–28.0% of children having had this experience.[1] Food allergies and food intolerances have become severe public health problems worldwide.[2] A food allergy is an abnormal immunological reaction to specific food(s) that results in the development of symptoms.[3] Food allergies include three types of immunological reactions: IgE-mediated, non-IgE-mediated and Mixed IgE- and non-IgE-mediated.[1] Food intolerances are non-immune mediated adverse reactions to food; lactose intolerance

is one such example.[1] Food allergies and food intolerances impact multiple organs and systems, and are associated with a range of symptoms (eg, urticaria, eczema, colic, vomiting, reflux, diarrhoea or constipation, blood in stool and growth retardation).[1 3] Meanwhile, food allergies may lead to life-threatening anaphylaxis.[4]

Adverse food intake reactions bring great challenges to healthcare, education, food and catering industries in many countries. Critically, the physiological, psychological, social and financial burdens relating to food allergies and food intolerances also undermine health-related quality of life (HRQoL) of patients and their family caregivers.[5] Instruments measuring HRQoL can help quantify the impacts of food allergies and food intolerances, and may support better prevention and management of this problem. For example, healthcare professionals could use these instruments to assess the quality of life of patients and their caregivers to provide specific healthcare services and suggestions. Furthermore, researchers can use these instruments in studies to evaluate the quality of life of patients and their caregivers.

There are two types of HRQoL instruments: generic and disease-specific.[5] Disease-specific HRQoL instruments are more likely to have a higher level of sensitivity compared with generic HRQoL instruments.[5 6] A number of disease-specific HRQoL instruments for food allergies has been developed. Some of them have been validated and adapted into multiple versions for different cultures and languages.[6 7] Since 2014, there have also been some studies reporting on the overall development of instruments for food intolerances.[8–10] Two literature reviews relating to HRQoL instruments for food allergies were published in 2009 and 2014, respectively.[6 7] The 2009 review summarised and described generic and disease-specific instruments for food allergies in children and adults.[6] However, this review was limited by its use of a narrative review approach, rather than a systematic review following corresponding guidelines, which can lead to the omission of some important literature. The 2014 review systematically summarised and evaluated all disease-specific HRQoL instruments for IgE-mediated food allergies.[7] However, this systematic review failed to include other types of food allergies (non-IgE-mediated, and Mixed IgE- and non-IgE-mediated), as well as food intolerances. These omiited types of adverse reactions to food also have significant influences on quality of life of patients and caregivers, and as such, should be included in a broader systematic review of the literature. Furthermore, the 2014 review did not follow specific guidelines for evaluating the methodological quality of the included studies and the quality of the included instruments. In 2018, the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) guideline was developed for systematic reviews of Patient-Reported Outcome Measures (PROMs).[11–14] This guideline could improve the quality of the systematic review of PROMs, offering researchers a critical and comprehensive evaluation of the available instruments.

Therefore, the overall aim of this systematic review is to critically describe, appraise, and summarise the existing disease-specific HRQoL instruments for patients with food allergies and/or food intolerances and their family caregivers, based on the COSMIN guideline for systematic reviews of PROMs.[13] The specific objectives of the proposed systematic review are to: (1) identify and describe all existing validated disease-specific HRQoL instruments for patients with food allergies and/or food intolerances and their caregivers; (2) evaluate the methodological quality of studies on measurement properties of the instruments and (3) assess and compare the psychometric properties and other key characteristics of these instruments.

This systematic review will answer the following questions: (1) What are existing disease-specific HRQoL instruments for patients with food allergies and/or food intolerances and their caregivers? (2) What are the characteristics of these instruments? (3) What is the methodological quality of studies on measurement properties of these instruments? (4) What are the measurement properties, interpretability and feasibility of these instruments? (5) What are the similarities and differences among these instruments? and (6) What are the knowledge and research gaps in this area?

METHODS

This is a protocol for a systematic review following Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols (PRISMA-P) 2015 checklist.[15] Based on COSMIN guideline for systematic reviews of PROMs, we adapted the 'Objectives' section of the PRISMA-P 2015 checklist by replacing 'Participants, Interventions, Comparators and Outcomes' with 'Construct, Population(s), Type of Instrument(s) and Measurement properties'.[11] We registred the protocol in the International Prospective Register of Systematic Reviews (PROSPERO), and the title of the protocol has been registered on the Joanna Briggs Institute website. We will conduct this systematic review based on the COSMIN guideline for systematic reviews of PROMs[13] and report it following the updated PRISMA 2020 checklist.[16]

Inclusion and exclusion criteria of studies

Inclusion criteria

Studies will be included if they: (1) report disease-specific HRQoL instrument(s) designated for patients with food allergies and/or food intolerances, and/or their caregivers; (2) describe the processes of development and/or evaluation of one or more measurement properties for eligible instrument(s) and (3) have full-text availability. The authors of the articles will be approached if a full-text version is not available online. However, if the authors' contact information is not available or the authors do not respond to the inquiry, these studies will be excluded but their information will be recorded in online supplemental file of the formal systematic review.

Exclusion criteria

Studies will be excluded if they: (1) are not primary studies (eg, discussion papers, letters and editorials) or case studies or (2) are reports that used the instruments only for outcome measurement.

Search strategy

Between 1 December 2021 and 31 December 2021, we will search PubMed, EMBASE, Web of Science, Scopus, CINAHL and ProQuest (Health & Medical Collection) using comprehensive and sensitive search strategies that combine Medical Subject Heading and free text words. All databases will be searched from the date of inception to the date of searching. The major search concepts will be quality of life (Construct), food allergies/food intolerances (Population), PROMs (Type of instrument(s)) and measurement properties. Three comprehensive and sensitive keyword search strategies developed by other researchers for reviews of the concepts will be used in this literature search. They are: (1) the search filter of 'quality of life' for medical and health bibliographic databases developed by Dutch medical information specialists[17]; (2) the search filter for finding PROMs developed by the University of Oxford[18]; (3) the sensitive PubMed search filter for measurement properties developed by Terwee *et al*, and corresponding search filters adapted for other databases. These literature search filters will improve the comprehensiveness, effectiveness and quality of the literature search in this study.[19] Furthermore, a health science librarian had been consulted for developing the search strategies. Online supplemental table S1 and online supplemental tables S9–S13 show the search strategies we developed for the databases searched. The search will not be limited to a specific language; that is, we will include eligible publications in any language, and a translation service will be used if needed. Database searches will be carried out again to provide a final update of the searches after the systematic review is accepted by a journal. The systematic review will be updated if new eligible studies are identified.

Study screening

Endnote and Covidence will be used to manage the references screening. First, we will use EndNote to recognise and remove duplicates, and then conduct manual screening.[20] Following this initial screening, titles, abstracts and full-text articles will be reviewed and screened independently by two researchers with the support of Covidence. Disagreements between the two researchers will be addressed through discussion. Consultation of a third researcher will be adopted where necessary. Reference lists from all eligible papers will also be screened using the aforementioned inclusion and exclusion criteria. The processes of study screening are shown in figure 1.

Data extraction

Two researchers will conduct the data extraction independently. A third researcher will review the extracted data and address the discrepancies between the two researchers if identified. We will extract data on (1) basic characteristics of the included instruments (online supplemental table S2, including: the name of the instrument, developer(s)/year developed, construct(s), targeted population, mode of administration, recall period, (sub) scale(s)/(number of items), response options, range of scores/scoring, original language and available translations); (2) characteristics of the included study populations (online supplemental table S3, including sample size, age, gender, disease, disease duration and severity, setting, country, language); (3) results of measurement properties of the included instruments (Result columns in online supplemental table S5, including content validity, structural validity, internal consistency, cross-cultural validity, measurement invariance, reliability, measurement error, criterion validity, construct validity, responsiveness and (4) interpretability (online supplemental table S7) and feasibility (online supplemental table S8) of the included instruments.

Quality appraisal and data synthesis

Two researchers will conduct the quality assessment for included studies and instruments independently. A third researcher will be consulted if consensus could not be reached. The COSMIN guideline will be used to assess each subscale of a multidimensional PROM separately.[21] Therefore, the measurement properties for subscale scores and the entire PROM will be rated separately in this study.

In the first step, COSMIN standards for design requirements and preferred statistical methods will guide the evaluation of the methodological quality of the included studies on the development and measurement properties of the instruments.[11 13 14] The following COSMIN resources will be used in this phase: the COSMIN Risk of Bias checklist for PROMs,[12] the COSMIN methodology for systematic reviews of PROMs—User manual,[11] and the COSMIN methodology for assessing the content validity of PROMs—User manual.[22] In this step, an Excel sheet file named 'Scoring form COSMIN boxes' will be used to manage the evaluation records (refer to online supplemental additional file 3; this file is also available at https://www.cosmin.nl/tools/guideline-conducting-systematic-review-outcome-measures/); this file is provided by the COSMIN guideline. The final consensus on the results of the methodological quality will be presented in online supplemental table S4 and Meth qual column in online supplemental table S5.

In the second step, we will evaluate the results associated with measurement properties of identified instruments according to the COSMIN quality criteria for good measurement properties.[21 23] The corresponding results will be reported in the rating columns in online supplemental table S5. However, the rating results of content validity will be separately presented in online supplemental table S5-1 given that criteria and rating systems

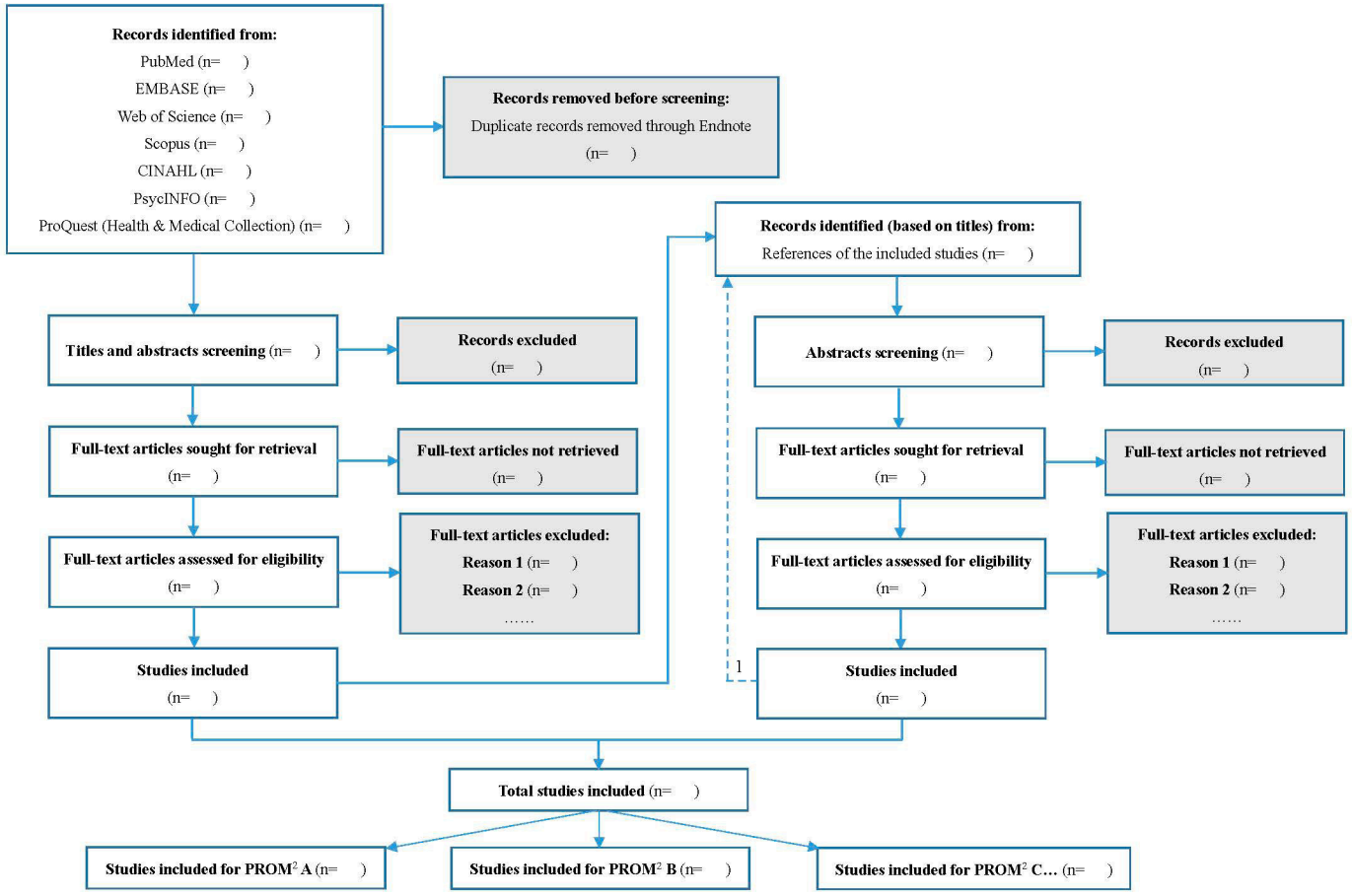

Figure 1 Flow chart of literature selection process. PROM, patient-reported outcome measure.

for evaluation of content validity of PROMs differ from other measurement properties in COSMIN guideline.[23]

The COSMIN guideline also provides a separate table (online supplemental table S6) to synthesise evidence and results associated with measurement properties. In the third step, we will statistically pool or qualitatively summarise the results on measurement properties from different studies providedand show the summarised or pooled results in the column of Summary or pooled results of online supplemental table S6. A meta-analysis approach (weighted means and 95% CIs) will be used when possible. We will evaluate the pooled or summarised results per measurement property for each PROM according to the COSMIN quality criteria for good measurement properties; the corresponding results will be shown in the Overall rating columns in online supplemental table S6.

Finally, we will use the modified GRADE (Grading of Recommendations Assessment, Development and Evaluation) approach developed by the COSMIN guideline for systematic reviews of PROMs to grade the quality of the evidence. The quality of the evidence will be defined as the level of confidence based on the level of trustworthiness of the pooled or summarised result (shown in Quality of evidence columns in online supplemental table S6). The COSMIN guideline classifies the quality of the evidence into four levels: 'high', 'moderate', 'low' or 'very low'.[21]

These findings will enable us to formulate recommendations on the usage of existing disease-specific HRQoL instruments for patients with food allergies and/or food intolerances and their caregivers. Our findings will also identify knowledge gaps in this area and inform future research.

Patient and public involvement

Neither patients nor the public will be involved in this study.

Ethics and dissemination

Ethical approval is not applicable for this study. We will share the findings from the study at national and/or international conferences and in a peer-reviewed journal in the fields of food allergies or food intolerances.

DISCUSSION

To our knowledge, this review will be the first PRISMA and COSMIN guidelines-guided systematic review of disease-specific HRQoL instruments for patients with food allergies and/or food intolerances and their caregivers. This review will identify, describe, evaluate and compare all eligible instruments. The methodological quality of all included studies on the measurement properties of these instruments and the psychometric properties of

all included instruments will be evaluated based on the COSMIN guideline for systematic reviews of PROMs. The findings will facilitate the formulation of the recommendations on the usage of the targeted instruments for clinical practice and research. We will also identify knowledge gaps associated with measurements of HRQoL for patients with food allergies and/or food intolerances and their caregivers. This review has the potential to clearly identify opportunities for further research, and therefore supports future studies on the development and improvement of disease-specific HRQoL instruments for these populations.

Author affiliations
[1]Xiangya School of Nursing, Central South University, Changsha, Hunan, China
[2]Xiangya Center for Evidence-Based Practice & Healthcare Innovation: A Joanna Briggs Institute Affiliated Group, Central South University, Changsha, Hunan, China
[3]Ingram School of Nursing, McGill University, Montreal, Quebec, Canada
[4]Third Xiangya Hospital, Central South University, Changsha, Hunan, China

Twitter Aimee Castro @AimeeRCastro

Contributors Study design: QC, JD and ST; Study supervision: JD, ST and HL; Manuscript writing: QC, ZL, JD and AC; Critical revisions for important intellectual content: AC, JC, CH, JX and HL.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iDs
Qirong Chen http://orcid.org/0000-0003-1908-290X
Aimee Castro http://orcid.org/0000-0002-6461-0866
Siyuan Tang http://orcid.org/0000-0001-9940-5072
Jinnan Xiao http://orcid.org/0000-0003-2403-8495
Jinfeng Ding http://orcid.org/0000-0002-8783-8919

REFERENCES

1 Manuyakorn W, Tanpowpong P. Cow milk protein allergy and other common food allergies and intolerances. *Paediatr Int Child Health* 2019;39:32–40.
2 Hossny E, Ebisawa M, El-Gamal Y, *et al*. Challenges of managing food allergy in the developing world. *World Allergy Organ J* 2019;12:100089.
3 Muraro A, Werfel T, Hoffmann-Sommergruber K, *et al*. EAACI food allergy and anaphylaxis guidelines: diagnosis and management of food allergy. *Allergy* 2014;69:1008–25.
4 Muraro A, Roberts G, Worm M, *et al*. Anaphylaxis: guidelines from the European Academy of allergy and clinical immunology. *Allergy* 2014;69:1026–45.
5 Dunn Galvin A, Hourihane J O'B. Health-related quality of life in food allergy : Impact, correlates, and predictors. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 2016;59:841–8.
6 Flokstra-de Blok BMJ, Dubois AEJ. Quality of life in food allergy: valid scales for children and adults. *Curr Opin Allergy Clin Immunol* 2009;9:214–21.
7 Salvilla SA, Dubois AEJ, Flokstra-de Blok BMJ, *et al*. Disease-Specific health-related quality of life instruments for IgE-mediated food allergy. *Allergy* 2014;69:834–44.
8 DunnGalvin A, Barnett J, Begen FM, *et al*. Development and preliminary validation of the food intolerance quality of life questionnaire (FIQLQ): adult form. *Qual Life Res* 2018;27:1109–16.
9 Hourihane JO, Greenhawt MJ, Galvin AD. A new valid and reliable parent proxy questionnaire to measure the impact of food protein enterocolitis syndrome on children: the Fpies quality of life questionnaire, parent form. *J Allergy Clin Immunol* 2015;135:AB280.
10 Greenhawt MJ, Schultz F, Galvin AD. Caregiver quality of life in food protein enterocolitis syndrome. *J Allergy Clin Immunol* 2015;135:AB280.
11 Mokkink LB, Prinsen CAC, Patrick DL, *et al*. COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) - user manual Netherlands: COSMIN, 2018. Available: https://cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf [Accessed 5 May 2021].
12 Mokkink LB, de Vet HCW, Prinsen CAC, *et al*. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27:1171–9.
13 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.
14 Terwee CB, Prinsen CAC, Chiarotto A, *et al*. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res* 2018;27:1159–70.
15 Shamseer L, Moher D, Clarke M, *et al*. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015;350:g7647.
16 Page MJ, Moher D, Bossuyt PM, *et al*. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ* 2021;372:n160.
17 Vissers T, Vries RD. Quality of life (QOL) search block Amsterdam: Afdeling Biomedische Informatiespecialisten; 2020, 2020. Available: https://blocks.bmi-online.nl/catalog/294 [Accessed 5 May 2021].
18 Mackintosh A, Comabella CCI, Hadi M, *et al*. PROM GROUP CONSTRUCT & INSTRUMENT TYPE FILTERS Oxford: University of Oxford; 2010, 2010. Available: https://cosmin.nl/wp-content/uploads/prom-search-filter-oxford-2010.pdf [Accessed 5 May 2021].
19 Terwee CB, Jansma EP, Riphagen II, *et al*. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res* 2009;18:1115–23.
20 Bramer WM, Giustini D, de Jonge GB, *et al*. De-duplication of database search results for systematic reviews in endnote. *J Med Libr Assoc* 2016;104:240–3.
21 Mokkink LB, Prinsen CA, Patrick DL, *et al*. COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs) – user manual Amsterdam, the Netherlands COSMIN, 2018a. Available: https://cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf [Accessed 27 Apr 2021].
22 Terwee CB, Prinsen CA, Chiarotto A, *et al*. COSMIN methodology for assessing the content validity of PROMs - user manual Netherlands: COSMIN; 2018, 2018. Available: https://cosmin.nl/wp-content/uploads/COSMIN-methodology-for-content-validity-user-manual-v1.pdf [Accessed 5 May 2021].
23 Terwee CB, Prinsen CA, Chiarotto A, *et al*. COSMIN methodology for assessing the content validity of PROMs - User manual Amsterdam, The Netherlands: COSMIN, 2018b. Available: https://cosmin.nl/wp-content/uploads/COSMIN-methodology-for-content-validity-user-manual-v1.pdf [Accessed 27 Apr 2021].

Table S1: Search strategy for PubMed

1	#1	Search: Quality of Life[Mesh] OR quality of life[tiab] OR life qualit*[tiab] OR living qualit*[tiab] OR quality of living[tiab] OR Activities of Daily Living[Mesh] OR activities of daily living[tiab] OR activity of daily living[tiab] OR activities of daily life[tiab] OR activity of daily life[tiab] OR daily living activit*[tiab] OR daily life activit*[tiab] OR adl[tiab] OR chronic limitation of activity[tiab] OR self care*[tiab] OR Health Status[Mesh] OR health status[tiab] OR level of health[tiab] OR health level*[tiab] OR qol[tiab] OR hrql[tiab] OR hrqol[tiab]
2	#2	Search: food hypersensitivity[Mesh] OR food intolerance[Mesh] OR food allerg*[tw] OR food hypersensitivit*[tw] OR food intolerance*[tw] OR food sensitivit*[tw]
3	#3	Search: (HR-PRO[tiab] OR HRPRO[tiab] OR HRQL[tiab] OR HRQoL[tiab] OR QL[tiab] OR QoL[tiab] OR quality of life[tw] OR life quality[tw] OR health index*[tiab] OR health indices[tiab] OR health profile*[tiab] OR health status[tw] OR ((patient[tiab] OR self[tiab] OR child[tiab]OR parent[tiab] OR carer[tiab] OR proxy[tiab]) AND ((report[tiab] OR reported[tiab] OR reporting[tiab]) OR (rated[tiab] OR rating[tiab] OR ratings[tiab]) OR based[tiab] OR (assessed[tiab] OR assessment[tiab] OR assessments[tiab]))) OR ((disability[tiab] OR function[tiab] OR functional[tiab] OR functions[tiab] OR subjective[tiab] OR utility[tiab] OR utilities[tiab] OR wellbeing[tiab] OR well being[tiab]) AND (index[tiab] OR indices[tiab] OR instrument[tiab] OR instruments[tiab] OR measure[tiab] OR measures[tiab] OR questionnaire[tiab] OR questionnaires[tiab] OR profile[tiab] OR profiles[tiab] OR scale[tiab] OR scales[tiab] OR score[tiab] OR scores[tiab] OR status[tiab] OR survey[tiab] OR surveys[tiab])))
4	#4	Search: (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

Reference:
1. Vissers T, Vries RD. Quality of life (QoL) search block Amsterdam: Afdeling Biomedische Informatiespecialisten; 2020 [updated March 23, 2020; cited 2021 May 5]. Available from: https://blocks.bmi-online.nl/catalog/294 accessed May 5 2021.
2. Mackintosh A, Comabella CCI, Hadi M, et al. PROM GROUP CONSTRUCT & INSTRUMENT TYPE FILTERS Oxford: University of Oxford; 2010 [updated Febrary, 2010; cited 2021 May 5]. Available from: https://cosmin.nl/wp-content/uploads/prom-search-filter-oxford-2010.pdf accessed May 5 2021.
3. Terwee CB, Jansma EP, Riphagen II, et al. Development of a Methodological PubMed Search Filter for Finding Studies on Measurement Properties of Measurement Instruments. Quality of Life Research 2009;18(8):1115-23. doi: 10.1007/s11136-009-9528-5

		intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intraassay[tiab] OR intra-assay[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[tiab] OR kappa's[tiab] OR kappas[tiab] OR repeatab*[tw] OR ((replicab*[tw] OR repeated[tw]) AND (measure[tw] OR measures[tw] OR findings[tw] OR result[tw] OR results[tw] OR test[tw] OR tests[tw])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR known group[tiab] OR factor analysis[tiab] OR factor analyses[tiab] OR factor structure[tiab] OR factor structures[tiab] OR dimension*[tiab] OR subscale*[tiab] OR (multitrait[tiab] AND scaling[tiab] AND (analysis[tiab] OR analyses[tiab])) OR item discriminant[tiab] OR interscale correlation*[tiab] OR error[tiab] OR errors[tiab] OR individual variability[tiab] OR interval variability[tiab] OR rate variability[tiab] OR (variability[tiab] AND (analysis[tiab] OR values[tiab])) OR (uncertainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR standard error of measurement[tiab] OR sensitiv*[tiab] OR responsive*[tiab] OR (limit[tiab] AND detection[tiab]) OR minimal detectable concentration[tiab] OR interpretab*[tiab] OR ((minimal[tiab] OR minimally[tiab] OR 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])
5	#5	Search: (addresses[pt] OR biography[pt] OR case reports[pt] OR comment[pt] OR directory[pt] OR editorial[pt] OR festschrift[pt] OR interview[pt] OR lectures[pt] OR legal cases[pt] OR legislation[pt] OR letter[pt] OR news[pt] OR newspaper article[pt] OR patient education handout[pt] OR popular works[pt] OR congresses[pt] OR consensus development conference[pt] OR consensus development conference, nih[pt] OR practice guideline[pt]) NOT (animals[Mesh] NOT humans[Mesh])
6	#6	#1 AND #2 AND #3 AND #4 NOT #5

Note:

#1: The search blocks of quality of life for medical and health bibliographic databases complied by Dutch medical information specialists is accessible from https://blocks.bmi-online.nl/catalog/294

#3: The search filter for finding PROMs developed by the University of Oxford is accessible from https://cosmin.nl/wp-content/uploads/prom-search-filter-oxford-2010.pdf

#4 and #5: The sensitive PubMed search filter for measurement properties developed by Terwee et al., and corresponding translated search filters for other databases are accessible from https://www.cosmin.nl/tools/pubmed-search-filters/?portfolioCats=14

Reference:

1. Vissers T, Vries RD. Quality of life (QoL) search block Amsterdam: Afdeling Biomedische Informatiespecialisten; 2020 [updated March 23, 2020; cited 2021 May 5]. Available from: https://blocks.bmi-online.nl/catalog/294 accessed May 5 2021.
2. Mackintosh A, Comabella CCI, Hadi M, et al. PROM GROUP CONSTRUCT & INSTRUMENT TYPE FILTERS Oxford: University of Oxford; 2010 [updated Febrary, 2010; cited 2021 May 5]. Available from: https://cosmin.nl/wp-content/uploads/prom-search-filter-oxford-2010.pdf accessed May 5 2021.
3. Terwee CB, Jansma EP, Riphagen II, et al. Development of a Methodological PubMed Search Filter for Finding Studies on Measurement Properties of Measurement Instruments. Quality of Life Research 2009;18(8):1115-23. doi: 10.1007/s11136-009-9528-5

Table S2. Characteristics of the included PROMs[1]

PROM[1]	Developer(s)/ year developed	Construct(s)	Target population	Mode of administration	Recall period	(Sub)scale (s) (number of items)	Response options	Range of scores/scoring	Original language	Available translations
A										
B										
……										

Note: 1. PROM(s) = Patient-reported outcome measure(s). In this study, PROM(s) refers to the disease-specific HRQL instrument(s) for patients with food allergy or/and food intolerance and their caregivers.

Table S3. Characteristics of the included study populations

		Population			Disease characteristics			Instrument administration			
PROM[1]	Reference	N	Age Mean (SD, range) year	Gender % female	Disease	Disease duration mean (SD) year	Disease severity	Setting	Country	Language	Response rate
A	1										
	2										
	3										
……											
B	1										
……											

Note: 1. PROM(s) = Patient-reported outcome measure(s). In this study, PROM(s) refers to the disease-specific HRQL instrument(s) for patients with food allergy or/and food intolerance and their caregivers.

References:
1. Mokkink LB, Prinsen CAC, Patrick DL, et al. COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) - user manual Netherlands: COSMIN; 2018 [updated February 2018; cited 2021 May 5]. Available from: https://cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf accessed May 5 2021.
2. Prinsen CA, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Quality of Life Research 2018;27(5):1147-57. doi: 10.1007/s11136-018-1798-3

Table S4. Rating[1] of the PROMs[2] development

PROM[2]	PROM design							Cognitive interview (CI) study[4]				TOTAL PROM DEVELOPMENT	Reference
	General design requirements					Concept elicitation[3]	Total PROM design	General design requirements	Comprehen-sibility	Comprehen-siveness	Total CI study		
	Clear construct	Clear origin of construct	Clear target population for which the PROM was developed	Clear context of use	PROM developed in sample representing the target population			CI study performed in sample representing the target population					
A													
B													
……													

Note: 1. Ratings (filled in cells): V = very good, A = adequate, D = doubtful, I = inadequate.

2. PROM(s) = Patient-reported outcome measure(s). In this study, PROM(s) refers to the disease-specific HRQL instrument(s) for patients with food allergy or/and food intolerance and their caregivers.

3.The concept elicitation will not be further rated if the PROM(s) was not developed in the sample representing the target population;

4. Empty cells indicate that a CI study (or part of it) was not performed.

References:
1. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res 2018;27(5):1159-70. doi: 10.1007/s11136-018-1829-0
2. Terwee CB, Prinsen CA, Chiarotto A, et al. COSMIN methodology for assessing the content validity of PROMs - User manual Amsterdam, The Netherlands: COSMIN; 2018 [updated February 2018; cited 2021 April 27]. Available from: https://cosmin.nl/wp-content/uploads/COSMIN-methodology-for-content-validity-user-manual-v1.pdf accessed April 27 2021.

Table S5. Methodology qualities of the studies on measurement properties of the PROMs[1], and results of and rating on measurement properties of the PROM(s)[1]

PROM[1] (Reference)	Content validity[2]										Structural validity			Internal consistency			Cross-cultural validity			Reliability			Measurement error			Criterion validity			Construct validity						Responsiveness											
	Asking patients						Asking experts																						Convergent validity			Known groups validity			Comparison with gold standard			Comparison with other instruments			Comparison between subgroups			Comparison before and after intervention		
	Relevance		Comprehensiveness		Comprehensibility		Relevance		Comprehensiveness																																						
	n	Meth qual[3]	n	Meth qual	n	Meth qual	n	Meth qual	n	Meth qual	n	Meth qual	Result (rating[4])	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)	n	Meth qual	Result (rating)
A (Ref 1)																																														
A (Ref 2)																																														
A (Ref 3)																																														
……																																														
B (Ref 1)																																														
……																																														

Note: 'n' means the sample size. 'Meth qual' means 'methodology quality'. Empty cells indicate that the information is not provided by the corresponding reference.

1. PROM(s) = Patient-reported outcome measure(s). In this study, PROM(s) refers to the disease-specific HRQL instrument(s) for patients with food allergy or/and food intolerance and their caregivers.

2. Given that the criteria and rating systems for evaluating the content validity of PROMs are different from those for other measurement properties, the rating results of content validity are not included in this table but separately shown in following Table S5-1.

3. Ratings (filled in cells) for Methodological quality: 'V' = very good, 'A' = adequate, 'D' = doubtful, 'I' = inadequate.

4. Ratings (filled in cells) for measurement properties of the PROMs: ' + '= sufficient, ' − '= insufficient, '?' =indeterminate.

References:
1. Mokkink LB, Prinsen CAC, Patrick DL, et al. COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) - user manual Netherlands: COSMIN; 2018 [updated February 2018; cited 2021 May 5]. Available from: https://cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf accessed May 5 2021.
2. Prinsen CA, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Quality of Life Research 2018;27(5):1147-57. doi: 10.1007/s11136-018-1798-310.1007/s11136-017-1765-4
3. Mokkink LB, De Vet HC, Prinsen CA, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Quality of Life Research 2018;27(5):1171-79. doi: 10.1007/s11136-017-1765-4

Table S5-1. Rating of the content validity of PROMs[1]

PROM (Reference – study type/Rating of reviewers)	Content Validity													CONTENT VALIDITY RATING[3]
	Relevance[2]						Comprehensiveness[2]		Comprehensibility[2]					
	1. Are the included items relevant for the construct of interest?	2. Are the included items relevant for the target population of interest?[4]	3. Are the included items relevant for the context of use of interest?[4]	4. Are the response options appropriate?	5. Is the recall period appropriate?	RELEVANCE RATING[3]	6. Are all key concepts included?	COMPREHENSIVENESS RATING[3]	7. Are the PROM instructions understood by the population of interest as intended?	8. Are the PROM items and response options understood by the population of interest as intended?	9. Are the PROM items appropriately worded?	10. Do the response options match the question?	COMPREHENSIBILITY RATING[3]	
A (Ref 1- PROM development study)														
A (Ref 2 - Content validity study)														
A (Ref 3 - Content validity study)														
Rating of reviewers														
B (Ref 1- PROM development study)														
B (Ref 2 - Content validity study)														
Rating of reviewers														
……														

Note: 1. PROM(s) = Patient-reported outcome measure(s). In this study, PROM(s) refers to the disease-specific HRQL instrument(s) for patients with food allergy or/and food intolerance and their caregivers.

2. Ratings (filled in white cells) for the 10 criteria for relevance, comprehensiveness, comprehensibility can be + / − /±/ ?: ' + '= sufficient, ' − '= insufficient, '±' = inconsistent, '?' =indeterminate.

3. The RELEVANCE, COMPREHENSIVENESS, COMPREHESIBILITY, AND CONTENT VALIDITY ratings (filled in gray cells) can be + / − /±/ ? : ' + '= sufficient, ' − '= insufficient, '±' = inconsistent, '?' =indeterminate.

References:
1. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res 2018;27(5):1159-70. doi: 10.1007/s11136-018-1829-0
2. Terwee CB, Prinsen CA, Chiarotto A, et al. COSMIN methodology for assessing the content validity of PROMs - User manual Amsterdam, The Netherlands: COSMIN; 2018 [updated February 2018; cited 2021 April 27]. Available from: https://cosmin.nl/wp-content/uploads/COSMIN-methodology-for-content-validity-user-manual-v1.pdf accessed April 27 2021.

Table S6. Quality of the PROMs[1] and quality of the evidence for measurement properties of the PROMs[1] (Summary of findings)

Measurement properties	PROM[1] A			PROM[1] B			……		
	Summary or pooled results	Overall rating[2,3]	Quality of evidence[4]	Summary or pooled results	Overall rating[2,3]	Quality of evidence[4]	Summary or pooled results	Overall rating[2,3]	Quality of evidence[4]
Content validity[2]									
Relevance[2]									
Comprehensiveness[2]									
Comprehensibility[2]									
Structural validity[3]									
Internal consistency[3]									
Cross-cultural validity /measurement invariance[3]									
Reliability[3]									
Measurement error[3]									
Criterion validity[3]									
Construct validity[3]									
Responsiveness[3]									

Note: Empty cells indicate that the information is not provided by included studies.

1. PROM(s) = Patient-reported outcome measure(s). In this study, PROM(s) refers to the disease-specific HRQL instrument(s) for patients with food allergy or/and food intolerance and their caregivers.

2. Overall ratings (filled in gray cells) for the content validity (relevance, comprehensiveness, comprehensibility) can only be + / − /±: ' + '= sufficient, ' − '= insufficient, '±' = inconsistent.

3. Overall ratings (filled in white cells) for other measurement properties can be + / − /±/ ?: ' + '= sufficient, ' − '= insufficient, '±' = inconsistent, '?' =indeterminate.

4. Ratings for quality of evidence: High, Moderate, Low, Very low.

References:

1. Mokkink LB, Prinsen CAC, Patrick DL, et al. COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) - user manual Netherlands: COSMIN; 2018 [updated February 2018; cited 2021 May 5]. Available from: https://cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf accessed May 5 2021.
2. Terwee CB, Prinsen CA, Chiarotto A, et al. COSMIN methodology for assessing the content validity of PROMs - User manual Amsterdam, The Netherlands: COSMIN; 2018 [updated February 2018; cited 2021 April 27]. Available from: https://cosmin.nl/wp-content/uploads/COSMIN-methodology-for-content-validity-user-manual-v1.pdf accessed April 27 2021.

Table S7. Information on interpretability of the PROMs[1]

PROM (Reference)	Distribution of the instruments scores in the study population	Percentage of missing items and percentage of missing total scores	Floor and ceiling effects	Scores and change scores available for relevant (sub)groups	Minimal important change (MIC) or minimal important difference (MID)	Information on response shift
A (Ref 1)						
A (Ref 2)						
A (Ref 3)						
……						
B (Ref 1)						
……						

Note: 1. PROM(s) = Patient-reported outcome measure(s). In this study, PROM(s) refers to the disease-specific HRQL instrument(s) for patients with food allergy or/and food intolerance and their caregivers.

References:
1. Mokkink LB, Prinsen CAC, Patrick DL, et al. COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) - user manual Netherlands: COSMIN; 2018 [updated February 2018; cited 2021 May 5]. Available from: https://cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf accessed May 5 2021.

Table S8. Information on feasibility of the PROMs[1]

Feasibility aspects	**PROM A**	**PROM B**	**……**
Patient's comprehensibility			
Clinician's comprehensibility			
Type and ease of administration			
Length of the instrument			
Completion time			
Patient's required mental and physical ability level			
Ease of standardization			
Ease of score calculation			
Copyright			
Cost of an instrument			
Required equipment			
Availability in different settings			
Regulatory agency's requirement for approval			

Note: 1. PROM(s) = Patient-reported outcome measure(s). In this study, PROM(s) refers to the disease-specific HRQL instrument(s) for patients with food allergy or/and food intolerance and their caregivers.

References:
1. Mokkink LB, Prinsen CAC, Patrick DL, et al. COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) - user manual Netherlands: COSMIN; 2018 [updated February 2018; cited 2021 May 5]. Available from: https://cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf accessed May 5 2021.

Table S9: Search strategy for Embase

1	#1	Search: 'quality of life'/exp OR (life NEXT/1 qualit*):ab,ti OR 'quality of life':ab,ti OR 'daily life activity'/exp OR 'activities of daily living':ab,ti OR ('daily living' NEXT/1 activit*):ab,ti OR ('daily live' NEXT/1 activit*):ab,ti OR 'adl':ab,ti OR 'chronic limitation of activity':ab,ti OR (self NEXT/1 care*):ab,ti OR 'health status'/exp OR 'health status':ab,ti OR 'level of health':ab,ti OR (health NEXT/1 level*):ab,ti OR 'qol':ab,ti OR 'hrql':ab,ti OR 'hrqol':ab,ti
2	#2	Search: 'food allergy'/exp OR 'nutritional intolerance'/exp OR 'food allerg*' OR 'food hypersensitivit*' OR 'food intolerance*' OR 'food sensitivit*'
3	#3	Search: 'HR-PRO':ab,ti OR 'HRPRO':ab,ti OR 'HRQL':ab,ti OR 'HRQoL':ab,ti OR 'QL':ab,ti OR 'QoL':ab,ti OR 'quality of life':ab,ti OR 'life quality':ab,ti OR (health NEXT/1 index*):ab,ti OR 'health indices':ab,ti OR (health NEXT/1 profile*):ab,ti OR 'health status' OR (('patient':ab,ti OR 'self':ab,ti OR 'child':ab,ti OR 'parent':ab,ti OR 'carer':ab,ti OR 'proxy':ab,ti) AND (('report':ab,ti OR 'reported':ab,ti OR 'reporting':ab,ti) OR ('rated':ab,ti OR 'rating':ab,ti OR 'ratings':ab,ti) OR 'based':ab,ti OR ('assessed':ab,ti OR 'assessment':ab,ti OR 'assessments':ab,ti))) OR (('disability':ab,ti OR 'function':ab,ti OR 'functional':ab,ti OR 'functions':ab,ti OR 'subjective':ab,ti OR 'utility':ab,ti OR 'utilities':ab,ti OR 'wellbeing':ab,ti OR 'well being':ab,ti) AND ('index':ab,ti OR 'indices':ab,ti OR 'instrument':ab,ti OR 'instruments':ab,ti OR 'measure':ab,ti OR 'measures':ab,ti OR 'questionnaire':ab,ti OR 'questionnaires':ab,ti OR 'profile':ab,ti OR 'profiles':ab,ti OR 'scale':ab,ti OR 'scales':ab,ti OR 'score':ab,ti OR 'scores':ab,ti OR 'status':ab,ti OR 'survey':ab,ti OR 'surveys':ab,ti))
4	#4	Search: 'intermethod comparison'/exp OR 'data collection method'/exp OR 'validation study'/exp OR 'feasibility study'/exp OR 'pilot study'/exp OR 'psychometry'/exp OR 'reproducibility'/exp OR reproducib*:ab,ti OR 'audit':ab,ti OR psychometr*:ab,ti OR clinimetr*:ab,ti OR clinometr*:ab,ti OR 'observer variation'/exp OR 'observer variation':ab,ti OR 'discriminant analysis'/exp OR 'validity'/exp 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 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
5	#5	Search: ('addresses':it OR 'biography':it OR 'case reports':it OR 'comment':it OR 'directory':it OR 'editorial':it OR 'festschrift':it OR 'interview':it OR 'lectures':it OR 'legal cases':it OR 'legislation':it OR 'letter':it OR 'news':it OR 'newspaper article':it OR 'patient education handout':it OR 'popular works':it OR 'congresses':it OR 'consensus development conference':it OR 'consensus development conference':it OR 'practice guideline':it) NOT ('animal'/exp NOT 'human'/exp)
6	#6	#1 AND #2 AND #3 AND #4 NOT #5

Table S10: Search strategy for Web of Science

1	#1	Search: TS=(“quality of life”) OR TS=("life qualit*") OR TS=("living qualit*") OR TS=("quality of living") OR TS=("activities of daily living") OR TS=("activity of daily living") OR TS=("activities of daily life") OR TS=("activity of daily life") OR TS=("daily living activit*") OR TS=("daily life activit*") OR TS=(adl) OR TS=("chronic limitation of activity") OR TS=("self care*") OR TS=("health status") OR TS=("level of health") OR TS=("health level*") OR TS=(qol) OR TS=(hrql) OR TS=(hrqol)
2	#2	Search: TS=(“food allerg*”) OR TS=(“food hypersensitivit*”) OR TS=(“food intolerance*”) OR TS=(“food sensitivit*”)
3	#3	Search: TS=(HR-PRO) OR TS=(HRPRO) OR TS=(HRQL) OR TS=(HRQoL) OR TS=(QL) OR TS=(QoL) OR ALL=(“quality of life”) OR ALL=(“life quality”) OR TS=(“health index*”) OR TS=(“health indices”) OR TS=(“health profile*”) OR ALL=(“health status”) OR ((TS=(patient) OR TS=(self) OR TS=(child) OR TS=(parent) OR TS=(carer) OR TS=(proxy)) AND (TS=(report) OR TS=(reported) OR TS=(reporting) OR TS=(rated) OR TS=(rating) OR TS=(ratings) OR TS=(based) OR TS=(assessed) OR TS=(assessment) OR TS=(assessments))) OR ((TS=(disability) OR TS=(function) OR TS=(functional) OR TS=(functions) OR TS=(subjective) OR TS=(utility) OR TS=(utilities) OR TS=(wellbeing) OR TS=(“well being”)) AND (TS=(index) OR TS=(indices) OR TS=(instrument) OR TS=(instruments) OR TS=(measure) OR TS=(measures) OR TS=(questionnaire) OR TS=(questionnaires) OR TS=(profile) OR TS=(profiles) OR TS=(scale) OR TS=(scales) OR TS=(score) OR TS=(scores) OR TS=(status) OR TS=(survey) OR TS=(surveys)))
4	#4	Search: TS=(instrumentation) OR TS=(methods) OR TS=(“validation stud*”) OR TS=(“comparative stud*”) OR TS=(psychometrics) OR TS=(psychometr*) OR ALL=(clinimetr*) OR ALL=(clinometr*) OR TS=(“outcome assessment”) OR TS=(“outcome measure”) OR TS=(“observer variation”) OR TS=(“observer variation”) OR TS=(“health status indicators”) OR TS=(“reproducib*”) OR TS=(“discriminant analysis”) OR TS=(reliab*) OR TS=(unreliab*) OR TS=(valid*) OR TS=(“coefficient of variation”) OR TS=(coefficient) OR TS=(homogeneity) OR TS=(homogeneous) OR TS=(“internal consistency”) OR ((TS=(alpha) OR TS=(alphas)) AND TS=(cronbach*)) OR ((TS=(correlation*) OR TS=(selection*) OR TS=(reduction*)) AND TS=(item)) OR TS=(agreement) OR TS=(precision) OR TS=(imprecision) OR TS=(precise values) OR TS=(test-retest) OR (TS=(test) AND TS=(retest)) OR ((TS=(test) OR TS=(retest)) AND TS=(reliab*)) OR TS=(stability) OR TS=(interrater) OR TS=(inter-rater) OR TS=(intrarater) OR TS=(intra-rater) OR TS=(intertester) OR TS=(inter-tester) OR TS=(intratester) OR TS=(intra-tester) OR TS=(interobserver) OR TS=(inter-observer) OR TS=(intraobserver) OR TS=(intra-observer) OR TS=(intertechnician) OR TS=(inter-technician) OR TS=(intratechnician) OR TS=(intra-technician) OR TS=(interexaminer) OR TS=(inter-examiner) OR

		TS=(intraexaminer) OR TS=(intra-examiner) OR TS=(interassay) OR TS=(inter-assay) OR TS=(intraassay) OR TS=(intra-assay) OR TS=(interindividual) OR TS=(inter-individual) OR TS=(intraindividual) OR TS=(intra-individual) OR TS=(interparticipant) OR TS=(inter-participant) OR TS=(intraparticipant) OR TS=(intra-participant) OR TS=(kappa) OR TS=(kappa's) OR TS=(kappas) OR TS=(repeatab*) OR ((ALL=(replicab*) OR ALL=(repeated)) AND (ALL=(measure) OR ALL=(measures) OR ALL=(findings) OR ALL=(result) OR ALL=(results) OR ALL=(test) OR ALL=(tests))) OR TS=(generaliza*) OR TS=(generalisa*) OR TS=(concordance) OR (TS=(intraclass) AND TS=(correlation*)) OR TS=(discriminative) OR TS=(known group) OR TS=("factor analysis") OR TS=("factor analyses") OR TS=("factor structure") OR TS=("factor structures") OR TS=(dimension*) OR TS=(subscale*) OR ((TS=(analysis) OR TS=(analyses)) AND TS=(scaling) AND TS=(multitrait)) OR TS=("item discriminant") OR TS=("interscale correlation*") OR TS=(error) OR TS=(errors) OR TS=("individual variability") OR TS=("interval variability") OR TS=("rate variability") OR ((TS=(values) OR TS=(analysis)) AND TS=(variability)) OR ((TS=(measurement) OR TS=(measuring)) AND TS=(uncertainty)) OR TS=("standard error of measurement") OR TS=(sensitiv*) OR TS=(responsive*) OR (TS=(limit) AND TS=(detection)) OR TS=("minimal detectable concentration") OR TS=(interpretab*) OR ((TS=(minimal) OR TS=(minimally) OR TS=(clinical) OR TS=(clinically)) AND (TS=(important) OR TS=(significant) OR TS=(detectable)) AND (TS=(change) OR TS=(difference))) OR (TS=(small) AND (TS=(real) OR TS=(detectable)) AND (TS=(change) OR TS=(difference))) OR TS=("meaningful change") OR TS=("ceiling effect") OR TS=("floor effect") OR TS=("Item response model") OR TS=(IRT) OR TS=(Rasch) OR TS=("differential item functioning") OR TS=(DIF) OR TS=("computer adaptive testing") OR TS=("item bank") OR TS=("cross-cultural equivalence")
5	#5	Search: DT=(Art Exhibit Review) OR DT=(Biographical-Item) OR DT=(Chronology) OR DT=(Correction) OR DT=(Correction, Addition) OR DT=(Dance Performance Review) OR DT=(Fiction, Creative Prose) OR DT=(Film Review) OR DT=(Hardware Review) OR DT=(Letter) OR DT=(Meeting Abstract) OR DT=(Meeting Summary) OR DT=(Music Performance Review) OR DT=(Music Score) OR DT=(Music Score Review) OR DT=(News Item) OR DT=(Note) OR DT=(Poetry) OR DT=(Retracted Publication) OR DT=(Retracted Publication) OR DT=(Script) OR DT=(Software Review) OR DT=(Theater Review) OR DT=(TV Review, Radio Review) OR DT=(TV Review, Radio Review Video) OR DT=(Withdrawn Publication)
6	#6	#1 AND #2 AND #3 AND #4 NOT #5

Table S11: Search strategy for Scopus

1	#1	The search strategy of #1 is the same as that in Web of Science, considering there is no use of Mesh words or Emtree words in these databases.
2	#2	The search strategy of #2 is the same as that in Web of Science, considering there is no use of Mesh words or Emtree words in these databases.
3	#3	The search strategy of #3 is the same as that in Web of Science.
4	#4	The search strategy of #4 is the same as that in Web of Science, considering there is no use of Mesh words or Emtree words in these databases.
5	#5	#1 AND #2 AND #3 AND #4

Note: The exclusion of some publication types will not be included in the literature search strategy in Scopus, because there is no corresponding filter. This step will be included in Literature Screening step.

Table S12: Search strategy for CINAHL

1	#1	The search strategy of #1 is the same as that in PubMed.
2	#2	The search strategy of #2 is the same as that in PubMed.
3	#3	The search strategy of #3 is the same as that in PubMed.
4	#4	The search strategy of #4 is the same as that in PubMed.
5	#5	The search strategy of #5 is the same as that in PubMed.
6	#6	#1 AND #2 AND #3 AND #4 NOT #5

Table S13: Search strategy for ProQuest (Health & Medical Collection)

1	#1	The search strategy of #1 is the same as that in PubMed.
2	#2	The search strategy of #2 is the same as that in PubMed.
3	#3	The search strategy of #3 is the same as that in PubMed.
4	#4	The search strategy of #4 is the same as that in PubMed.
5	#5	#1 AND #2 AND #3 AND #4

Note: The exclusion of some publication types will not be included in the literature search strategy in ProQuest (Health & Medical Collection), because there is no corresponding filter. This step will be included in Literature Screening step.

COSMIN content validity methodology

Instructions for completing the COSMIN boxes for content validity

1 CHECK the COSMIN website if the quality of the PROM development was already rated in another review. In that case, you can skip box 1A and use the quality
2 We recommend to score all PROMS with two raters, indepedently, and reach consensus afterwards. You can change "rater 1" and "rater 2" into the names of the raters
4 Add extra rows, columns or tables if needed
5 Tables 1, 2, and 3 will be filled automatically (you may need to add links to the other tabs). They can be included in a systematic review

COSMIN box 1. Standards for evaluating the quality of PROM development

Check the COSMIN website to see if the quality of the PROM development was already rated in another review

Ratings: V= very good; A = adequate; D = doubtful; I = inadequate; N= not applicable

1a. PROM design

	PROM ref			PROM ref			PROM ref		
General design requirements	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus
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									
4 Is a clear description provided of the context of use (i.e. discriminative, evaluative									
5 Was the PROM development study performed in a sample representing the target population for which the PROM was developed?									
Concept elicitation (relevance and comprehensiveness)	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus
6 Was an appropriate qualitative data collection method used to identify relevant items for									
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: was the sample size appropriate?									
SUBTOTAL QUALITY CONCEPT ELICITATION STUDY *Lowest score of items 6-13*									
TOTAL QUALITY OF THE PROM DESIGN *Lowest score of items 1-13*									

1b. Cognitive interview study or other pilot test

	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus
14 Was a cognitive interview study or other pilot test performed? *If NO skip items 15-35*									
General design requirements	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus
15 Was the cognitive interview study or other pilot test performed in a sample representing the target population?									
Comprehensibility	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus
16 Were patients asked about the comprehensibility of the PROM? *If NO or not clear, skip*									
	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus
17 Were all items tested in their final form?									
18 Was an appropriate qualitative method used to assess the comprehensibility of the PROM									
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?									
	SUBTOTAL QUALITY OF COMPREHENSIBILITY STUDY *Lowest score of items 15-25*									
	Comprehensiveness	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus
26	Were patients asked about the comprehensiveness of the PROM? *If NO or not clear, skip items 27-35*									
		Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus	Rater 1	Rater 2	Consensus
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?									
	SUBTOTAL QUALITY OF COMPREHENSIVENESS STUDY *Lowest score of items 15, 26-35*									
	TOTAL QUALITY OF THE PILOT STUDY *Lowest score of items 14-35*									
	TOTAL QUALITY OF THE PROM DEVELOPMENT STUDY *Lowest score of items 1-35*									

COSMIN box 2. Standards for evaluating the quality of content validity studies of PROMs

Only those parts of the box need to be completed for which information is available

Score: V= very good; A = adequate; D = doubtful; I = inadequate; N= not applicable

	PROM			PROM			PROM		
	ref			ref			ref		
2a. Asking patient about relevance	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus
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?									
SUBTOTAL QUALITY OF RELEVANCE STUDY *Lowest score of items 1-7*									
2b. Asking patients about comprehensiveness	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus
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?									
SUBTOTAL QUALITY OF COMPREHENSIVENESS STUDY *Lowest score of items 8-14*									
2c. Asking patients about comprehensibility	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus
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?									
SUBTOTAL QUALITY OF COMPREHENSIBILITY STUDY *Lowest score of items 15-21*									
2d. Asking professionals about relevance	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus
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?									
SUBTOTAL QUALITY OF RELEVANCE STUDY *Lowest score of items 22-26*									

2e. Asking professionals about comprehensiveness	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus
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?									
SUBTOTAL QUALITY OF COMPREHENSIVENESS STUDY *Lowest score of items 27-31*									

Rating the content validity of the PROM

Complete one tabel per PROM (subscale)

Criteria for content validity

To fill in ratings use apostrophe (') before the + / - / ± / ? signs

Score: + = sufficient; - = insufficient; ? = indeterminate; ± = inconsistent

PROM (subscale)	PROM development study	PROM development study	PROM development study	Content validity study 1	Content validity study 1	Content validity study 1	Content validity study 2[2]	Content validity study 2[2]	Content validity study 2[2]	Rating of reviewers	Rating of reviewers	Rating of reviewers	OVERALL RATINGS PER PROM[3]	OVERALL RATINGS PER PROM[3]	OVERALL RATINGS PER PROM[3]	QUALITY OF EVIDENCE	QUALITY OF EVIDENCE	QUALITY OF EVIDENCE
	(+ / - / ± / ?)	(+ / - / ± / ?)	(+ / - / ± / ?)	(+ / - / ± / ?)	(+ / - / ± / ?)	(+ / - / ± / ?)	(+ / - / ± / ?)	(+ / - / ± / ?)	(+ / - / ± / ?)	(+ / - / ± / ?)	(+ / - / ± / ?)	(+ / - / ± / ?)	+ / - / ±	+ / - / ±	+ / - / ±	High, moderate, low, very low	High, moderate, low, very low	High, moderate, low, very low
	rater 1	rater 2	consensus	rater 1	rater 2	consensus	rater 1	rater 2	consensus	rater 1	rater 2	consensus	rater 1	rater 2	consensus	rater 1	rater 2	consensus
Relevance																		
1 Are the included items relevant for the construct of interest?[1]																		
2 Are the included items relevant for the target population of interest?[1]																		
3 Are the included items relevant for the context of use of interest?[1]																		
4 Are the response options appropriate?																		
5 Is the recall period appropriate?																		
RELEVANCE RATING (+ / - / ± / ?)																		
Comprehensiveness																		
6 Are all key concepts included?																		
COMPREHENSIVENESS RATING (+ / - / ± / ?)																		
Comprehensibility																		
7 Are the PROM instructions understood by the population of interest as intended?																		
8 Are the PROM items and response options understood by the population of interest as intended?																		
9 Are the PROM items appropriately worded?																		
10 Do the response options match the question?																		
COMPREHENSIBILITY RATING (+ / - / ± / ?)																		
CONTENT VALIDITY RATING (+ / - / ± / ?)																		

[1] These criteria refer to the construct, population, and context of use of interest in the systematic review.

[2] Add more columns if more content validity studies are available

[3] If ratings are inconsistent between studies, consider using separate tables for subgroups of studies with consistent results.

COSMIN Risk of Bias checklist

Only those parts of the boxes need to be completed for which information is available

Score: V= very good; A = adequate; D = doubtful; I = inadequate; N= not applicable

	PROM			PROM			PROM		
Article reference:	ref			ref			ref		
3. Structural validity	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus
unidimensionality or 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?									
TOTAL *Lowest score of items 1-4*			V						
4. Internal consistency	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus
1 Was an internal consistency statistic calculated for each unidimensional (sub)scale 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									
5 Were there any other important flaws?									
TOTAL *Lowest score of items 1-5*			V						
5. Cross-cultural validity\measurement invariance	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus
1 Were the samples similar for relevant characteristics except for the group									
2 Was an adequate approach used to analyse the data?									
3 Was the sample size included in the analysis adequate?									
4 Were there any other important flaws?									
TOTAL *Lowest score of items 1-4*			V						
6. Reliability	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus
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?									
TOTAL *Lowest score of items 1-8*			V						
7. Measurement error	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus
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?								
	TOTAL *Lowest score of items 1-6*			v					

8. Criterion validity	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus
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?									
TOTAL *Lowest score of items 1-3*			v						

9. Hypotheses testing for construct validity									
9a. Comparison with other outcome measurement instruments (convergent validity)	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus
1 Is it clear what the comparator instrument(s) measure(s)?									
2 Were the measurement properties of the comparator instrument(s) adequate?									
3 Was the statistical method appropriate for the hypotheses to be tested?									
4 Were there any other important flaws?									
TOTAL *Lowest score of items 1-4*			v						
9b. Comparison between subgroups (discriminative or known-groups validity)	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus
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?									
TOTAL *Lowest score of items 5-7*			v						

10. Responsiveness									
10a. Criterion approach (i.e. comparison to a gold standard)	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus
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?									
TOTAL *Lowest score of items 1-3*			v						
10b. Construct approach (i.e. hypotheses testing; comparison with other outcome m	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus
4 Is it clear what the comparator instrument(s) measure(s)?									
5 Were the measurement properties of the comparator instrument(s) adequate?									
6 Was the statistical method appropriate for the hypotheses to be tested?									
7 Were there any other important flaws?									
TOTAL *Lowest score of items 4-7*			v						
10c. Construct approach: (i.e. hypotheses testing: comparison between subgroups)	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus
8 Was an adequate description provided of important characteristics of the									
9 Was the statistical method appropriate for the hypotheses to be tested?									
10 Were there any other important flaws?									

TOTAL *Lowest score of items 8-10*			v						
10d. Construct approach: (i.e. hypotheses testing: before and after intervention)	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus	rater 1	rater 2	Consensus
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?									
TOTAL *Lowest score of items 11-13*			v						

Rating the measurement properties of the PROM

Use one Table per PROM
Add additional columns (studies) if necessary

PROM	Study 1			Study 2			Study 3			OVERALL					
	RATING	RATING	RATING	RATING	RATING	RATING	RATING	RATING	RATING	OVERALL RATING	OVERALL RATING	OVERALL RATING	QUALITY OF EVIDENCE	QUALITY OF EVIDENCE	QUALITY OF EVIDENCE
	+/-/?	+/-/?	+/-/?	+/-/?	+/-/?	+/-/?	+/-/?	+/-/?	+/-/?	+/-/±/?	+/-/±/?	+/-/±/?	High, moderate, low, very low	High, moderate, low, very low	High, moderate, low, very low
	rater 1	rater 2	consensus	rater 1	rater 2	consensus	rater 1	rater 2	consensus	rater 1	rater 2	consensus	rater 1	rater 2	consensus
Structural validity															
Internal consistency															
Cross-cultural validity															
Measurement invariance															
Reliability															
Measurement error															
Criterion validity															
Construct validity															
Responsiveness															

PROM	Study 1			Study 2			Study 3			OVERALL					
	RATING	RATING	RATING	RATING	RATING	RATING	RATING	RATING	RATING	OVERALL RATING	OVERALL RATING	OVERALL RATING	QUALITY OF EVIDENCE	QUALITY OF EVIDENCE	QUALITY OF EVIDENCE
	+/-/?	+/-/?	+/-/?	+/-/?	+/-/?	+/-/?	+/-/?	+/-/?	+/-/?	+/-/±/?	+/-/±/?	+/-/±/?	High, moderate, low, very low	High, moderate, low, very low	High, moderate, low, very low
	rater 1	rater 2	consensus	rater 1	rater 2	consensus	rater 1	rater 2	consensus	rater 1	rater 2	consensus	rater 1	rater 2	consensus
Structural validity															
Internal consistency															
Cross-cultural validity															
Measurement invariance															
Reliability															
Measurement error															
Criterion validity															
Construct validity															
Responsiveness															