Recommendations on the use of item libraries for patient-reported outcome measurement in oncology trials

Piccinin, Claire ; Basch, Ethan; Bhatnagar, Vishal ; Calvert, Melanie; Campbell, Alicyn; Cella, David ; Cleeland, Charles; Coens, Corneel; Darlington, Anne-Sophie ; Dueck, Amylou C; Groenvold, Mogens; Herold, Ralf; King-Kallimanis, Bellinda L. ; Kluetz, Paul G; Kulić, Dagmara ; O'Connor, Daniel; Oliver, Kathy; Pe, Madeline; Reeve, Bryce B.; Reijneveld, Jaap C

DOI:
10.1016/S1470-2045(22)00654-4

Document Version
Peer reviewed version

Citation for published version (Harvard):
Piccinin, C, Basch, E, Bhatnagar, V, Calvert, M, Campbell, A, Cella, D, Cleeland, C, Coens, C, Darlington, A-S, Dueck, AC, Groenvold, M, Herold, R, King-Kallimanis, BL, Kluetz, PG, Kulić, D, O'Connor, D, Oliver, K, Pe, M, Reeve, BB, Reijneveld, JC, Wang, XS & Bottomley, A 2023, 'Recommendations on the use of item libraries for patient-reported outcome measurement in oncology trials: findings from an international, multidisciplinary working group', *The Lancet Oncology*, vol. 24, no. 2, pp. e86–95. https://doi.org/10.1016/S1470-2045(22)00654-4

TITLE

Recommendations on the use of item libraries for patient-reported outcome measurement in oncology trials: Findings from an international, multidisciplinary working group

AUTHORS

Claire Piccinin[1] MSc, Ethan Basch[2] MD, Vishal Bhatnagar[3] MD, Melanie Calvert[4,5,6,7,8] PhD, Alicyn Campbell[9] MPh, David Cella[10] PhD, Charles S. Cleeland[11] PhD, Corneel Coens[1] MSc, Anne-Sophie Darlington[12] PhD, Amylou C. Dueck[13] PhD, Mogens Groenvold[14,15] MD, Ralf Herold[16] MD, Bellinda L. King-Kallimanis[17] PhD, Paul G. Kluetz[3] MD, Dagmara Kuliś[1] MA, Daniel O'Connor[18] MBChB, Kathy Oliver[19] BA, Madeline Pe[1] PhD, Bryce B. Reeve[20] PhD, Jaap C. Reijneveld[21] MD, Xin Shelley Wang[11] MD, & Andrew Bottomley[1] PhD

AFFILIATIONS

[1]Quality of Life Department, European Organisation for Research and Treatment of Cancer, Brussels, Belgium
[2]Department of Medicine, University of North Carolina, Chapel Hill, North Carolina, USA
[3]Oncology Center of Excellence, U.S. Food and Drug Administration, Silver Spring, Maryland, USA
[4]Centre for Patient-Reported Outcomes Research, Institute of Applied Health Research, University of Birmingham, Birmingham, UK
[5]National Institute for Health and Care Research (NIHR) Applied Research Collaboration West Midlands, Birmingham, UK
[6]Birmingham Health Partners Centre for Regulatory Science and Innovation, University of Birmingham, Birmingham, UK
[7]NIHR Birmingham Biomedical Research Centre, University of Birmingham, Birmingham, UK
[8]NIHR Surgical Reconstruction and Microbiology Research Centre, University of Birmingham, Birmingham, UK
[9]Digital Health Oncology R&D, AstraZeneca, Gaithersburg, Maryland, USA
[10]Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, Evanston, Illinois, USA
[11]Department of Symptom Research, M.D. Anderson Cancer Center, University of Texas, Houston, Texas, USA
[12]School of Health Sciences, University of Southampton, Southampton, UK
[13]Department of Quantitative Health Sciences, Mayo Clinic, Scottsdale, Arizona, USA
[14]Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark
[15]University of Copenhagen, Copenhagen, Denmark
[16]European Medicines Agency, Amsterdam, Netherlands
[17]LUNGevity Foundation, Chicago, Illinois, USA
[18]Medicines and Healthcare Products Regulatory Agency, London, UK
[19]International Brain Tumour Alliance, Surrey, UK
[20]Department of Population Health Sciences, Duke University School of Medicine, Durham, North Carolina, USA
[21]Brain Tumor Center Amsterdam, Amsterdam University Medical Center, Amsterdam, Netherlands

CORRESPONDING AUTHOR

Claire Piccinin
Quality of Life Department, European Organisation for Research and Treatment of Cancer
Avenue E. Mounier 83/11, 1200 Brussels, Belgium
claire.piccinin@eortc.org
+32 2 774 1592

SUMMARY

The use of item libraries for patient-reported outcome (PRO) assessment in oncology allows for the customization of PRO assessment to measure key health-related quality of life (HRQOL) concepts of relevance to the target population and intervention. However, no high-level recommendations exist to guide users on the design and implementation of these customized PRO measures (item lists) across different PRO measurement systems. To address this issue, a working group was set up, including international stakeholders (academic, independent, industry, health technology assessment, regulatory, and patient advocacy), with the goal of creating recommendations for the use of item libraries in oncology trials. A scoping review was carried out to identify relevant publications and highlight any gaps. Stakeholders commented on the available guidance for each research question, proposed recommendations on gaps, and came to an agreement using discussion-based methods. Nine primary research questions were identified that formed the scope and structure of the recommendations on how to select items and implement item lists created from item libraries. These recommendations address methods to drive item selection, plan the structure and analysis of item lists, and facilitate their use in conjunction with other measures. The findings resulted in high-level, instrument-agnostic recommendations on the use of item library-derived item lists in oncology trials.

BACKGROUND

It is increasingly recognized that patient-reported outcome (PRO) measures designed to assess patients' symptoms, functioning, and general health status are critical for capturing the impact of disease and treatment on patients' lives. [1–3] Most standard PRO measures are static, i.e., they present the same set of items (questions) at every assessment for all patients. However, as the use of PRO measures becomes more widespread, there is also an increasing recognition that standard, static PRO measures sometimes fail to measure key health domains that are relevant for specific studies, contexts, populations, and stakeholders. This may be especially true for innovative treatments, given faster evaluation and approval times or for rare cancer groups, for whom questionnaire development may be challenging. The rise in the availability of item libraries addresses the need for a flexible approach to assess specific symptoms, functioning, health status, and other health-related quality of life (HRQOL) domains [4] for oncology research and clinical care.

PRO item libraries are collections of single items and/or multi-item scales that measure HRQOL domains including disease-related symptoms, symptomatic adverse events (AE), functioning, and overall health status. In contrast to static questionnaires, researchers and clinicians can select specific items from the library to measure only relevant PRO domains for a given context or target population (glossary of keywords in web appendix, page 1). In case of administration of multiple PRO measures, the flexibility afforded by item libraries may help to minimize patient burden through use of customized measures. In this publication, we refer to this customized item selection as an item list. While some item libraries are derived from existing, validated questionnaires (e.g., European Organisation for Research and Treatment of Cancer (EORTC) Item Library, [5] Functional Assessment of Chronic Illness Therapy (FACIT) Searchable Library, [6] MD Anderson Symptom Inventory (MDASI) Symptom Library [7]), and allow for the flexible use of items originally validated within the scope of standard questionnaire development, others have been designed with the specific aim of creating a flexible library of items (e.g., Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) [8]).

Item banks (e.g., Patient-Reported Outcomes Measurement Information System (PROMIS) [9], EORTC Computerized Adaptive Testing (CAT) Core [10]) are a special case of item libraries in that all the items included for each HRQOL domain have been calibrated with an item response theory (IRT) model. Item banks allow investigators to generate multiple short forms from the same item bank and they allow for CAT, which tailors the PRO measure based on how a patient answers each item. CAT measures or short forms derived from the same item bank can be compared or combined with each other. Within these recommendations, we distinguish item libraries from item banks. Although higher level recommendations related to outcome and tool selection/implementation may also be applicable for CAT-derived measures, CAT-specific recommendations are beyond the scope of this work.

Inclusion of PRO measures with the goal of capturing patient views can provide insight into the assessment of endpoints like safety, efficacy, and tolerability. [11] Flexible approaches such as item libraries can help capture outcomes of importance to different stakeholders that may be missing from static PRO measures. Given the number of choices end users face regarding which item libraries to use, how to select and analyse items, and how to combine these with other measures, there is an important need for guidance.

To address the issue of implementing PRO assessment with fit-for-purpose item selection from item libraries in oncology trials, an international, multidisciplinary group was established, including various key developers of item libraries along with representatives from industry, academia, regulatory, health technology assessment, and patient advocacy organizations. The primary goal of the working group (WG) was to develop best practice recommendations for the use of oncology-specific item libraries in controlled clinical trials through a combination of evidence- and discussion-based methods.

DEVELOPMENT OF RECOMMENDATIONS

Identifying relevant stakeholders and scope of work. A WG was convened by the EORTC members, with the aim of creating a balanced group of stakeholders and ensuring representation from the key PRO item libraries (EORTC Item Library, FACIT Searchable Library, MDASI Symptom Library, PRO-CTCAE, and PROMIS). Seventeen external (i.e., non-EORTC) members were invited to collaborate and all joined the WG. The final group included representatives with various backgrounds: health technology assessment/regulatory (4), academic/independent (15), industry (1), and patient advocacy (2). Two meetings were held with WG members to plan the scope, refine aims, and determine topics for recommendations.

Scoping review. Once the specific aims were identified and agreed upon, a scoping review was carried out to identify relevant publications and other sources for the recommendations, and to highlight any gaps in the literature. [12,13]

Search strategy and selection criteria. The initial search was performed in PubMed® (using the terms "cancer", "patient-reported outcome", "item library"), and publications (up until 01 December 2020) were retained if they provided explicit high-level recommendations (i.e., not simply reporting an example of use) on creating PRO measures derived from item libraries in oncology populations. Given the paucity of existing published recommendations, the criteria were broadened to include recommendations for PRO measurement in general, that could also be applicable to item libraries (up until 01 May 2022). Item library websites and platforms were searched for available resources and recommendations (including grey literature) and additional publications/sources were shared by WG members. Data were then extracted and compiled into a matrix highlighting research question/topic, available recommendation, and source (reference). In the first version of the matrix, data were listed

per unique source (or research group), for transparency. In a second version, overlapping pieces of evidence and recommendations were merged, with the relevant source(s) listed alongside. Recommendations deemed likely to require more detailed review and agreement (e.g., missing and/or conflicting evidence) were highlighted.

Stakeholders' feedback on current recommendations. Data from the second version of the matrix were inserted into a table with an additional column for feedback. Recommendations identified in the previous stage as requiring additional review remained highlighted. The table was circulated to all WG members for comments (e.g., support or disagreement with a suggested recommendation), questions, and additional references. Online cloud storage was utilised to encourage simultaneous collaboration.

Following review by WG members, comments were analysed and merged, with recommendations adapted accordingly. Given the lack of existing recommendations, the less formal approach, which relied on discussion to reach agreement (instead of formal consensus) was deemed appropriate. [14] The results then formed the basis for the content of the manuscript. A third meeting was held to agree upon the overall structure and contents of the manuscript and address comments and edits.

RECOMMENDATIONS FROM THE WORKING GROUP

The WG identified 9 primary research questions (Table 1) to guide recommendations on how to select items and implement item lists from item libraries in oncology trials. Results of the scoping review confirmed the lack of high-level item library-specific guidelines, with only one article initially retained and 51 articles/sources added following the use of broader inclusion criteria and review by WG members (Figure 1) (full list in web appendix, pages 2-5). The research questions and accompanying recommendations are ordered to reflect a chronological course of events, starting with methods to drive item selection, followed by the structure and analysis of item lists, and ending with the use of item lists in conjunction with other measures and measurement systems.

INSERT TABLE 1

INSERT FIGURE 1

1. Which methods should be used to drive item selection?

In general, clinical trial investigators should assess key issues (i.e., PROs) that inform the evaluation of treatment safety, tolerability, and efficacy. Selection should focus on clinically meaningful disease- and treatment-specific concepts that inform treatment decision-making and patient care, including symptom management. Sources for selecting items can come from the published literature, formal interviews with stakeholders (including patients), and expert input, including patient and public involvement (PPI). The use of clinically meaningful items based on evidence from similar cohorts in

oncology practice may also be considered. Selection of relevant items within an item library should have face validity with regard to the study's aims and research questions and be suitable for the patient population under investigation. [6] Although items may be derived from validated questionnaires and measurement systems, pilot testing of the selected items is recommended, to ensure relevance and comprehensibility for the target population. [15] When a large list of items is identified, an approach to reduce the full list, involving prioritization by both patients and healthcare professionals (HCPs) may be required. [16]

In cases where items are missing and may need to be generated, it is important to liaise with the relevant instrument developers.

The approaches listed below can help guide item selection:

Literature reviews. Systematic literature reviews of HRQOL impact in the study population, including reviews of existing questionnaires and PRO tools, should be carried out to identify possible issues and symptoms. When systematic reviews are not feasible, scoping reviews may be considered.

Interviews and focus groups. One-to-one structured and semi-structured interviews with patients, and specifically patients with the relevant condition(s) and stage(s) of disease and treatment (if possible), should be conducted to elicit relevant issues and ensure patient centricity. A variety of different demographic characteristics (e.g., gender, age, ethnicity, nationality, literacy level) should be covered, to ensure representativeness and wide-spread applicability. [15,17–19] Structured and semi-structured interviews with HCPs (e.g., physicians, nurses, and psychologists) with the relevant clinical expertise [20–22] along with focus groups with HCPs, patients, and patient advocacy organizations, and early meetings with relevant stakeholders (e.g., regulators) may also generate valuable information regarding PROs of interest.

Patient and public involvement. The need to obtain input through PPI should be considered during all stages of research design. [23–28]

Publicly available data and registries. Where available, public sources of PRO data may also provide insight as to the prevalence of symptomatic AEs and disease-related HRQOL issues for specific patient groups. [29] Investigator brochures and existing drug labelling (if applicable) can serve as valuable sources of safety information. Registries maintained by patient advocacy organisations may also highlight important issues.

Retrospective chart reviews. In cases where data are available from prior trials involving human subjects, retrospective chart reviews can be carried out to identify concepts of interest.

The symptoms and HRQOL issues identified using these various approaches can then inform the selection of relevant items.

Specifications for interventional research: early phase trials. In addition to the above-mentioned examples, during early phase (I/II) trials, prior phase I studies and data from compounds using the same mechanism of action should be considered. Symptomatic AEs associated with the most common treatments for the disease, those identified by multi-stakeholder groups, and those that are known to be relevant and burdensome to patients should be considered. [30] Where available, the use of preclinical data may also be relevant. The use of free text reporting may help to capture unexpected or previously unidentified symptomatic AEs. [31] Moreover, at this stage, a broader set of items may be needed to help capture the range of possible symptomatic AEs and issues. It is also important to consider symptomatic AEs which may be specifically linked to the mode of treatment administration (e.g., injection). [32]

Specification for interventional research: late phase trials. During late phase trials (III/IV), inclusion of symptomatic AEs, disease-related symptoms, and HRQOL issues identified during earlier phase trials is recommended. Consultation of investigator brochures and recommendations derived from multi-stakeholder groups may also be considered. During these stages, assessment of overall impact and burden of symptomatic AEs is advised. As in early phase research, the use of free text reporting may be considered. For trials that may be submitted to regulatory agencies, it is important to take available regulatory guidance into account (e.g., [33,34]) and to consider engaging with regulators in seeking scientific input on item library use for the concerned trial.

2. *When should single items vs. multi-item scales be used and what are the benefits and limitations of both?*

When it comes to psychometric properties, there is considerable evidence suggesting that multi-item scales generally outperform single items. In general, multi-item scales have a higher level of precision and are more informative. [35–37] They tend to demonstrate better reliability and content validity and are less prone to floor/ceiling effects, compared to single items. In cases where a concept is intended to discriminate between patients, a multi-item scale may be better suited to capture differences. [38,39] Complex types of functioning and multi-domain concepts (e.g., physical functioning) with different attributes generally require several items to ensure content validity, in the form of a multi-item scale. [20,40,41] Moreover, if the symptom or issue represents a key aspect of the disease or treatment, or if in-depth knowledge of the domain(s) is required, it may be favourable to include a multi-item scale to ensure robust assessment.

However, for pragmatic reasons (e.g., when screening multiple symptoms simultaneously and frequently), to minimize the likelihood of patient burden due to large item lists, it is important to consider which concepts may be sufficiently captured by single items. For example, when assessment of multiple symptoms is the goal, then single items may suffice for most symptoms and may be easier to interpret than complex multi-item scales. In addition to research questions, the study endpoint (e.g., primary, secondary, exploratory) should also be considered.

In cases where burden linked to instrument length is an issue, single items may be favourable given that a broader set of concepts can be covered with fewer items. Ultimately, the choice of single versus multiple items depends on the symptom or domain under investigation, as well as the specific research questions and study design. While some concepts might be captured by one item alone (e.g., symptom presence), others may require more, especially if symptom severity, functional impact, and interference with daily activities are also targeted.

3. *How should different types of psychometric properties be considered and tested, based on the item list/measure and the context of its use?*

When item lists are derived from item libraries that contain validated questionnaires, it may not be necessary to conduct additional psychometric testing. Instead, it is recommended that single items be treated as such and multi-item scales remain intact, unless there is a strong rationale for removing items. We caution against the creation of new multi-item scales unless such work is carried out in close collaboration with the item library developers. If the item list is intended to be administered in a new population, further comparative validation testing may be required to ensure that the psychometric properties are retained. Relevant scores can then be compared to those in the published literature.

A general list of psychometric properties and tests for consideration in evaluating validity, reliability, and responsiveness to change is provided in Table 2. Users should refer to the various sources for more information where necessary. [20–22,37,42–45]

INSERT TABLE 2

4. *How can bias be minimized in the design of item lists?*

Although flexibility in item selection helps to ensure that important symptoms and HRQOL concepts are included, there is also the possibility to omit these, leading to underreporting of symptomatic AEs and other HRQOL concerns. Adopting a rigorous approach to item selection using the methods detailed above can help to minimize bias by incorporating various perspectives and types of evidence. Transparency regarding the item selection process is crucial. Investigators and other item library users should carefully document their methods for item selection, including how the literature was reviewed and which decision rules were applied. [30] Statistical and psychometric methods (e.g., differential item functioning) can also help to address the issue by evaluating whether a measure performs similarly across different subgroups.

For multi-arm treatment clinical trials, researchers should ensure transparency by describing how the selected items relate to each of the study arms and which symptomatic AEs are attributed to each of the study regimens. It is important that the same items (or the same item banks for CAT) be included in all treatment arms in order to minimize the potential for bias, avoid underreporting of symptomatic

AEs, and ensure comparability. [38,39,46,47] Researchers should also describe how the selected item list compares to those used in other studies investigating the same treatment regimen.

5. How can unexpected issues be measured?

Free text reporting. The use of free text response options can help to elicit unexpected symptomatic AEs and issues, which may be particularly useful in certain study contexts and populations (e.g., early phase trials). [48] The newly generated issues can then be translated to item(s) that can be included in PROs in future trials of the same intervention, pending appropriate testing. Within the scope of real-world evidence and in the assessment of novel treatments when longer-term follow-up is required, free text response options can help to ensure that important symptomatic AEs and issues are captured, while avoiding potentially lengthy test batteries. [49] In specific populations for whom standard questionnaires must be kept short (e.g., palliative care), the use of free text response options may be particularly relevant. [50]

Predictive text reporting. Studies of some measurement systems, like the PRO-CTCAE, have incorporated drop-down menus using terms from the PRO-CTCAE and the Medical Dictionary for Regulatory Activities (MedDRA) to ensure meaningfulness of concepts and comparability. [47]

Even in PRO-CTCAE studies where unstructured free text entries are used, the majority of these can be mapped onto the PRO-CTCAE and MedDRA terminology, but this does add additional work for the researchers involved. [47] Research on the use of the write-in three symptoms/problems (WISP) has also shown that additional symptoms and problems reported by patients using unstructured free text reporting can be qualitatively coded and summarized. [50]

It is important to note that free text reporting may be more feasible within the context of measures that assess symptomatic AEs alone (versus psychosocial impact and functioning), given that these are easier to map onto standardized medical terminology and frameworks. Moreover, in large-scale international studies, the analysis of data generated from free text responses may be complicated by translation issues and a lack of standardization.

6. How should item lists be ordered?

Grouping similar items and response formats together. As with standard questionnaires, items should be integrated such that similar formats (with matching response/time scales) remain grouped together. [39] Items should generally be grouped within a single HRQOL domain and not intermixed across multiple domains. In many cases, it may be worth considering whether key constructs and issues should be included first to ensure completeness of data.

Controlling for possible priming effects. Items should be ordered in such a way that they avoid influencing subsequent responses. Items which are sensitive in nature (e.g., those capturing sexual

functioning) should generally be placed at the end of a measure in the event that they might upset patients in such a way that subsequent responses could be impacted. [38,39,51]

Preserving psychometric properties. When administered in conjunction with a standard static questionnaire or questionnaires, item lists should be presented in a distinct manner from the former to preserve the psychometric properties of the static measure(s) and clearly distinguish the item lists. [7,38,39]

7. How should appropriate recall periods be selected?

In general, it is recommended that items be administered with the recall periods with which they were developed and validated. In cases where more flexibility is sought and alternative recall periods are selected, it is important to consider research questions and available evidence. It should be noted that the use of alternative response scales/categories is beyond the scope of this paper, given that such modifications alter the items themselves.

It is generally recommended to use recall periods that capture events and symptoms occurring within the last week. Responses are likely to be influenced by the patient's overall state during recall and measures which rely heavily on memory may undermine content validity and reliability. [52] However, in some trial settings, specific symptoms (e.g., pain) may be best measured daily, particularly when these symptoms represent endpoints. [53] Also for clinical monitoring, for example, in patients with acute conditions or undergoing aggressive therapies, capturing daily changes using a 24-hour recall period may be most appropriate. Although longer recall periods (e.g., 2-4 weeks) tend to be associated with increasing rates of recall bias, some domains and types of functioning, especially those that may not be expected to occur daily (e.g., sexual functioning) may be best measured by a longer recall period. [38,39]

Moreover, studies of some PRO measures and measurement systems have found little impact of recall period. [54–57] As such, it is important to consider the available evidence for the specific item library, as well as study design and timing of instrument/item list administration. If investigators need to capture the patient experience over the entire time course of treatment, it is important for frequency of assessment to coincide with recall periods. In general, the choice of recall period depends on several factors, including the measure's intended use, the study's research questions, the schedule of events, [33] and the timing of PRO administration. Depending on the specific outcome of interest (e.g., symptom variability vs. overall assessment of impact), different recall periods may be relevant. [58] While the use of new recall periods may complicate comparability across studies, such an approach may still be necessary in some cases.

8. What are some of the determinants of patient burden and how can it be minimized?

Although length of measures may be linked to patient burden, the issue of burden is more complex than a simple threshold for number of items. When multiple instruments are administered, it is important to avoid duplication of concepts, which may be frustrating to patients. Completion time should also be considered, as longer completion may lead to higher burden. Timing of questionnaire administration is also relevant. For example, patients may be more willing to complete a longer questionnaire if they know that it will not occur frequently. When frequent (e.g., weekly) administration is planned, then measures should generally be relatively short. [20,38,39]

Formatting of the questionnaire, patients' literacy levels, administration mode (e.g., paper, phone, electronic), and sensitive content; may all be linked to burden. [20] Other underlying factors like perceived difficulty of measure(s), lower cognitive functioning and dexterity problems, cognitive demands related to PRO administration, as well as disease stage and severity may all play a role in contributing to burden. However, for relevant issues, patients may specify that additional items are required. [59,60] When patients are assured that their responses provide a meaningful contribution, completion of measures may be perceived as less burdensome. The following approaches are recommended for minimizing patient burden:

Considering PPI. The need to obtain input through PPI should also be considered when assessing possible determinants of burden.

Robust approach to item selection and pilot testing of provisional list. Measures and items should be selected in a thoughtful way, minimizing redundancy and highlighting relevance by focusing on key symptomatic AEs and issues and ensuring meaningfulness for patients. [46] Pilot testing the item list and battery of measures may also help to determine level of burden and feasibility. [38,39]

9. *How should item lists be used in conjunction with static measures and/or other measurement systems?*

Inclusion of core outcomes. Use of a core set of common symptoms, not specific to disease or treatment, has been recommended by various stakeholders when initiating the item selection process during the design phase of a cancer clinical trial.[61] Further work to refine final symptom selection requires consideration of the expected disease- and treatment-related symptoms that are meaningful to patients, adding items not found in the core set and removing items which are not expected to occur or be relevant to the trial context. Moreover, the FDA currently recommends the use of a minimum core outcome set when designing a PRO strategy for clinical trials with regulatory intent, that can be expanded depending on study objectives.[33]

Engagement with regulators and patient groups. It is important to consider clear and early engagement with regulators and patient groups. For example, the FDA and EMA recommend selection of measures that allow for measurement of symptomatic AEs, disease-related symptoms, and physical

functioning as concepts that should be a key focus, although other concepts may also be included where relevant. [23,33,34,40,41,46]

Avoiding duplication of concepts and ensuring relevance of items. If a symptomatic AE is covered within a standard PRO questionnaire used in the trial, it should not additionally be included within a separate item list used in the same trial (unless there is a strong rationale). When patient burden is potentially an issue or when sections of a static questionnaire are clearly not relevant to the target population, items may be removed from a questionnaire (with the instrument developer's consent). However, this should be approached with caution and the resulting measure should be distinguished as an item list and not a full questionnaire. [6,7,62,63] Where relevant, it is recommended that item removal occur at the scale level, to preserve multi-item scales and facilitate scoring and interpretation.

A resource list highlighting key recommendations for specific recommendations is provided in Table 3.

INSERT TABLE 3

DISCUSSION

This work aimed to develop evidence- and discussion-based recommendations on the use of PRO assessment from item libraries in oncology trials. As highlighted in the results, the use of item libraries allows for flexibility in PRO measurement, helping to ensure a patient-centred approach to the assessment of important issues and concepts. With this added flexibility comes the need to ensure robust measurement and minimal bias whenever feasible, preserving the rigour learned from the development of static questionnaires. It is crucial that every investigator account for the development of their item list in a transparent way that builds on the existing evidence base and promotes an objective and comprehensive approach. This helps to avoid possible cherry-picking of items, which could favour some treatment regimens and potentially lead to underreporting of symptomatic AEs and other important HRQOL issues.

The recommendations provide guidance on the use of customized item lists, from item selection through implementation and integration with standard questionnaires. Although it was possible to achieve high-level, instrument-agnostic recommendations for many of the research questions, as described in the results, recommendations may need to be adapted based on the specific context of use and population(s) under investigation. Throughout all stages of item list implementation, it is also critical to consider the role of various stakeholders.

Although the absence of a formal consensus approach (e.g., Delphi exercise) may be viewed as a limitation, given the very high level of these recommendations, the less formal approach which relied

on evidence from the scoping review and discussion among the WG members, was deemed appropriate. Since many of the recommendations depend on the context of use, and the consideration of other additional factors, it was simply not feasible to create very specific recommendations for each research question.

Confirmed participation from various item library developers and stakeholders is a strength of this work, as it helped to ensure balanced perspectives and relevance of recommendations across different measurement systems. While the recommendations described here were developed largely within the framework of controlled clinical trials, most can be extended to other types of trials within oncology (e.g., supportive care and observational). Furthermore, the general principles of these recommendations on item library use and implementation are also applicable outside of oncology trials.

CONCLUSION

These recommendations address a wide range of issues that are relevant for the use of item libraries to assess PROs in oncology trials, with the role of patients and other key stakeholders emphasized throughout.

REFERENCES

1. Chen J, Ou L, Hollis SJ. A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. BMC Health Services Research. 2013.
2. FDA. Guidance for industry: Patient-reported outcome measures: Use in medical product development to support labeling claims: Draft guidance. Health Qual Life Outcomes. 2006.
3. Yang LY, Manhas DS, Howard AF, Olson RA. Patient-reported outcome use in oncology: a systematic review of the impact on patient-clinician communication. Supportive Care in Cancer. 2018.
4. Wilson IB, Cleary PD. Linking Clinical Variables With Health-Related Quality of Life: A Conceptual Model of Patient Outcomes. JAMA J Am Med Assoc. 1995.
5. EORTC. EORTC Item Library [Internet]. [cited 2021 Jun 14]. Available from: https://www.eortc.be/itemlibrary/
6. FACIT. FACIT Searchable Library and Custom Form Developer (Build-a-PRO) [Internet]. Available from: https://wizard.facit.org/
7. MDASI. MDASI Symptom Library [Internet]. Available from: https://www.mdanderson.org/content/dam/mdanderson/documents/Departments-and-Divisions/Symptom-Research/MDASI symptom library.pdf
8. PRO-CTCAE. PRO-CTCAE Instruments & Form Builders [Internet]. [cited 2021 Jun 14]. Available from: https://healthcaredelivery.cancer.gov/pro-ctcae/instrument.html
9. PROMIS. PROMIS Health Measures [Internet]. [cited 2021 Jun 14]. Available from: https://www.healthmeasures.net/explore-measurement-systems/promis?AspxAutoDetectCookieSup=
10. Petersen MA, Aaronson NK, Conroy T, Costantini A, Giesinger JM, Hammerlid E, et al. International validation of the EORTC CAT Core: a new adaptive instrument for measuring core quality of life domains in cancer. Qual Life Res. 2020.
11. Kluetz PG, O'Connor DJ, Soltys K. Incorporating the patient experience into regulatory decision making in the USA, Europe, and Canada. The Lancet Oncology. 2018.
12. Grant MJ, Booth A. A typology of reviews: An analysis of 14 review types and associated methodologies. Health Information and Libraries Journal. 2009.
13. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol. 2018.
14. Brahmer JR, Lacchetti C, Schneider BJ, Atkins MB, Brassil KJ, Caterino JM, et al. Management of immune-related adverse events in patients treated with immune checkpoint inhibitor therapy: American society of clinical oncology clinical practice guideline. J Clin Oncol. 2018.
15. DeWalt DA, Rothrock N, Yount S, Stone AA. Evaluation of item candidates: The PROMIS qualitative item review. Med Care. 2007.
16. Williams LA, Whisenant MS, Mendoza TR, Haq S, Keating KN, Cuffel B, et al. Modification of existing patient-reported outcome measures: qualitative development of the MD Anderson Symptom Inventory for malignant pleural mesothelioma (MDASI-MPM). Qual Life Res. 2018.

17. PROMIS. PROMIS Measure Development & Research [Internet]. Available from: https://www.healthmeasures.net/explore-measurement-systems/promis/measure-development-research

18. Klem M, Saghafi E, Abromitis R, Stover A, Dew MA, Pilkonis P. Building PROMIS item banks: Librarians as co-investigators. Qual Life Res. 2009.

19. Calvert MJ, Cruz Rivera S, Retzer A, Hughes SE, Campbell L, Molony-Oates B, et al. Patient reported outcome assessment must be inclusive and equitable. Nat Med. 2022.

20. FDA. Guidance for Industry; Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Clin Fed Regist. 2009.

21. Johnson C, Aaronson N, Blazeby JM, Bottomley A, Fayers P, Koller M, et al. EORTC Quality of Life Group Guidelines for Developing Questionnaire Modules. 2011;4th Ed.(April).

22. Bjordal K, Bottomley A, Gilbert A, Martinelli F, Pe M, Sztankay M, et al. EORTC Quality of Life Group Module Development Guidelines [Internet]. 5th ed. Brussels, Belgium: EORTC; 2021. Available from: https://qol.eortc.org/manuals/

23. Turner G, Aiyegbusi OL, Price G, Skrybant M, Calvert M. Moving beyond project-specific patient and public involvement in research. Journal of the Royal Society of Medicine. 2020.

24. Cruz Rivera S, Stephens R, Mercieca-Bebber R, Retzer A, Rutherford C, Price G, et al. "Give Us the Tools!": development of knowledge transfer tools to support the involvement of patient partners in the development of clinical trial protocols with patient-reported outcomes (PROs), in accordance with SPIRIT-PRO Extension. BMJ Open. 2021.

25. Haywood K, Lyddiatt A, Brace-McDonnell SJ, Staniszewska S, Salek S. Establishing the values for patient engagement (PE) in health-related quality of life (HRQoL) research: an international, multiple-stakeholder perspective. Qual Life Res. 2017.

26. FDA. CDER Patient-Focused Drug Development [Internet]. [cited 2021 Dec 2]. Available from: https://www.fda.gov/drugs/development-approval-process-drugs/cder-patient-focused-drug-development

27. FDA. FDA-led Patient-Focused Drug Development (PFDD) Public Meetings [Internet]. [cited 2021 Dec 2]. Available from: https://www.fda.gov/industry/prescription-drug-user-fee-amendments/fda-led-patient-focused-drug-development-pfdd-public-meetings

28. European Medicines Agency & Committee for Human Medicinal Products. ICH guideline E8 (R1) on general considerations for clinical studies. Ema/Chmp/Ich/544570/1998 [Internet]. 2019;8(October 2021):1–25. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-8-general-considerations-clinical-trials-step-5_en.pdf

29. FDA. FDA Project Patient Voice [Internet]. Available from: https://www.fda.gov/about-fda/oncology-center-excellence/project-patient-voice

30. Retzer A, Aiyegbusi OL, Rowe A, Newsome PN, Douglas-Pugh J, Khan S, et al. The value of patient-reported outcomes in early-phase clinical trials. Nature Medicine. 2022.

31. Trask PC, Dueck AC, Piault E, Campbell A. Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events: Methods for item selection in industry-sponsored oncology clinical trials. Clin Trials. 2018;15(6):616–23.

32. Shepshelovich D, McDonald K, Spreafico A, Razak ARA, Bedard PL, Siu LL, et al. Feasibility Assessment of Using the Complete Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) Item Library. Oncologist. 2019.

33. FDA. Core Patient-Reported Outcomes in Cancer Clinical Trials Guidance for Industry DRAFT GUIDANCE. 2021;(June 2021). Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/core-patient-reported-outcomes-cancer-clinical-trials

34. European Medicines Agency Committee for Medicinal Products for Human Use (CHMP). Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man: The use of patient-reported outcome (PRO) measures in oncology studies [Internet]. 2016 [cited 2021 Nov 4]. Available from: https://www.ema.europa.eu/en/documents/other/appendix-2-guideline-evaluation-anticancer-medicinal-products-man_en.pdf

35. Diamantopoulos A, Sarstedt M, Fuchs C, Wilczynski P, Kaiser S. Guidelines for choosing between multi-item and single-item scales for construct measurement: A predictive validity perspective. J Acad Mark Sci. 2012.

36. Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR, Young SL. Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. Frontiers in Public Health. 2018.

37. Fayers PM, Machin D. Quality of Life: The Assessment, Analysis and Interpretation of Patient-reported Outcomes. 3rd Ed. John Wiley & Sons; 2016.

38. Piccinin C, Kulis D, Bottomley A, Bjordal K, Coens C, Darlington AS, et al. PCN296 Development of scientific guidelines for use of the EORTC Item Library in cancer clinical trials. Value Heal. 2019.

39. Piccinin C, Kuliś D, Bottomley A, Bjordal K, Coens C, Darlington AS, et al. EORTC Quality of Life Group Item Library User Guidelines. 1st Ed. Brussels: EORTC; 2022.

40. Kluetz PG, Slagle A, Papadopoulos EJ, Johnson LL, Donoghue M, Kwitkowski VE, et al. Focusing on core patient-reported outcomes in cancer clinical trials: Symptomatic adverse events, physical function, and disease-related symptoms. Clin Cancer Res. 2016.

41. Kluetz PG, Papadopoulos EJ, Johnson LL, Donoghue M, Kwitkowski VE, Chen WH, et al. Focusing on core patient-reported outcomes in cancer clinical trials - Response. Clinical Cancer Research. 2016.

42. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Quality of Life Research. 2018.

43. Nunnally JC, Bernstein IH. Psychometric Theory. 3rd Ed. New York: McGraw-Hill; 1994.

44. Portney LG, Watkins MP. Foundations of Clinical Research: Applications to Practice. 3rd Ed. Upper Saddle River, New Jersey: Pearson/Prentice Hall; 2009.

45. Dueck AC, Mendoza TR, Mitchell SA, Reeve BB, Castro KM, Rogak LJ, et al. Validity and reliability of the us national cancer institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). JAMA Oncol. 2015;1(8):1051–9.

46. Basch E, Rogak LJ, Dueck AC. Methods for Implementing and Reporting Patient-reported Outcome (PRO) Measures of Symptomatic Adverse Events in Cancer Clinical Trials. Clin Ther. 2016.

47. Chung AE, Shoenbill K, Mitchell SA, Dueck AC, Schrag D, Bruner DW, et al. Patient free text reporting of symptomatic adverse events in cancer clinical research using the National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). J Am Med Informatics Assoc. 2019.

48. FDA. Patient-Focused Drug Development: Methods to Identify What Is Important to Patients [Internet]. 2022. Available from: https://www.fda.gov/drugs/development-approval-process-drugs/fda-patient-focused-drug-development-guidance-series-enhancing-incorporation-patients-voice-medical

49. Calvert MJ, O'Connor DJ, Basch EM. Harnessing the patient voice in real-world evidence: the essential role of patient-reported outcomes. Nat Rev Drug Discov. 2019;18(10):731–2.

50. Rojas-Concha L, Hansen MB, Petersen MA, Groenvold M. Which symptoms and problems do advanced cancer patients admitted to specialized palliative care report in addition to those included in the EORTC QLQ-C15-PAL? A register-based national study. Support Care Cancer. 2020.

51. Claessen FMAP, Mellema JJ, Stoop N, Lubberts B, Ring D, Poolman RW. Influence of Priming on Patient-Reported Outcome Measures: A Randomized Controlled Trial. Psychosomatics. 2016.

52. Mendoza TR, Dueck AC, Bennett A V., Mitchell SA, Reeve BB, Atkinson TM, et al. Evaluation of different recall periods for the US National Cancer Institute's PRO-CTCAE. Clin Trials. 2017.

53. Basch EM, Scholz M, de Bono JS, Vogelzang N, de Souza P, Marx G, et al. Cabozantinib Versus Mitoxantrone-prednisone in Symptomatic Metastatic Castration-resistant Prostate Cancer: A Randomized Phase 3 Trial with a Primary Pain Endpoint. Eur Urol. 2019.

54. Lai JS, Cook K, Stone A, Beaumont J, Cella D. Classical test theory and item response theory/Rasch model to assess differences between patient-reported fatigue using 7-day and 4-week recall periods. J Clin Epidemiol. 2009.

55. Thavarajah N, Bedard G, Zhang L, Cella D, Beaumont JL, Tsao M, et al. The Functional Assessment of Cancer Therapy - Brain (FACT-Br) for assessing quality of life in patients with brain metastases: A comparison of recall periods. J Pain Manag. 2013.

56. Condon DM, Chapman R, Shaunfield S, Kallen MA, Beaumont JL, Eek D, et al. Does recall period matter? Comparing PROMIS® physical function with no recall, 24-hr recall, and 7-day recall. Qual Life Res. 2020.

57. Peipert JD, Chapman R, Shaunfield S, Kallen MA, Schalet BD, Cella D. Do You Recall?: Results From a Within-Person Recall Study of the Patient-Reported Outcomes Measurement Information System (PROMIS) Short Form v2.0 – Physical Function 8c. Value Heal. 2022.

58. Stull DE, Leidy NK, Parasuraman B, Chassany O. Optimal recall periods for patient-reported outcomes: Challenges and potential solutions. Current Medical Research and Opinion. 2009.

59. Rolstad S, Adler J, Rydén A. Response burden and questionnaire length: Is shorter better? A review and meta-analysis. Value Heal. 2011.

60. Atkinson TM, Schwartz CE, Goldstein L, Garcia I, Storfer DF, Li Y, et al. Perceptions of Response Burden Associated with Completion of Patient-Reported Outcome Assessments in Oncology. Value Heal. 2019.

61. Reeve BB, Mitchell SA, Dueck AC, Basch E, Cella D, Reilly CM, et al. Recommended patient-reported core set of symptoms to measure in adult cancer treatment trials. J Natl Cancer Inst. 2014.

62. Groenvold M, Aaronson NK, Darlington ASE, Fitzsimmons D, Greimel E, Holzner B, et al. Focusing on core patient-reported outcomes in cancer clinical trials - Letter. Clinical Cancer Research. 2016.

63. Kulis D, Piccinin C, Bottomley A, Darlington AS, Grønvold M. PCN45 What to choose?

creating a patient-reported outcomes (PRO) assessment strategy with EORTC measures and Item Library that can meet the needs of regulators. Value Heal. 2019.

CONTRIBUTORS

All authors contributed to the conceptualization of this work, along with data interpretation and writing (review & editing). CP and AB were responsible for data collection/extraction and data analysis. CP wrote the original draft of the manuscript, and the final version was approved by all authors.

DECLARATION OF INTERESTS

EB reports receiving personal consulting fees (as consultant/scientific advisor) from AstraZeneca, Carevive Systems, Navigating Cancer, Sivan Healthcare, and Resilience Health. MC has received funds for her institution (University of Birmingham, UK) from the NIHR Birmingham Biomedical Research Centre, NIHR Surgical Reconstruction and Microbiology Research Centre, NIHR Birmingham-Oxford Blood and Transplant Research Unit (BTRU) in Precision Transplant and Cellular Therapeutics, NIHR ARC West Midlands at the University of Birmingham and University Hospitals Birmingham NHS Foundation Trust, Health Data Research UK, Innovate UK (part of UK Research and Innovation), Macmillan Cancer Support, SPINE UK,UKRI, UCB Pharma, Janssen, GSK, Gilead; reports personal consulting fees from Aparito Ltd, CIS Oncology, Takeda, Merck, Daiichi Sankyo, Gilead, Glaukos, GSK, the Patient-Centered Outcomes Research Institute (PCORI); has a family member who owns stock in GSK; and is Director of the Birmingham Health Partners Centre for Regulatory Science and Innovation, Director of the Centre for the Centre for Patient Reported Outcomes Research, and is a National Institute for Health Research (NIHR) Senior Investigator. AC is employed by AstraZeneca and reports grants from Carevive, Takeda, Clinical Outcomes Solutions, and LUNGevity. DC reports receiving royalties or licenses as President of FACIT.org and as President-Elect and Board Member of PROMIS Health Organization. CSC reports receiving licensing fees paid to both MD Anderson Cancer Center and his SAS, LLC for Brief Pain Inventory. BLKK has received grants from AstraZeneca, G1 Therapeutics, Bristol-Myers Squibb, Merck, BluePrint Medicine, Eli Lilly, Genentech, Takeda, Jazz Pharmaceuticals; consulting fees from Eli Lilly, Health Outcomes Solutions, University of South Florida, Atheneum; and has participated on a Data Safety Monitoring Board or Advisory Board for Bristol-Myers Squibb. KO reports receiving honoraria (2010 to 2020) from GSK for her involvement with the Healthcare Advisory Board and receiving an honorarium as a speaker at the Sharing Progress in Cancer Care webinar (October 2021). All other authors declare no competing interests.

DISCLAIMER

This publication reflects the views of the individual authors and should not be construed to represent official views or policies of the US Food and Drug Administration, US National Cancer Institute, Medicines and Healthcare products Regulatory Agency, the UK National Health Service, the National Institute for Health Research, the UK Department of Health and Social Care, or the European Medicines Agency. The views expressed in this article are the personal views of the author(s) and may not be understood or quoted as being made on behalf of or reflecting the position of the regulatory agency/agencies or organisations with which the author(s) is/are employed/affiliated.

This publication has not been submitted to another journal or published in whole or in part elsewhere previously.

No authors are employed by NIH.

APPENDICES

Table 1. Research questions to guide recommendations

Research Question/Topic	**Specification**
Which methods should be used to drive item selection?	In general (irrespective of study phase)
	Based on clinical trial phase
When should single items vs. multi-item scales be used and what are the benefits and limitations of each approach?	Use of single items vs. multi-item scales
How should different types of psychometric properties be considered and tested, based on the item list/measure and the context of its use?	Single items & multi-item scales - validity
	Single items & multi-item scales - reliability
	Responsiveness to change
How can bias be minimized in the design of item lists?	In general
	For use in multi-arm clinical trials
How can unexpected issues be measured by item lists?	Using free text and predictive text reporting
How should item lists be ordered?	To ensure comprehensibility
	To account for possible priming effects and potentially sensitive issues
	To preserve psychometric properties, where relevant
How should appropriate recall periods be selected?	In general
	Considering PRO and study/clinical characteristics
What are some of the determinants of patient burden and how can it be minimized?	Determinants of patient burden
	Methods to minimize patient burden
How should item lists be used in conjunction with static measures and/or other measurement systems?	To assure measurement of core outcomes
	To achieve a flexible and balanced approach to PRO measurement

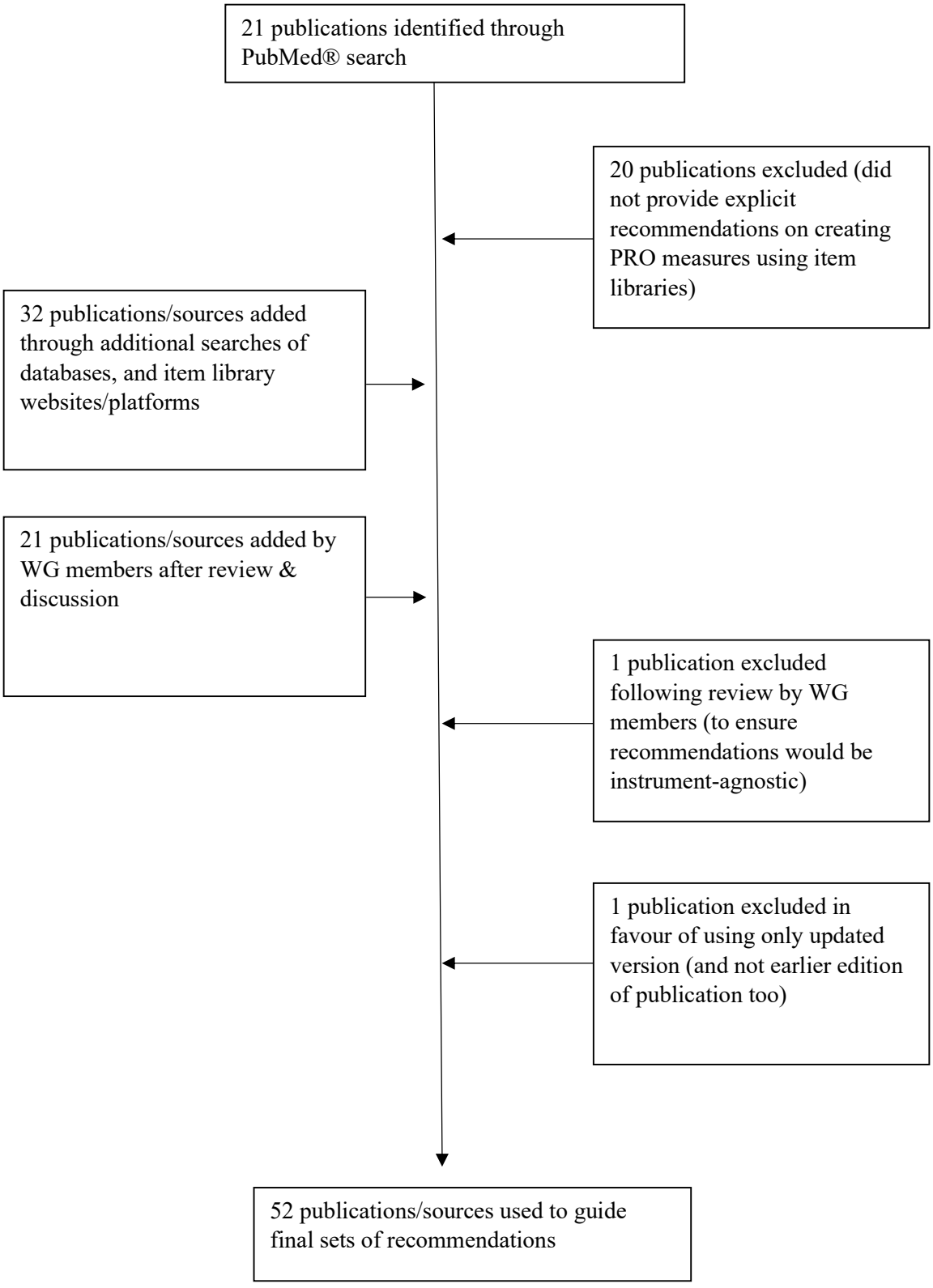

Figure 1. Flowchart of sources included in recommendations

Table 2. Psychometric properties for consideration in item list development

<table>
<tr><th colspan="4">Single items and multi-item scales - validity</th></tr>
<tr><td colspan="4">Content validity</td></tr>
<tr><td colspan="4">Patient-centred approaches (e.g., interviews and focus groups) can help to ensure inclusion and measurement of meaningful concepts. [20–22] It is important to establish content validity before evaluating other measurement properties (e.g., construct validity; reliability), since evidence of other types of validity cannot overcome issues related to content validity. COSMIN has developed criteria and a checklist which can be used to evaluate content validity in PRO measures. [42]</td></tr>
<tr><td colspan="4">Construct validity</td></tr>
<tr><td>• Construct validity should be assessed by comparing results from the new measure with existing instruments and outcomes (e.g., other questionnaires, clinician reports, clinical data) which can serve as anchors to evaluate whether results are consistent with established relationships (i.e., convergent and discriminant validity).</td><td>• Convergent validity assesses whether a PRO measure is correlated with a similar measure (i.e., of the same or similar construct), using correlation coefficients.</td><td>• Known groups validity assesses the extent to which the PRO measure can distinguish between different groups known to differ on the domain of interest.</td><td>• Structural validity (multi-item scales only) confirms that the items that make up a multi-item scale are associated with each other in a way that confirms the dimensionality of the domain(s) being assessed. Typically factor analytic methods are used to evaluate structural validity. [43]</td></tr>
<tr><td colspan="4">Criterion validity</td></tr>
<tr><td colspan="4">Criterion validity can be evaluated by comparing the measure to a known gold standard measure of the same concept, but it is rare that this is applicable to PROs, since most concepts measured using PROs would not have a gold standard equivalent.</td></tr>
<tr><th colspan="4">Single items and multi-item scales – reliability</th></tr>
<tr><td colspan="4">Test-retest reliability / stability</td></tr>
<tr><td colspan="4">Test-retest reliability or stability can be assessed using intra-class correlation coefficients (ICC) between assessments. Although there is some debate surrounding the issue, correlations of at least 0.70 are generally considered acceptable, while those exceeding 0.80 are "good". [43,44] If the measure is intended to be used for individual patient monitoring, a higher correlation would be recommended.</td></tr>
<tr><td colspan="4">Internal consistency (multi-item scales only)</td></tr>
<tr><td colspan="4">Cronbach's Coefficient alpha, along with item-total correlations, can be used to assess internal consistency.</td></tr>
<tr><td colspan="4">Item response theory (IRT) (multi-item scales only)</td></tr>
<tr><td colspan="4">IRT models allow a comprehensive evaluation of how well (in terms of information or standard error of measurement) the set of items within a scale captures the full range of HRQOL levels observed in the study sample.</td></tr>
<tr><td colspan="4">Skewness / floor and ceiling effects</td></tr>
<tr><td colspan="4">Possible skewness and floor/ceiling effects can be evaluated by assessing the distribution of scores.</td></tr>
<tr><th colspan="4">Responsiveness to change</th></tr>
<tr><td colspan="4">Comparison with criterion parameters</td></tr>
<tr><td colspan="4">Changes in PRO scores can be compared to changes in other similar measures (e.g., criterion parameters like performance status) that provide evidence that the PRO changes relate to the concept being investigated. [20–22,45]</td></tr>
<tr><td colspan="4">Comparison at different time points</td></tr>
<tr><td colspan="4">Changes in PRO scores can also be compared at different time points throughout the course of disease/treatment. [20–22,45]</td></tr>
</table>

Table 3. Key recommendations on the use of item libraries in oncology - resource list

<table>
<tr><td rowspan="2">Methods to drive item selection</td><td>Systematic literature review</td><td colspan="3">Interviews and focus groups</td><td>Patient and public involvement</td><td>Retrospective chart reviews</td><td colspan="3">Preclinical data</td></tr>
<tr><td></td><td>Patient interviews</td><td>HCP interviews</td><td>Focus groups</td><td></td><td></td><td colspan="3"></td></tr>
<tr><td rowspan="3">Psychometric tests to assess validity</td><td colspan="2" rowspan="2">Content validity</td><td colspan="6">Construct validity</td><td rowspan="2">Criterion validity</td></tr>
<tr><td rowspan="2">Anchor-based comparison</td><td>Convergent validity</td><td>Known groups validity</td><td colspan="3">Structural validity</td></tr>
<tr><td>Patient-centred approaches</td><td>COSMIN checklist</td><td>Correlation coefficients</td><td>Comparison with clinically different groups</td><td>Comparison with known gold standard measure of same concept</td><td>Exploratory factor analysis</td><td>Confirmatory factor analysis</td><td>Comparison with known gold standard measure of same concept</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td rowspan="3">Psychometric tests to assess reliability & responsiveness to change</td><td>Test-retest reliability</td><td>Skewness / floor and ceiling effects</td><td colspan="2">Internal consistency</td><td colspan="3">Aggregated data in summated scales</td><td colspan="2">Responsiveness to change</td></tr>
<tr><td>ICC</td><td>Distribution of scores</td><td>Cronbach’s alpha (for composite scores)</td><td>Item-total correlations</td><td>IRT</td><td>Factor analysis</td><td>Consultation with content experts</td><td colspan="2">Anchor-based methods</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="2"></td></tr>
<tr><td rowspan="2">Methods to measure unexpected issues & symptoms</td><td colspan="9">Free text reporting</td></tr>
<tr><td colspan="9"></td></tr>
</table>

Issues to consider when ordering item lists	Grouping similar items & formats together	Priming effects	Listing items with sensitive issues last	Preserving format and psychometric properties of standard questionnaires
Methods to minimize patient burden	Robust, patient-centred approach to item selection	Avoiding duplication of concepts	Focusing on key issues of relevance	Pilot testing provisional item list
Factors to consider when using item lists in conjunction with other questionnaires	Inclusion of core outcomes	Communication with regulators and patient groups	Avoiding duplication of concepts	Ensuring relevance of items