Using patient-reported outcome measures in clinical trials: perspectives for and against a modular approach

Carrie-Anne Ng[1]
Jessica Roydhouse[2]
Tim Luckett[3]
Richard De Abreu Lourenco[1]
Brendan Mulhern[1]

1 Centre for Health Economics Research and Evaluation, University of Technology Sydney, Sydney, NSW.
2 Menzies Institute for Medical Research, University of Tasmania, Hobart, TAS.
3 Improving Palliative, Aged and Chronic Care through Clinical Research and Translation (IMPACCT), University of Technology Sydney, Sydney, NSW.

brendan.mulhern@uts.edu.au

doi: 10.5694/mja2.52629

There is growing acknowledgement among international stakeholders, including regulators, health technology assessment (HTA) bodies and professional societies, that the inclusion of patient-reported outcome measures (PROMs) in clinical trials should be guided by clear rationales.[1] In this regard, existing PROMs may overreach or fall short in measuring domains most relevant for a specific study's context, population, treatments and stakeholders.

In response to the need for flexibility in assessing specific health-related quality of life (HRQoL) domains and regulatory recommendations,[1] PROM developers and stakeholders have provided guidance on a modular approach.[2-4] Although most guidance has focused on cancer clinical trials, the same principles are applicable to any context where PROMs are used (eg, clinical practice). In trials, the modular approach has been defined as the selection and assessment of specific patient-relevant and clinically relevant domains of interest, purposefully selected from multidomain PROMs, and then independently scored and interpreted.[3,4]

Selection of the most relevant disease- and treatment-specific domains can be based on previous research, consumer advice and clinical experience. Stakeholders have also recommended using conceptual frameworks and core patient-reported outcome (PROs) sets to guide the selection of PROM domains.[5] HRQoL concepts identified as primary or key secondary outcomes should ideally be assessed using the most valid, reliable and dedicated measures, which are often more in-depth than domains in multidomain PROMs used to provide a broad overview. For example, if pain is a primary outcome in a trial investigating an anti-cancer treatment, the Brief Pain Inventory might be a comprehensive substitution for the two-item domain of pain from the widely used cancer-specific HRQoL PROM, the European Organisation for Research and Treatment of Cancer Quality of Life Core Questionnaire (EORTC QLQ-C30).[6] Secondary outcomes, such as appetite loss and constipation, might be adequately assessed by the QLQ-C30's single-item domains for these outcomes. Thus, we propose three possible applications of the modular approach (Box 1): (i) using a study-specific conceptual framework, incorporating dedicated PROMs or domains to measure relevant HRQoL concepts; (ii) using a full-length PROM, removing domains less relevant to the study context or not covered by a core PRO set; or (iii) using a full-length PROM, substituting domains that are primary or key secondary outcomes with dedicated PROMs or domains.

Box 2 presents the cases for and against the use of a modular approach. In this perspective article, we discuss some of these arguments in the context of assessing clinical efficacy and conducting economic evaluations.

The case for the modular approach

A modular approach can address respondent burden associated with lengthy PROMs and the potential for missing or poor quality PRO data in clinical trials[7] (Box 3, Example 1). However, the aim of a modular approach is to prioritise assessment of the most important domains in each context, rather than to reduce PROM length. Indeed, patients may be willing to complete longer measures if items are relevant to their experience.[10] Additionally, administering domains relevant to the clinical trial context can improve sensitivity to clinically important changes in PROs (Box 3, Example 2).

The modular approach also enables flexibility in assessing different domains at various timepoints. Domains that are conceptually proximal to the disease and treatment (eg, nausea severity) are more likely to change over a short period of time than more distal domains (eg, emotional function), and may therefore require more frequent assessment.[11]

To preserve the psychometric properties of domains resulting from a modular approach, working with PROM developers and following published recommendations is encouraged.[2] If selected domains are obtained from PROMs validated in the target population, they are likely to retain psychometric properties, such as content and construct validity. As additional psychometric testing may not be required, this can be beneficial for novel treatments, given faster evaluation and approval times,[12] or for patients with rare cancers, where developing or validating bespoke disease-specific PROMs may be challenging.

The case against the modular approach

The assessment of a broad set of items promotes the integration of PROMs into early-phase trials (phase 1 or phase 2) for capturing potential symptomatic adverse events and HRQoL problems. However, once these issues are identified, investigators must decide which domains to assess in late-phase trials, and to what extent. It is unsurprising that many may default to using full-length PROMs to maintain comparability with existing studies, avoid missing unexpected effects and meet HTA agencies' requirements for economic evaluation.

To complicate matters, the modular approach necessitates greater scrutiny in the balance between the depth and breadth of issues and respondent burden (ie, time taken). Although this choice can be aided by international guidance and patient and public involvement,[13,14] it becomes less clear

1 Suggested applications of the modular approach for patient-reported outcome measures (PROM) inclusion in clinical trials

	Example of application	Specific considerations
Context	Dr Smith was developing a randomised phase 2 study of drug X + Y versus drug X alone for the management of symptoms in a population with disease Z. The examples below represent three hypothetical scenarios during the development of a patient-reported outcome strategy.	
Approach A: using a study-specific conceptual framework, incorporating dedicated patient-reported outcome measures (PROMs) and/or domains to measure relevant health-related quality of life (HRQoL) concepts	Dr Smith identified patient-reported concepts that are relevant and important to patients with disease Z, and which are expected to be affected by drug X +/- Y. This identified ten key concepts of interest. After reviewing several candidate PROMs, no suitable, validated PROM that covered all ten concepts was identified. As a result, the team used: • PROM A-1 (six subscales assessing six key concepts) and • four multi-item domains from PROM A-2 (PROM A-2 consists of 13 subscales – three of which overlap with PROM A-1, six assess concepts that are not relevant to the study, and four assess the remaining four concepts).	Match the concepts of interest with the selected PROM domains and respective items to ensure they are conceptually similar.
Approach B: using a full-length PROM, removing domains less relevant to the study context	Dr Smith planned to use PROM B in his study. PROM B is disease-Z-specific and consists of five multi-item domains: fatigue, pain, body image, weight change and musculoskeletal symptoms. However, the literature indicates that musculoskeletal symptoms are only relevant in patients with advanced disease Z. As the study includes only patients with early-stage disease Z, the team decided to use PROM B, excluding the musculoskeletal symptoms subscale.	Consider if the subscale that was removed covers a concept that may arise at a different timepoint during the trial. Administering this subscale at specific timepoint(s) may be required.
Approach C: using a full-length PROM, substituting domains that are primary or key secondary outcomes with dedicated PROMs or domains.	Dr Smith planned to use PROM C-1 in his study. PROM C-1 is disease-Z-specific and consists of ten domains, all of which cover the concepts of interest that are important to his study. Fatigue is a common and important symptom in patients with disease Z, and likely to be most improved by drug X + Y. However, the fatigue subscale of PROM C-1 has only two items, and Dr Smith is concerned that this may not be sensitive to meaningful changes. Thus, the team decided to use PROM C-2, which is an eight-item fatigue-specific instrument, to assess fatigue. To avoid duplication, PROM C-1 was used without its two-item fatigue subscale.	Evidence showing that dedicated PROM/domain psychometrically performs better than the substituted domain may help to the strengthen the rationale for substitution.

2 Summary of the cases for and against use of the modular approach for patient-reported outcome measures (PROMs) in clinical trials

Case for the modular approach	Case against the modular approach
Promotion of scientific rigor	
• Investigators are required to justify the choice of domains, such as through a conceptual framework with specific hypotheses, rather than "trawling" for effects. • Enables novel combinations of patient-reported outcome measures (PROMs) and domains, including more thorough assessment of primary patient-reported outcomes (PROs) compared with secondary PROs of less interest.	• Ensuring scientific rigor requires greater time and effort to select and administer domains relevant to a clinical trial context. • Unable to record unexpected effects that may otherwise be captured by full-length PROMs. • Potential for bias in selecting domains to favour specific effects while minimising others.
Respondent burden	
• Measurement is focused on the most clinically relevant domains from multidomain PROMs, reducing respondent burden by removing less relevant subscales. • Reduced potential for unnecessary duplication of domains when using multiple PROMs.	• Demands certainty and justification that particular domains are irrelevant and can therefore be excluded.
Psychometric/measurement properties	
• Selected domains retain many established psychometric properties, if the domains were obtained from PROMs that have already been validated in a target context.	• Item order effects may impact psychometric performance of domains when administered separately from the full-length PROM.
Trade-off between flexibility and comparability	
• Flexibility in substituting less informative domains for other PROMs or domains dedicated to measuring specific quality of life concepts. • Domains of interest that are not available in a PROM can be appended. • Flexibility in administering different domains at timepoints when they can capture the most meaningful effects.	• Limited comparability with other studies that have used full-length PROMs. • Acceptability of the modular approach by health technology assessment agencies and other key stakeholders (eg, the Pharmaceutical Benefits Advisory Committee) is unclear.

3 Examples of how a modular approach can improve patient-reported outcome data quality

Example 1:

The Expanded Prostate Cancer Index Short Form (EPIC-26) was developed based on a comprehensive set of symptoms associated with various treatment types for men with localised prostate cancer.

However, a content analysis of the EPIC-26 revealed significant missing data in the urinary function and urinary bother domains for patients using urinary catheters, as the response options were not relevant.[8] Excluding these subscales for use in this patient population would improve the instrument's precision while enhancing the relevance of its items.

Example 2:

An analysis of the United States Food and Drug Administration-registered trials for new immune checkpoint inhibitors found that patient-reported outcome (PRO) strategies did not assess all eight adverse events unique to these novel immunotherapy agents.[9] Commonly used cancer-specific patient-reported outcome measures, such as the European Organisation for Research and Treatment of Cancer Quality of Life Core Questionnaire (EORTC QLQ-C30), are supplemented by relevant PRO domains (eg, assessing rash and itching) to better inform treatment toxicity while minimising the potential for biased assessment of the patient experience.

for domains amidst an abundance of options from various PROMs for the same HRQoL concept. Studies comparing the psychometric properties of PROMs in various study contexts can assist in this selection[15] but, traditionally, conclusions have been drawn by assessing full-length PROMs instead of their domains. Psychometric evidence for specific combinations of subscales and items obtained from the EORTC suite are emerging,[16] but evidence for the combination of subscales across different PROMs is lacking. This raises several questions requiring further conceptual and empirical consideration: Should only domains that have demonstrated the best psychometric evidence be selected and combined? How will a mixture of recall periods and response options across PROMs impact face validity and response distributions? What strategies can avoid biased selection of domains that may favour certain effects while downplaying others?

Guidance on the modular approach suggests that item order can affect the validity of subscales because responses may be influenced by the preceding questions.[2,4] These potential biases are usually addressed during early development of a PROM, which supports the argument for keeping PROMs in their original format unless there is potential for retesting. However, item order effects are mostly theoretical, with limited research available that has tested these directly.[17] Order effects have primarily been considered at the questionnaire level and conclusions regarding their presence have only been conservatively applied to the PROMs assessed.[17]

The use of the modular approach in economic evaluation

The evolution of PROM design should progress alongside other fields that rely on PROMs. In economic evaluations, preference-based measures (PBMs) are commonly used to acquire utility values for calculating quality-adjusted life years in cost–utility analyses. PBMs can be derived from PROMs using a subset of items from the existing measure (eg, the Short Form-6 Dimension (version 2) [SF-6Dv2] from the generic 36 Item Short Form Survey (version 2) [SF-36v2];[18] the EORTC Quality of Life Utility – Core 10 Dimensions [QLU-C10D] from the EORTC QLQ-C30[19]).

A key consideration in applying a modular approach is that the resulting measure may lack the full subset of items necessary for estimating utility values for economic evaluations. For instance, in response to the United States Food and Drug Administration's recommendation of a core PRO set, the EORTC developed the Core Function Questionnaire QLQ-F17, which includes only the functional domains of the QLQ-C30[20] (ie, our proposed Approach B in Box 1). However, the QLQ-F17 does not enable estimation of utility values using the QLU-C10D algorithm.

It is uncertain if the same considerations and recommendations of the modular approach used to determine clinical efficacy can be applied to PBMs. Similar to domains in the modular approach, derived PBMs can be independently scored, interpreted and validated. Improving the flexibility of PBMs has also been explored by administering derived PBMs as standalone measures. This may reduce respondent burden when the full-length PROM adds little additional information, and when the derived PBM has been shown to be valid in the population of interest. For instance, the SF-6Dv2 can either be derived from responses to the SF-36v2, or administered as standalone measures.[21] These approaches have yielded different utility values in patients with breast cancer, with evidence supporting the use of the standalone versions for enhanced sensitivity and discriminative power.[22,23] However, derived PBMs are typically not designed as standalone measures, nor do they serve the same purpose as modules intended to improve clinical relevance.

Future discussions of the modular approach should strive for further consensus on the definition, ensuring it is both clinically relevant and suitable for economic evaluations. Additionally, it is unclear if these novel approaches to PROM administration will be acceptable for economic evaluations submitted to HTA agencies (eg, the Pharmaceutical Benefits Advisory Committee). A key challenge for these HTA agencies will be balancing comparability across different health conditions, relevance to patient populations, and the validity and responsiveness of PBMs.

Conclusion

The modular approach involves replacing existing full-length PROMs with a focused selection of HRQoL domains most relevant to a particular clinical trial context. Several considerations must be addressed before integrating the modular approach into clinical trials, including achieving broader acceptance from regulatory bodies (eg, HTA agencies). These considerations include expanding and clarifying the definition of what constitutes a module and how to apply the modular approach, while ensuring its

suitability for economic evaluations. Further evidence that alleviates concerns associated with a modular approach, particularly evidence supporting its psychometric validity, would improve its acceptability in the field.

Open access: Open access publishing facilitated by University of Technology Sydney, as part of the Wiley – University of Technology Sydney agreement via the Council of Australian University Librarians.

Competing interests: No relevant disclosures.

Provenance: Not commissioned; externally peer reviewed. ■

1 US Food and Drug Administration. Principles for selecting, developing, modifying, and adapting patient reported outcome instruments for use in medical device evaluation: Guidance for Industry and Food and Drug Administration Staff, And Other Stakeholders. FDA, 2022. https://www.fda.gov/media/141565/download (viewed Apr 2024).

2 Piccinin C, Basch E, Bhatnagar V, et al. Recommendations on the use of item libraries for patient-reported outcome measurement in oncology trials: findings from an international, multidisciplinary working group. *Lancet Oncol* 2023; 24: e86-e95.

3 Brady KJS, Peipert JD, Atkinson TM, et al. International Society for Quality of Life Research commentary on the US Food and Drug Administration draft guidance for industry on core patient-reported outcomes in cancer clinical trials. *Qual Life Res* 2023; 32: 2155-2163.

4 Serrano D, Cella D, Husereau D, et al. Administering selected subscales of patient-reported outcome questionnaires to reduce patient burden and increase relevance: a position statement on a modular approach. *Qual Life Res* 2024; 33: 1-10.

5 US Food and Drug Administration. Core patient-reported outcomes in cancer clinical trials: Guidance for Industry. FDA, 2024. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/core-patient-reported-outcomes-cancer-clinical-trials (viewed May 2024).

6 Diamantopoulos A, Sarstedt M, Fuchs C, et al. Guidelines for choosing between multi-item and single-item scales for construct measurement: a predictive validity perspective. *J Acad Mark Sci* 2012; 40: 434-449.

7 Aiyegbusi OL, Roydhouse J, Rivera SC, et al. Key considerations to reduce or address respondent burden in patient-reported outcome (PRO) data collection. *Nat Commun* 2022; 13: 6026.

8 Atkinson TM, Schwartz CE, Goldstein L, et al. Perceptions of response burden associated with completion of patient-reported outcome assessments in oncology. *Value Health* 2019; 22: 225-230.

9 King-Kallimanis BL, Bhatnagar V, Horodniceanu EG, et al. Timing is everything: the importance of patient-reported outcome assessment frequency when characterizing symptomatic adverse events. *Clin Trials* 2022; 19: 267-273.

10 Wang T, McAuslane N, Liberti L, et al. Companies' health technology assessment strategies and practices in Australia, Canada, England, France, Germany, Italy and Spain: an industry metrics study. *Front Pharmacol* 2020; 11: 594549.

11 Crossnohere NL, Brundage M, Calvert MJ, et al. International guidance on the selection of patient-reported outcome measures in clinical trials: a review. *Qual Life Res* 2021; 30: 21-40.

12 Philip J, Collins A, Phillips J, et al. The development of the Australian National Palliative Care Clinical Studies Collaborative "Integrating Qualitative Research into Clinical Trials Framework". *J Palliat Med* 2021; 24: 331-337.

13 King MT, Agar M, Currow DC, et al. Assessing quality of life in palliative care settings: head-to-head comparison of four patient-reported outcome measures (EORTC QLQ-C15-PAL, FACT-Pal, FACT-Pal-14, FACT-G7). *Support Care Cancer* 2020; 28: 141-153.

14 Regnault A, Pompilus F, Ciesluk A, et al. Measuring patient-reported physical functioning and fatigue in myelodysplastic syndromes using a modular approach based on EORTC QLQ-C30. *J Patient Rep Outcomes* 2021; 5: 60.

15 Novotny PJ, Dueck AC, Satele D, et al. Effects of patient-reported outcome assessment order. *Clin Trials* 2022; 19: 307-315.

16 Brazier JE, Mulhern BJ, Bjorner JB, et al. Developing a new version of the SF-6D health state classification system from the SF-36v2: SF-6Dv2. *Med Care* 2020; 58: 557-565.

17 King MT, Costa DSJ, Aaronson NK, et al. QLU-C10D: a health state classification system for a multi-attribute utility measure based on the EORTC QLQ-C30. *Qual Life Res* 2016; 25: 625-636.

18 European Organisation for Research and Treatment of Cancer. EORTC Quality of Life Group Item Library User Guidelines, 1st edition. EORTC, 2022. https://www.eortc.org/app/uploads/sites/2/2022/10/EORTC-Item-Library-User-Guidelines.pdf (viewed May 2024).

19 Poder TG, Fauteux V, He J, et al. Consistency between three different ways of administering the Short Form 6 Dimension Version 2. *Value Health* 2019; 22: 837-842.

20 Ameri H, Safari H, Poder T. Exploring the consistency of the SF-6Dv2 in a breast cancer population. *Expert Rev Pharmacoecon Outcomes Res* 2021; 21: 1017-1024.

21 Nahvijou A, Safari H, Ameri H. Comparing the performance of the EQ-5D-5L with two versions of the SF-6Dv2 in patients with breast cancer. *Health Serv Outcomes Res Methodol* 2020; 20: 183-194.

22 Talvitie A-M, Ojala H, Tammela T, et al. The expanded prostate cancer index composite short form (EPIC-26) for measuring health-related quality of life: content analysis of patients' spontaneous comments written in survey margins. *Qual Life Res* 2022; 31: 1-10.

23 King-Kallimanis BL, Howie LJ, Roydhouse JK, et al. Patient reported outcomes in anti-PD-1/PD-L1 inhibitor immunotherapy registration trials: FDA analysis of data submitted and future directions. *Clin Trials* 2019; 16: 322-326. ■