The future of cancer therapy

EORTC QUALITY OF LIFE GROUP GUIDELINES FOR DEVELOPING QUESTIONNAIRE MODULES

Fifth Edition

2021

Sally Wheelwright, Kristin Bjordal, Andrew Bottomley, Alexandra Gilbert, Francesca Martinelli, Madeline Pe, Monika Sztankay, Kim Cocks, Corneel Coens, Anne-Sophie Darlington, Peter Fayers, Johannes Giesinger, Michael Koller, Dagmara Kuliś, Morten Aa. Petersen, Jaap Reijneveld, Susanne Singer, Krzysztof Tomaszewski, Deborah Fitzsimmons

on behalf of the EORTC Quality of Life Group

EORTC Quality of Life Group

Module Development Guidelines

5th Edition

PREFACE

The modular approach of the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Group (QLG) was created initially in 1988 (Aaronson et al., 1988); the first version of the guidelines for module development was subsequently published in 1993 (Sprangers et al., 1993). These guidelines have been shown to be a useful tool for module development since they were first developed. Modules that have been produced following these guidelines have exhibited good levels of psychometric and cross-cultural validity. The guidelines also allow those who use modules to understand the rigorous methodology of module development.

Experience with module development since the last revision in 2011 has highlighted areas where the guidelines require further development or refinement. These areas include: (i) alternative methods of identifying relevant Quality of Life (QoL) issues, (ii) links to the newly developed EORTC QLG Item Library held in the EORTC Quality of Life Department at the EORTC Headquarters, (iii) translation of modules, and (iv) changes to the methods used to develop validated modules. These amendments are included in the current version. New sections on procedures for updating modules and merging two related modules are also included.

We would like to thank all members of the EORTC QLG who have contributed to this document, and in particular, all authors of previous versions. We hope that these updated guidelines will continue to ensure uniformly high quality across modules. Users who have comments or questions are encouraged to contact the authors or the Project and Module Development Committee (PMDC) to enable them to further improve the guidelines.

January, 2021

Authors of the current version

PMDC members 2020:

Kristin Bjordal
Andrew Bottomley
Alexandra Gilbert
Francesca Martinelli
Madeline Pe
Monika Sztankay
Sally Wheelwright

Other QLG members:

Kim Cocks
Corneel Coens
Anne-Sophie Darlington
Peter Fayers
Deborah Fitzsimmons
Johannes Giesinger
Michael Koller
Dagmara Kuliś
Morten Aa. Petersen
Jaap Reijneveld
Susanne Singer
Krzysztof Tomaszewski

Contributors to be acknowledged:

Carmen Peuters
Samantha Sodergren

Authors from all previous versions:

Neil Aaronson
Kristin Bjordal
Jane M Blazeby
Andrew Bottomley
Ann Cull
Peter Fayers
Mogens Groenvold
Colin Johnson
Michael Koller
Dagmara Kuliś
John Ramage
Mirjam Sprangers
Galina Velikova
Teresa Young

CONTENTS

ABBREVIATIONS

CAT = Computer Adaptive Testing

CRF = Case Report Form

DIF = Differential Item Functioning

DOG = Disease Oriented Group (an EORTC group that focusses on a specific disease site from the clinical point of view)

EC = Executive Committee of the EORTC Quality of Life Group

EORTC = European Organisation for Research and Treatment of Cancer

EORTC QLQ-C30 = EORTC QLG Core Questionnaire, 30-item version = QLQ-C30

IRT = Item Response Theory

PI = Principal Investigator

PMDC = Project and Module Development Committee (EORTC Quality of Life Group committee responsible for overseeing all projects in the EORTC QLG portfolio)

PRO = Patient-Reported Outcome

QLD = EORTC Quality of Life Department, based at the EORTC Headquarters in Brussels

QLG = EORTC Quality of Life Group (an EORTC group that focusses on Quality of Life)

QoL = Quality of Life

SAC = Scientific Audit Committee (reviews activity of EORTC groups and provides recommendations biennially)

SSG = Statistical Support Group (a group of sub-Quality of Life Group members with an expertise in statistics and who might be consulted for statistical advice)

1 INTRODUCTION

An essential aspect of the 'modular' approach to QoL assessment adopted by the EORTC QLG is the development of modules to be administered in addition to the core questionnaire (the EORTC QLQ-C30).

The EORTC QLQ-C30 + modules are patient-reported outcome (PRO) measures. As stated in the guidelines from the FDA[1], an internationally recognized definition of PROs is:

"any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else. The outcome can be measured in absolute terms (e.g., severity of a symptom, sign, or state of a disease) or as a change from a previous measure. In clinical trials, a PRO instrument can be used to measure the effect of a medical intervention on one or more concepts (i.e., the thing being measured, such as a symptom or group of symptoms, effects on a particular function or group of functions, or a group of symptoms or functions shown to measure the severity of a health condition)".

The overall aim of a module is to improve the sensitivity and specificity of assessments of QoL in specific groups of patients. This allows the patient experience to be assessed in a more precise way so that both differences between groups and changes over time can be captured.

Modules differ in their content. They may relate, for example, to QoL issues affecting particular tumour types (e.g. primary site, metastatic site), aspects of care (e.g. patient satisfaction) or patients' psychological needs or experiences (e.g. information). Modules have been developed for defined patient groups (e.g. elderly), generic cancer symptoms (e.g. fatigue), treatment side effects (e.g. radiation proctitis), or other areas of relevance and importance to specific groups (e.g. activities of daily living in people with brain tumours). In principle, most modules are developed to be used at diagnosis and then during treatment and follow-up.

The modules, like the core questionnaire, are primarily designed for use in cancer clinical trials, but can be used in other research settings and studies as well (e.g. prospective observational studies).

[1] https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf

1.1 Guidelines for Module Development

1.1.1 What is, and what is not, covered by these guidelines

These guidelines will 1) assist and guide module developers who wish to develop a new module (or update an existing module, or merge existing modules), and 2) standardize the module development process in order to ensure uniformly high quality across the portfolio of modules.

Based on the various steps in the initial part of the development process, the important aspects of the patients' QoL are identified. Following this, issues, items and scales are created in order to assess symptoms, problems or functional status.

Definition of ISSUES

Issues are the experiences identified as having an impact on the QoL of cancer patients: for example, "*difficulties walking*".

Definition of ITEMS

An **item** is a question with a timeframe and a reply scale created to assess an issue: for example,
[During the last week]
"*Did you have difficulties walking?*" [Not at all / A little / Quite a bit / Very much]

Definition of SCALES

Various items measuring the same issue can be combined to form **scales**. For example, three different items concerning assessment of pain could be combined into a pain scale. A scale is a group of two or more items which have been combined to assess different aspects of an issue.

These guidelines focus primarily on the traditional development of modules, which typically include a fixed number of preset items, i.e. **'static'** modules.

Different considerations apply to other kinds of questionnaires, for example, to Computer Adaptive Tests (CATs), which are based on and scored according to Item Response Theory (IRT).

These guidelines do not address the construction of CATs. However, the use of IRT and Rasch models in terms of assessing the validity and the relative efficiency of items in multi-item scales is included.

In addition, modules are sometimes developed as '**stand-alone questionnaires**' (for example, to assess patient satisfaction). The development of stand-alone questionnaires is not explicitly described in these guidelines. The approach taken should largely mirror that for other modules.

The EORTC QLG has implemented an approach for creating '**static + flexible**' PRO measurement for cancer patients, based on the EORTC QLG Item Library[2]. An overview of this approach to developing 'flexible' modules will be summarized here, but the full guidelines for the static + flexible approach will be published separately.

Other guidelines are also available for more specific parts of the process. The EORTC QLG Translation Manual[3] (Kuliś et al., 2017), for example, should also be consulted.

Providing specific recommendations about the use in clinical practice is outside the scope of these guidelines. The EORTC QLG Clinical Practice User Manual[4] (Wintner et al., 2016) is available on the EORTC QLG website.

These guidelines cover the full life cycle of module development including:

- Oversight of module development
- General principles of module development
- Gaining approval for a module developed following these guidelines
- Quality assurance procedures for module development
- Publication of a module's development and final approval of an EORTC QLG module
- Submission of data to the EORTC QLG data repository
- Copyright and permission issues
- Specific guidance for the updating or combining of existing modules

1.1.2 Oversight of module development by the EORTC QLG

The first point of contact before developing a new module (or updating an existing module or combining existing modules) is the EORTC QLG Project and Module Development Committee (PMDC).

What is the PMDC?

The PMDC coordinates the process of presenting new proposals to the EORTC QLG and the EC and oversees the development process of all EORTC QLG modules and non-modules, on behalf of the EORTC QLG. Only proposals which receive EC endorsement can apply for EORTC QLG

[2] https://qol.eortc.org/item-library/
[3] https://www.eortc.org/app/uploads/sites/2/2018/02/translation_manual_2017.pdf
[4] https://www.eortc.org/app/uploads/sites/2/2018/02/EORTC_QLQ_Clinical_Practice_User_Manual-1.0.pdf

funding and name the final product an 'EORTC QLG' instrument. The committee has two co-chairs, of which one is a clinician and the other of which has methodological expertise.

In addition, the PMDC is composed of active members of the EORTC QLG and staff members from the EORTC Quality of Life Department (QLD).

The PMDC co-chairs

The PMDC co-chairs are active member of the EORTC QLG, with experience and expertise in module development and an established track record (such as international peer-reviewed publications) in PRO measurement. The PMDC co-chairs are also full voting members of the EORTC QLG EC for their term of office.

The PMDC co-chairs work together, with the support of the other PMDC members.

All reports that are submitted by developers at the end of different phases of module development are reviewed by EORTC QLG members, with oversight from the PMDC co-chairs, to ensure that all modules within the EORTC QLG portfolio meet the required standard to be designated as an official EORTC QLG module.

The PMDC co-chairs are also available to give advice and support to module developers during the process, e.g. to provide responses to ad-hoc queries.

Where a major protocol deviation is identified (such as failure to recruit in all specified countries/languages recommended), the module coordinator must discuss this with the PMDC as soon as possible. However, where a decision is required to amend a submitted protocol and/or grant (including any financial issues), the PMDC co-chairs will need to discuss this with the EC.

Retrospective endorsement of any module as an official EORTC QLG module **cannot be given**. At any time during the process of module development, the EORTC QLG reserves the right to withdraw support/endorsement of a module, e.g. where there has been a fundamental breach in following the guidelines.

The names of the current PMDC co-chairs are listed on the EORTC QLG website[5]. All enquiries regarding module development should be sent to pmdc@eortc.org.

[5] https://qol.eortc.org/

1.2 Different Phases of Development

Patient interviews

Inclusion of qualitative interviews at the earliest stage of module development is essential to ensure content validity. This has been a cornerstone of EORTC QLG module development for many years and the inclusion of qualitative interviews is endorsed by the FDA guidance (2009). Patient interviews are the most important of four steps identified by Rothman et al. (2009) to ensure that high content validity is achieved and demonstrated. The other steps recommended are: good conceptual match between instrument and its purpose (including a clear definition of the module population); demonstration that the most relevant and important content is included; and good documentation of all modifications made to items or modules. Appropriate methods for developing conceptual issues and frameworks for qualitative interview research, developing the interview discussion guide, reaching saturation, analysis of data, developing a theoretical model, item generation and cognitive debriefing are available (Brod et al., 2009).

Specific guidance for conducting interviews using video conferencing or phone interviews is not provided in these guidelines. At the time of writing, module developers have been using these methods due to restrictions as a result of the COVID-19 pandemic. It is recognized that module developers may continue to use different approaches for pragmatic reasons, even if restrictions are lifted. It is essential that the mode of interview is always recorded and sensitivity analyses should be conducted to evaluate whether the interview mode has any effect on results.

Overview of review process coordinated by the PMDC

The development of a new module goes through four phases. Usually, three formal reports are required to be submitted to the PMDC: one after Phase 2, one after Phase 3a and one after Phase 4. A Phase 3b report is required if changes to the module are made as a consequence of Phase 3b results. A separate Phase 1 report may be requested on a case-by-case basis. Completion of each phase occurs when the relevant report is approved by the PMDC co-chairs.

The four phases of the traditional development of a new module are briefly summarized below. More detailed instructions follow later in these guidelines.

Phase 1

An exhaustive list of QoL issues, specific to the group of patients the module is intended for, is identified through literature reviews and interviews with healthcare professionals and patients.

The output of Phase 1 is a list of issues.

The term '*module that has completed Phase 1'* describes modules for which a list of QoL issues has been generated. Typically no report is produced for Phase 1 (unless requested by the PMDC); rather, a joint report is produced for Phases 1 and 2.

Phase 2

The module is constructed. Issues are converted into items and items are translated by the QLD Translation Unit into the languages required for Phase 3. Use of the EORTC QLG Item Library is mandatory in this phase to avoid duplication of existing items.

The output of Phase 2 is a provisional module.

A module will be considered to have '*completed Phase 2'* once the list of QoL issues has been converted into a set of items and a provisional version of the module has been prepared, reviewed and approved. This includes approval by the PMDC co-chairs of the Phase 1/2 report. On approval, the English version of the module will be designated as a Phase 2 module.

Phase 3

This comprises two parts. In Phase 3a, the provisional module is pre-tested to identify and solve potential problems in its administration (e.g. the phrasing of questions, the sequence of questions) and to identify missing or redundant issues. In Phase 3b, preliminary testing of the psychometric properties of the provisional module, including scale structure of the module, is performed (Phase 3b).

The output of Phase 3b is the semi-final module (designated as a completed Phase 3 module), which will be tested in Phase 4.

A module that has completed Phase 3a and 3b as described in these guidelines, and has received the formal approval of the Phase 3 reports by the PMDC co-chairs, is described as having '*completed Phase 3'*. Such modules may be used in clinical trials under certain conditions and with appropriate permission which are detailed on the EORTC QLG website (under "Questionnaires").

A module that has completed Phase 3 and is undergoing validation testing may be described as '*in Phase 4'*.

Phase 4

The module and its scale structure are field-tested with a large, international group of patients in order to determine its acceptability, reliability, validity, responsiveness and cross-cultural applicability.

The output of Phase 4 is the final module.

When a module has successfully completed Phase 4 and has received formal approval based on review by the EC and other peer reviewers from the EORTC QLG, the module is described as both '*Phase 4 completed'* and '*validated'*.

For all validated modules, the English version of the module, all official translations and a scoring/interpretation guideline are available.

Usually, a paper describing the module's development will be submitted for publication in a peer-reviewed international journal after Phase 3 and Phase 4, but sometimes also for the work in Phase 1/2. All papers should be published at least in Green Open Access.

Timelines

Developing a module according to these guidelines is a rigorous process, which is needed to ensure high-quality instruments.

Investigators will need to take several factors into account when estimating timelines.

Some examples are: ethical approval needs to be obtained; some of the collaborators might wish to have a contract in place; recruitment might take more time for rarer cancers; the report at the end of one phase needs to be approved before being able to continue with the next phase. This list is not exhaustive. Part of the approval process for a module development will be an assessment of the feasibility of the module development to be completed in a timely manner (see section 3.1).

Copyright and permission issues

For copyright please see chapter 9. Full details on how to gain permission to use the module are kept up to date on the EORTC QLG website.

2 COORDINATION OF MODULE DEVELOPMENT

To ensure uniformly high quality in modules, the entire development process is subject to monitoring, peer review and quality assurance within the EORTC QLG. These activities are managed by the PMDC. The purpose of these activities is to:

1) Monitor progress
2) Avoid unintended duplication of effort
3) Avoid unnecessary or unwanted variation between modules
4) Evaluate the process, i.e. whether any deviations from standard procedures were justified and whether any alternative procedures followed were sufficient to meet the standards set
5) Evaluate the quality, suitability and compatibility of the provisional and final modules and their translated versions
6) Ensure the highest scientific standards in module development.

2.1 Oversight of Module Development Processes

2.1.1 Role of the PMDC co-chairs

The PMDC co-chairs are responsible for the oversight of module development on behalf of the EORTC QLG Executive Committee (EC).

The PMDC co-chairs coordinate module development through:

1) Reviewing, with the support of the PMDC and of EORTC QLG members, proposals for modules to be developed
2) Verifying that:
    - the proposal for a new module (or updating or combining existing modules) has been approved by the EORTC QLG EC following the process of review outlined in section 3.2;
    - the composition of the module development group is commensurate with recommendations (see section 3.3);
    - the module development reports at all phases of development are commensurate with the guidelines;

- the lead Principal Investigator (PI) has provided a six-monthly report on the conduct of the module development process and attended the biannual EORTC QLG meetings in person (or has had representation at the meeting);
- a report on the progress of each module a) has been presented at the PMDC plenary meeting every six months and b) is made available on the EORTC QLG website (in the form of a summary);
- if the module development process is funded by an EORTC QLG grant, the terms and conditions (including adherence to agreed timescales) have been met; and
- all completed module developments are stored within the EORTC QLG data repository held at the QLD.

3) Being available throughout the module development process for advice

4) Keeping the guidelines for module development up to date by making revisions when needed.

The above list is meant to provide an overview and is not exhaustive.

All potential module developers are advised to contact the PMDC co-chairs at an early stage about their idea/proposal.

Module developers must consult the PMDC co-chairs if deviations from these guidelines are anticipated or are being made during the module development process.

2.1.2 Role of the reviewers

All module development reports will be peer reviewed by at least two members of the EORTC QLG. Peer-reviewers will be independent, i.e. not a named collaborator, of the module development activity.

The PMDC will obtain reviews of the reports produced by the module developers after Phases 1 and 2, Phase 3a, and Phase 4, (and for Phase 3b if there are changes), from at least two members of the EORTC QLG. The reviewers should not have been involved in the module development, although there may be circumstances in which a reviewer may have been previously consulted for advice. Where necessary (e.g. to avoid unnecessary delays in the review process), one or more PMDC members may be asked to peer review.

Manuscripts for publication, which should usually be written '*on behalf of the EORTC Quality of Life Group',* may be submitted for approval in place of reports, but additional appendices will usually be required to provide sufficient detail. These are reviewed by members of the EC, again excluding those who are involved with the specific module. This is a separate process, described in chapter 8.

Depending on the nature and scope of the comments made by reviewers, the report, manuscript for publication or new module may need revision. The module developers should respond in writing to the PMDC co-chairs to answer the issues raised by the reviewers alongside producing both a marked up and a clean copy of the revised report.

Reviews apply not only to reports and manuscripts, but also to translations. Review of the translation processes and resulting translated versions is coordinated by the EORTC QLD. The module developer and the QLD translation team leader are required to review the pilot-testing results of all translations to ensure the appropriateness of the translated version.

The PMDC will liaise as necessary with the QLD translation team leader to ensure all reviews have been completed.

2.2 Quality Assurance

When a new project starts, the ethics approval, informed consent form, protocol and a list of PIs and sites are collected and archived at the EORTC Headquarters.

In order to keep the module development process harmonized, standard templates are available for protocols and reports, and also for reporting the scoring algorithm. These templates are available on the EORTC QLG website.

In order to promote good data preservation strategies, pseudonymized data from Phase 3 and 4 studies are collected, harmonized and stored at the EORTC Headquarters, according to the rules established in the contract or Memorandum of Understanding.

3 GETTING STARTED

3.1 Three Types of Module Development

Module development may comprise (1) developing a new module, (2) updating an existing module, or (3) combining existing modules. In all cases, the development of a module follows the already-described four-phase process; full details of each type of module development are provided in the subsequent chapters.

3.2 Approval of Proposal and Protocol Design

Researchers who are considering carrying out module development should discuss this in the first instance with the PMDC who can advise on procedures and give a preliminary view of the suitability of the proposed module. They can also receive practical information from PMDC members via the email address pmdc@eortc.org.

All module development proposals must be presented (usually in person) at one of the biannual EORTC QLG meetings. Before the proposal can be presented at a meeting, a two-page proposal form, provided by PMDC staff, must be completed and returned, usually one month before an EORTC QLG meeting.

3.2.1 Two-page proposal

When a suitable instrument to assess QoL for a specific topic relevant to oncology research is missing, a new module might need to be developed. Module construction begins with a clear description of the research question and the target population for which the module will be designed. The need for a new module, for the revision of an existing module, or for the merging of two modules should be clearly demonstrated within a proposal. Proposals for new modules will be evaluated on the following criteria:

- Novelty: the new module addresses a current gap not adequately covered by existing instruments (inside or outside the EORTC QLG portfolio).
- Relevance: the module targets a specific topic (population, experience, aspect of care, etc.) that is relevant for the evaluation of oncology interventions.
- Feasibility: the topic of the module can be unambiguously defined and is not considered a temporary issue (i.e. it is not linked to a specific drug test or intervention that may become obsolete). There should be evidence that sufficient patients can be practically recruited to complete the module validation. For rare tumours, the module developers should provide a

clear indication of how they will be able to recruit sufficient numbers of patients during each of the phases, including the interviews of patients in Phases 1 and 3a. It is not possible to use interview data from previous studies in place of the interviews described in these guidelines.

- Strategy: the proposal must fit within the strategies of both the EORTC and the EORTC QLG. As these evolve, please check the EORTC and EORTC QLG website for more information.

When the proposal is returned, the PMDC will check that all the criteria listed above are met. The researchers will then be invited to present their proposal at the next EORTC QLG meeting, for discussion and feedback from the PMDC members, other members of the EORTC QLG and possibly EORTC Disease Oriented Groups (DOGs).

The PMDC will report on these discussions to the EC. The EC must endorse the proposal to allow the researcher to apply for an EORTC QLG grant and for the subsequently developed module to be accepted as an official EORTC QLG module.

Even if the researcher has independent funding and will not be applying to the EORTC QLG for funds, endorsement of the proposal and each phase is still required for the final module to be accepted as an official EORTC QLG module.

More details concerning this procedure are available on the EORTC QLG website. For questions, please contact pmdc@eortc.org.

3.3 Module Developers

Whether the new module is a brand new one, one resulting from the update of an existing module, or one resulting from the merging of existing modules, the following principles apply.

The module development group

Whilst the composition of a module development group will be determined by specific requirements (e.g. to ensure sufficient recruitment of patients from the designated populations to be covered by the module) and may evolve over the lifecycle of a module's development, it is mandatory that the following principles are adhered to (unless another arrangement is agreed by the PMDC).

- There is one lead PI who takes overall responsibility for the scientific conduct and administrative responsibility for the reporting of the module development process to the EORTC QLG. Normally, the lead PI has experience in PRO development and use, often gained through contribution to previous EORTC QLG module developments.

- A co-PI can be nominated. The co-PI may have specific expertise to support the module development process. The co-PI does not necessarily have to be an EORTC QLG member; collaboration with the relevant EORTC DOG or other area (e.g. Task Force) is strongly recommended. In the absence of the lead PI, the co-PI is expected to deputize.

- Other members of the module development group may be drawn from the EORTC QLG and/or involve collaboration across different countries/sites.

- The inclusion of a named statistician with experience in PRO development and analysis as a collaborator is strongly recommended, and indeed required during Phases 3 and 4 of module development.

- The composition of the module development group should reflect a multidisciplinary and multinational team. In some module developments, this could include engagement with an established clinical network, e.g. to support patient recruitment. All members of the module development group are encouraged to join the EORTC QLG so that they can attend a parallel session within the EORTC QLG biannual meeting where module development progress is discussed.

In short, a module development group should include at least:

- A methodologist with experience in developing PRO instruments

- At least one clinician with expertise regarding the specific field and patient group of interest. When developing a module which is relevant for a specific tumour type or population the relevant DOG or Task Force, if existing, should be informed and invited to collaborate

- A dedicated person (e.g. named study coordinator/researcher) with enough time to focus on the day-to-day practical, logistic and administrative tasks that the development of a module brings

- A statistician for Phases 3 and 4

- Ideally, a patient representative to provide input of the patient experience.

Any substantial changes to the composition of a module development group during the life cycle of a module's development should be reported to the PMDC co-chairs.

The module development group will collaborate, under the leadership of the PI(s), as a group of independent researchers, and should seek to achieve consensus after each crucial step. The developers should regularly review progress, data collection and analysis of the responses, and also contribute to any report writing in order to agree upon:

1) The formulation of the QoL domains to be included (e.g. symptoms, functioning) – a clear conceptual framework should be produced during Phase 1 to explain the rationale for the focus of the module;

2) The list of QoL issues derived from the literature to be put to patients and healthcare professionals in different countries;

3) The final list of QoL issues to be included in the provisional module;

4) The final list of items (provisional module) and tentative scoring system; and

5) The final validated module and publication.

Module development group meetings

It is mandatory that throughout the module development process the lead module developer(s) attend the biannual EORTC QLG meetings and convene a meeting to discuss module progress during a parallel session. These meetings are usually open to all members of the EORTC QLG to attend, although in some cases the meeting can be closed to attendees not actively involved in the module's development process.

4 DEVELOPMENT OF A NEW MODULE

Conditions required to proceed

1) The outline proposal has been approved by the EC
2) If relevant: the grant application for the first phase(s) of development has been approved by the EC

Standard protocol and report templates

In all phases of development, standard templates are available for protocols and reports. Their use is mandatory. More details are provided in Chapter 8.

4.1 Phase 1: Generation of QoL Issues

4.1.1 Aim & design

Phase 1 is aimed at compiling an exhaustive list of relevant QoL issues that cover the domain(s) of interest in order to achieve content validity. In the process of compiling this list, three sources are used: literature review (including existing questionnaires), interviews with healthcare professionals, and most importantly, patient interviews.

4.1.2 Literature review

A systematic approach is required and an account of the search process and selection of literature should be documented; however, a full systematic literature review is not mandatory.

4.1.3 Interview samples

Cross-cultural consistency

Phase 1 should include participants from at least four countries which represent the language groupings below.

At least one country from:

a) English-speaking countries (UK, US, Canada, Australia, etc.)

plus at least one country from at least three of the six following groupings (for example Germany, Poland, Greece):

b) West-Germanic languages: German (Germany, Austria), Dutch (the Netherlands, Belgium)

c) Scandinavian languages: Danish, Swedish, Norwegian, Icelandic

d) Slavic languages: Polish, Czech, Slovak, Russian, Belarusian, Ukrainian, Slovene, Serbian, Croatian, Macedonian, Bulgarian

e) Romance languages: French (France, Belgium), Italian, Spanish (Spain), Portuguese (Portugal), Romanian

f) Other European languages: Greek, Hungarian, Finnish, Lithuanian, Latvian, Estonian, Welsh, etc.

g) Non-European languages: Turkish, Hebrew, Chinese (China, Taiwan, Singapore, Malaysia, Hong Kong), Spanish (Spanish-speaking Latin American countries), Portuguese (Brazil), etc.

Healthcare professional sample

Healthcare professionals may be of any relevant discipline, such as physicians, nurses, psychologists or dieticians. They must have sufficient clinical expertise and experience with treating patients belonging to the target population, e.g. a track record of working in the area over several years.

At least eight healthcare professionals should be included, drawn from all countries represented in Phase 1. It is usually unnecessary to recruit more than 20 individuals.

Patient sample

Patients with the relevant condition, at all relevant stages of disease and treatment, are eligible for the patient interviews. Prior to enrolment, the target population and the characteristics that are expected to most influence QoL, e.g. disease stage and treatment type, should be clearly identified and a recruitment matrix developed (see Appendix 1 Matrix for assisting in targeting patient recruitment). Where possible, the recruitment matrix should be designed to ensure that the most diverse range of participants are recruited, whilst ensuring the targets are achievable and there are not too many cells in the matrix.

It is recommended that 5–10 patients should be interviewed from each cell in the matrix. A minimum of 20 patients in total should be interviewed. Usually no more than a total of 30 patient interviews for the intended module are required. Interviews should continue until no new issues arise, i.e. data saturation is achieved across the entire dataset. This can be defined as three consecutive interviews where no new issues arise (Francis et al., 2010).

In addition to the characteristics used to define the recruitment matrix, it is important to ensure that other characteristics in the recruited sample, e.g. gender and age, reflect the target

population. Strict eligibility criteria should be developed and the module development coordinator will need to devise a recruitment monitoring strategy.

Patients should be recruited from a variety of locations, such as hospital inpatient and outpatient clinics and, if relevant, community settings. The nature of the module and the target population will help identify the most appropriate sources. Care should be taken if patients are recruited from self-selecting sources such as self-help groups as this increases the risk of bias. Similar numbers of patients should be recruited from each country participating in Phase 1 with no more than 30% from a single country.

Exceptions

In some cases, for example for certain subgroups of patients with brain tumours, it might be difficult to interview patients. Family members or caregivers could be interviewed instead. This scenario constitutes an exception. If the plan is to interview family members or caregivers instead of patients this should be clearly stated in the proposal and it should be discussed with the PMDC prior to starting the work. All planned details, for example who fills in the module, should be outlined and justified in the proposal.

4.1.4 Method

The following text provides a guideline for the literature review and interviews with patients and healthcare professionals. Adaptations are permissible if justified. Module developers are encouraged to discuss any planned adaptations with the PMDC.

Literature review

Existing questionnaires and clinical literature

Literature searches should be conducted, following PRISMA guidelines, on the appropriate databases, e.g. MEDLINE, ScienceDirect, Web of Science, EMBASE, PSYCHINFO, SciElo, to produce a list of all potentially relevant QoL issues. In addition, searches of the same databases and other sources (such as PROQOLID) should be carried out to identify existing relevant questionnaires. This could include both general QoL questionnaires and disease-specific questionnaires.

Module developers who wish to publish their review should use a systematic approach for the identification and review of previous studies and questionnaires. All details of the literature review process (databases, key words, search strategy, date and language restrictions, inclusion/exclusion criteria for articles searched, flowchart depicting the flow of the search with

detailed reasons for study exclusion) should be presented in the Phase 1 report as per PRISMA guidelines[6].

In order to achieve content validity, the clinical literature around the domain should be reviewed. In the development of new modules or topics that are more abstract than symptoms (e.g. psychological or sociological concepts) careful attention must be given to the theoretical literature. Issues may also need to be selected based on their ability to reflect an accepted theoretical/conceptual framework.

The search may also identify existing questionnaires or relevant questions that may exist within general QoL questionnaires. The underlying issues should be extracted from the questions and added to the list derived from the theoretical and clinical literature.

Issues may arise which appear similar. In the early phases, rather than the researcher making a judgement to exclude possibly overlapping issues, it is better to seek input from healthcare professionals and patients if the number of issues on the list is not excessive.

Interviews

Healthcare professional and patients should be interviewed, in this order.

Interviews with healthcare professionals

The provisional list of issues and the core instrument, if appropriate, should be presented to healthcare professionals for feedback on appropriateness of content and breadth of coverage.

Interview technique

The list of QoL issues may be administered in the form of a structured, personal interview in which basically two questions are asked: (a) are issues included which the specialists consider irrelevant for this patient group, and if so, why do they consider these issues irrelevant? and (b) are there issues missing from this list that the specialists consider relevant, and if so, why do they consider these issues relevant?

To establish the relative importance of the QoL issues, the specialists should be asked to identify a subset (e.g. a quarter) of issues that, in their opinion, affect patients' QoL most profoundly and which should definitely be included in the final module. Appendix 2 Generation of relevant QoL issues in Phase 1: provides an example of a detailed interview protocol. The module developers will consider the comments of these specialists during selection of items for inclusion in the module for Phase 3.

[6] http://www.prisma-statement.org/

Amendments of list of issues

New issues identified by healthcare professionals should be added to the issue list but researchers are discouraged from removing issues at this stage, before patients have been consulted.

Patient interviews

Interviews with patients

The provisional list of issues and the core instrument, if appropriate, should be presented to patients for feedback on appropriateness of content and breadth of coverage.

Interview technique

Open-ended interviews, sometimes referred to as open-ended concept elicitation, should be used to explore patients' perceptions of their quality of life in relation to their cancer experience in as unbiased a manner as possible. During the interview, the patient should be encouraged to consider all issues which they believe to be relevant to the condition. The researcher may ask the patient to describe their experience in a way that allows patients to explore issues of relevance to them. The interview design and prompts used will be decided by the nature of the module under development. A flexible approach to the interview schedule should allow the researcher and patient to explore interesting issues in order to provide a rich account based on the patient's own experience, but note that the main output of these interviews should be a list of relevant QoL issues.

After the open-ended interview is completed, the patient should be shown the EORTC QLQ-C30 and the issue list to prompt the elicitation of further issues. For each issue on the list, patients should be asked to indicate the extent to which they have experienced it during their illness, preferably using the EORTC QLQ-C30 response scale (not at all, a little, quite a bit, very much), as a measure of relevance. For those issues which patients have experienced only, they should also be asked to indicate how much the issue has troubled them / had an impact on them, using the same response scale as a measure of importance. Patients may also be asked to identify issues which they think should definitely be included (e.g. a quarter of the issue list) or definitely excluded.

Appendix 3 Generation of relevant QoL issues in Phase 1: provides an example of a detailed patient interview protocol.

If possible, patient interviews should be recorded for the purpose of reporting accuracy and audit. When this is not possible, careful field notes should be taken at the time of the interview.

Final selection of issues

The lists of issues from all sources should be reviewed by the module developers in order to produce a single, comprehensive list of issues for formulation into the provisional module in Phase 2. Module developers should agree on the decision rules to be used before the selection of items takes place. If there is disagreement between the views of patients and professionals, the views of patients will usually take precedence. In every case, the reasons for inclusion or exclusion of items should be given in the Phase 1/2 report, with clear reporting of the results of the decisions rule applied to inform the final selection of issues.

Decision rules may vary somewhat across modules. Examples of decision rules suitable for adaptation are given in Appendix 4 Decision rules for selection of QoL issues in Phase 1. Issue lists should be reviewed for overlap between issues and care should be taken that potential new issues are not already covered in the core questionnaire (QLQ-C30). When there is a very large number of issues (e.g. more than 50), most weight should be given to the patient responses during selection of issues.

Variations in approach

The method explained provides a guideline for using literature review and interviews with patients and healthcare professionals, but adaptations are permissible provided the following are determined:

- Relevance: the extent to which patients have ever experienced issues on the list including problems, limitations and positive experiences.
- Breadth of coverage: that the list includes all significant issues.
- Importance: relative importance of issues, i.e. how much they impact the patient.

Module developers are encouraged to discuss any significant proposed adaptations with the PMDC. One modification which does not require discussion is the use of focus group interviews as an alternative or supplement to individual interviews. Focus groups may be particularly beneficial for the generation of new issues. However, if this approach is used, it is still important to collect information about relevance and importance from individuals, perhaps at the end of the focus group.

4.1.5 Output from Phase 1

An exhaustive list of QoL issues that are of relevance for the module.

The recruitment matrix defined for Phase 1 and the experience with the recruitment will inform the matrix used in Phase 3.

4.2 Phase 2: Construction of the Module (Provisional Module)

4.2.1 Conditions required to proceed

1) An exhaustive list of QoL issues (output of Phase 1)

2) Usually a report is required after the completion of Phases 1 and 2 together, but in some cases, the EC may require the module developers to produce a Phase 1 report e.g. as a condition for continued funding

4.2.2 Aim & design

Phase 2 is aimed at converting the list of QoL issues into questions or items, usually with the format and time frame compatible with the EORTC QLQ-C30. That is, items refer to the patient's experience *during the last week* and the response is recorded on a 4-point Likert scale.

Exceptions to this one-week timeframe or 4-point scale may be acceptable. For example, if the issue or problem area is unlikely to be captured with a one-week timeframe, it can be extended. If an issue is either absent or present, the question may be formulated in a binary format (e.g. No or Yes). Where other variations are required (further details are given in 4.2.3), these should be discussed with the PMDC and QLD translation team leader in advance. Any proposed change in timeframe or format needs to be clearly identified and justified.

4.2.3 Method

Transforming issues into items

At this stage, it is important to avoid duplication of effort and to ensure uniformity across modules. The wording of new items should, as far as possible, be consistent with existing EORTC QLG modules. Existing items should be used unless there are strong arguments not to do so and these should be discussed and approved by the QLD translation team leader. This maintains consistency and reduces the requirement for translation, as existing translations are available in the EORTC QLG Item Library.

Items can be sourced from the EORTC QLG Item Library or other questionnaires, or can be constructed as new items. Specific guidelines on how to select existing items and specific

guidelines on how to create new items are available on the EORTC QLG website and they should be consulted and followed.

The EORTC QLG Item Library

The EORTC QLG Item Library should be the first place to check for items relating to the issues from the Phase 1 list. The EORTC QLG Item Library is regularly updated with new items (from modules that have completed Phase 3) and translations. The EORTC QLG Item Library may yield several possible wordings that pertain to the same issue. The recommended wording is specified for such items and should be used in all new modules. If another wording seems more appropriate for the module under development, the decision should first be discussed with the QLD Translation Unit. If several items addressing similar issues are identified, it may be advisable to test these in samples of patients from the target population or test them in Phase 3a. Access to the EORTC QLG Item Library can be obtained by completing the 'Request Access' form.

Other resources

In order to save time and effort, existing questionnaires developed by other research groups may be consulted for their wording. Subscriptions to the PROQOLID database are available from the PMDC to assist in searches. This database is a comprehensive, searchable record of QoL and other patient-reported outcome measures. Note that the format of existing questionnaire items may require adaptation to achieve consistency within the module. The explicit consent of the questionnaire constructors should be obtained prior to including the items in the module.

Item construction

If a new item is needed for the new module, it is important to be aware of the major methodological considerations in item construction. Advice on item construction can be found in standard textbooks. Some key points are that the question should be clear, brief, unambiguous and ask about only one issue. It is important to formulate questions appropriately with respect to the response options, and quantifiers must be avoided (e.g. do you have a lot of pain).

For positive issues, the resulting items should be positively phrased (i.e. in terms of abilities, capacities, and positive experiences). Other items should be negatively worded (i.e. in terms of problems, limits in functioning, and negative experiences). Module developers should take steps to avoid possible confusion and biased responses due to differences in the orientation of items (negative versus positive). Patients' attention can be drawn to these differences, for example, by putting salient words in bold or underlining them. Items of similar orientation should be grouped together in the module.

Conditional questions

If the question relates to the impact of a certain symptom, intervention or side effect, consideration should be given to how patients who are not experiencing that particular issue will answer the question (e.g. if asking whether pain medication helps, how could patients who do not take pain medication respond?). Similarly, responses about the impact of a patient's ability or capacity may depend on whether the patient uses that attribute (e.g. sexual functioning). A "Not at all" response should be unambiguous.

When asking about the **impact** of a symptom, intervention, side effect, activity, etc., it is advised to first ask about the presence/frequency/severity of that topic (i.e. a screening question).

For example:

1. "Have you taken pain medication?" No/Yes

2. "If yes (if you have taken pain medication), how much did it help?" Not at all / A little / Quite a bit / Very much.

Some questions are only applicable under a specific condition. The most common examples are gender specific questions or questions only to be answered if a specific intervention or event happened (e.g. "answer only if you have a stoma bag"). In this case, the questions that only apply if the condition is satisfied must be clearly identified.

When reporting on conditional questions, a distinction must be made between those responses left blank due to absence of the condition (i.e. not applicable to the patient) and those responses left blank due to missing data (i.e. condition is present but the patient has not responded). It is important to discuss the scoring of screening questions clearly during module development. If conditional questions are used, the validation report should include the prevalence of these conditions.

Inserting a condition as a statement, rather than a screening question, should be avoided unless there is a not applicable (N/A) option.

For example, questions like "If you take pain medication, did it help?" should be avoided.

What is very important is that conditional questions are clearly identified.

A question like "Have you had sore skin around your stoma?" should only appear if there is an N/A option, or a clear statement like "Answer this question ONLY if you have a stoma bag", or a screening question like "Do you have a stoma bag?".

Provisional Module Consultation

A wide consultation process is essential to ensure high-quality items. The provisional list of items should be reviewed for clarity and overlap by individuals with expertise or knowledge of module development, individuals with experience of the target population other than those who were involved in Phase 1 (including patients and healthcare professionals), and all members of the module development group. It may be beneficial to consult with the relevant EORTC DOG to ensure breadth of coverage.

Review and Translatability Assessment by the QLD Translation Unit

The list of items chosen for inclusion in the Phase 3 module should be sent to the QLD Translation Unit to:

1) Ensure that items are worded in a correct and consistent way; and

2) Identify any items that might pose difficulties in the translation process, i.e. translatability assessment.

The QLG Translation Unit will liaise with the PMDC for the review of the provisional module.

Following review, the QLD Translation Unit provides comments and suggestions for changes. Once discussed and approved, the QLD Translation Unit prepares the provisional module in the standard EORTC QLG formatting and provides it to the PI.

Scale structure

The forming of multiple item scales, which must make clinical sense, should be anticipated and an adequate number of items included to cover the construct. As a loose guide, while physical symptoms and side effects can sometimes be captured by single items, psychological issues and functioning scales are perhaps more likely to require multi-item scales. Large summary (symptom) scales should be avoided. If the purpose of the module is to address symptoms only, other approaches, e.g. construction of an item list, may be preferable.

Scoring will be simplified if all items in a scale are negatively or positively phrased. Items hypothesized to form a scale should usually be adjacent to each other in the module.

Scales including conditional and non-conditional items should be avoided.

4.2.4 Output from Phase 2

On the basis of Phases 1 and 2, a conceptual list of the relevant items should be emerging. This will include the identification of groups of items that are hypothesized to be measuring particular issues. Some of these items may be highly correlated and redundant items will eventually be dropped. However, sometimes the inclusion of a number of correlated items may be needed, for

example to capture the breadth of complex issues, and these items will then form a multi-item scale.

As noted above, wherever possible, existing items and multi-item scales should be drawn from the EORTC QLG Item Library, simplifying the development of the module. The conceptual model should also indicate which constructs are symptoms and which are functional abilities. This will help inform the provisional scoring structure in Phase 3.

Once the Phase 1/2 report has been approved, the preliminary module must be translated into the languages needed for Phase 3. For more details on translation, see chapter 7. A table of issues and resulting items should be produced.

4.3 Phase 3: Pre-testing

The aim of pre-testing the module is to identify and solve potential problems in its administration (e.g. the phrasing of questions in English and other languages, the sequence of questions), to identify missing or redundant issues and to carry out a preliminarily evaluation of the psychometric properties.

Even if all items are taken from existing questionnaires, there is still an obligation to pre-test the module, because:

(a) Items may require adaptation when used in different languages and cultural settings to those of the initial development (that is, in Phases 1 and 2); and

(b) Questions developed for a particular target group may perform differently when applied in a new setting.

Pre-testing consists of two phases: structured interviews (Phase 3a) and preliminary psychometric evaluations (Phase 3b).

4.4 Phase 3a: Structured Interviews

4.4.1 Conditions required to proceed

1) Provisional module from Phase 2

2) Approved Phase 1/2 report

4.4.2 Aim & design

This step consists of:

1) Administering the EORTC QLQ-C30 and the provisional module to patients belonging to the target population, but who were not involved in Phase 1, to obtain a response score for each item, together with ratings of relevance and importance.

2) Conducting structured interviews with each patient after completing all items to ensure that items are clear and acceptable.

3) Exploring any other issues which emerged from Phase 2, such as testing different versions of the same item if relevant.

4.4.3 Sample

Strict eligibility criteria should be defined to ensure that participants adequately represent the target population for which the module is being devised. A sample matrix should be drawn up to include all relevant strata as identified in Phase 1.

In Phase 3a, each cell of the sample matrix should contain at least 15 patients. Examples of possible sample matrices are shown in Appendix 1 Matrix for assisting in targeting patient recruitment.

Cross-cultural consistency

Phase 3 should be conducted in a wider range of countries and regions than Phases 1 and 2: the use of at least one English language, one non-European language, and at least one language from at least three of the six language groupings described in 4.1.3 is recommended (at least five language groupings in total). Between 10% and 30% of patients should be recruited from each language grouping. If this is not possible, it is essential to discuss with the PMDC.

Regarding cultural acceptability and linguistic validation of items, the aim of Phase 3 is to make sure that an item is comprehensible in all the languages. Note that any changes suggested by results from one language should be discussed with the Translation Unit and might lead to changes to the English source module and all translations. For further details on this subject please check the Translation Manual or contact the Translation Unit.

4.4.4 Method

Each patient should usually complete the EORTC QLQ-C30 and the module. In addition, each item of the module should be rated by each patient for relevance and importance to that individual. For each issue on the list, patients should be asked to indicate the extent to which they have experienced it during their illness, preferably using the EORTC QLQ-C30 response scale (not at all, a little, quite a bit, very much), as a measure of relevance. For those issues **which patients have**

experienced, they should also be asked to indicate how much the issue has troubled them / had an impact on them using the same response scale, as a measure of importance. Alternatively, patients could be asked to select the subset of items which they consider to be the most important or they could rank items by importance.

The structured interview

The interviewee should, in principle, be directed to each item separately and should be invited to make further comments about:

(1) The particular experience to which the item refers (e.g. "Is this experience related to your disease or treatment?"); and

(2) The wording of the item itself (e.g. "Was the item difficult to respond to? Was the item annoying, confusing or upsetting? How would you have asked this question?").

If there are a large number of items (i.e. more than 20) the time involved in inquiring about each individual item may be prohibitive. In this case, the questions may be directed towards the entire module (e.g. "Were there questions that you found difficult to answer? Were there questions that you found annoying, confusing or upsetting? Do you have other comments about these questions?"). These general questions may then be supplemented by the further probing of selected module items that are expected to cause some difficulty or items that appear to be troublesome during the interview. The pre-testing interview should be completed with two questions directed at the entire module (i.e. core questionnaire plus module):

(1) "Were there questions that you found irrelevant?"

(2) "Can you think of additional issues that are relevant for you but are not included in this questionnaire?"

On the basis of this pre-testing phase, the module may require adaptation. Examples of a detailed interview protocol as well as decision rules for deletion, addition, and rewording of items are provided in Appendix 5 Example of a patient interview in Phase 3a. Results of the pre-testing, summarized by language and country, should be sent to the QLD Translation Unit. All changes in wording (addition, rewording, merging) should be discussed for feasibility with the QLD Translation Unit and the PMDC before being included in the Phase 3a report.

Analysis and retention/deletion of items

Each item should be considered for retention or rejection on the basis of the comments made by patients and their ratings: items which are viewed as irrelevant by a substantial number of

patients should be considered for rejection. Items which are upsetting may benefit from modification but do not necessarily have to be rejected outright.

Clear decision rules should be defined by the module developers before analysis of the Phase 3a responses. Examples are shown in Appendix 6 Decision rules for inclusion or exclusion of items in Phase 3. The threshold for retention of items in Phase 3a should be set relatively high (taking account of all the features described above), to retain only those items that are essential, thereby minimizing respondent burden. Typically, modules will have fewer than 30 items, preferably presented within two pages at most.

Where appropriate, the importance and relevance of items may be analyzed in specific subgroups (e.g. a gender only, advanced setting only, etc.) to investigate the applicability of specific items across the intended population.

4.5 Phase 3b: Preliminary Psychometric Evaluations

4.5.1 Conditions required to proceed

1) The output from Phase 3a, which will be the preliminary Phase 3 module and the approved Phase 3a report

4.5.2 Aim & design

The aim of Phase 3b is to carry out some preliminary evaluation of the psychometric properties of the preliminary module. It is likely that, during Phase 2, a number of items were generated that are hypothesized to form one or more multi-item scales. These scales will be tested fully in Phase 4, but preliminary formulation and testing of the hypothesized scales in Phase 3 should be used to identify any necessary modifications of the multi-item scales *before* Phase 4. This is necessary because Phase 4 comprises the testing and validation of the final module and as such, further changes of substance at that stage would then require further confirmatory testing (i.e. a second Phase 4 validation). Thus, during the planning for Phase 3, whenever new (i.e. not already existing in the EORTC QLG Item Library) multi-item scales are hypothesized to exist, preliminary psychometric evaluations are required. The evaluations should be of the preliminary module *resulting from* the preceding patient interviews (Phase 3a). Substantial revisions of items and addition of new items based on these psychometric evaluations should be kept to a minimum. It is also important to consider the possibility of removing redundant items. Relevant and important items that do not form multi-item scales should be kept as single items in the module in order to maintain content validity. This ensures that the psychometric properties of

all/most items entering Phase 4 have been preliminarily evaluated, thus reducing the risk of substantial changes during Phase 4.

4.5.3 Sample

Various rules of thumb can be applied to determine sample size (see Fayers and Machin, 2016). Based on their experience and knowledge of the relevant literature, the authors of these current guidelines recommend the following approach to sample size estimation. Note that the guidance, including numbers, provided in this section is intended to allow module developers to generate a rough estimate of sample size. A named statistician with relevant experience must be part of the Phase 3 module development group, and the module developers should work with the statistician to calculate and justify the actual sample size.

The patient sample size in Phase 3b must be large enough to allow for preliminary evaluation of the psychometric properties, in particular the hypothesized scale structure. Generally, the more complex the module, the larger the sample needed. When judging the required sample size, the following should (as a minimum) be taken into account:

1) The number of items (more items require a larger sample)
2) The number of scales (more scales require a larger sample)
3) Number of items per scale (shorter multi-item scales require a larger sample)
4) Homogeneity within scales (lower correlations among items within multi-item scales require a larger sample)
5) Homogeneity of patient sample (more homogeneity of patient scores requires larger sample).

In order to provide a rough estimate of the number of participants required, start with a minimum of 100 patients and for each of the following criteria fulfilled, add a further 50 patients:

1. >= 20 items in total
2. >= 8 scales (single + multi-item scales)
3. Expect Cronbach's alpha < 0.7 for one or more multi-item scales
4. Expect <= 5% in a response option for one or more items

Hence, if none of the four criteria are applicable, 100 patients may suffice, while if all four criteria apply, 300 patients are advisable.

For example, for a preliminary module with two homogenous scales, each consisting of five items and a heterogenous patient population (i.e. expect patients with different levels of symptomatology) it may suffice to include 100 patients. On the other hand, for a module with 30 items hypothesized to constitute 12 scales, each with 2–3 items and some low–moderate internal correlations, then 300 patients are advised, particularly if the patient population is homogenous.

If the module has not been changed substantially based on the interviews in Phase 3a (i.e. no substantial revisions of items or addition of new items) then the item responses collected for the interviews may be used for the psychometric analysis. Only the number of additional participants required to achieve the desired Phase 3b sample size will need to be recruited.

If substantial changes have been made (major rewording/addition of new items), the item responses from the interviews in Phase 3a cannot be used. The psychometric analysis will have to be based solely on new data.

4.5.4 Method

Any new patients included in Phase 3b will only be required to complete the EORTC QLQ-C30 + the provisional module + the debriefing questionnaire.

Analysis

The standard scoring method should be applied wherever possible: the scale outcome is a linear transformation of the average of all observed responses (if at least half of the responses are not missing) on a 0–100 range so that higher scores reflect either better functioning or worse symptoms. Any deviations from this standard should be clearly identified and justified. For example, instead of the average, a minimum or maximum can be considered if the items cover competing issues.

For multi-item scales, the default option to handle missing item data is that the scale outcome is based on the average of all observed individual responses if at least half of the responses are not missing. For example, in a 4-item scale, where two items are left blank, the scale outcome is calculated using the two observed answers only. For items with an N/A option or items linked to a screening question where the patient responds “No” to having the condition (and therefore the responses linked to the screening question become N/A), the default is to report the number of patients who responded N/A separately. The module development group should consider carefully whether items with a N/A option should be included in a multi-item scale and if they are, how scale scores will be differentiated for those who are able to answer all items and those who respond N/A for any item. Ideally this should be avoided.

Potential problems with floor and ceiling effects, insensitivity or compliance can be revealed using simple descriptive statistics (frequency tables, etc.) for the individual items. This can aid item selection.

Final testing of scale structure, reliability, validity and responsiveness to change over time requires a larger number of patients and is carried out in Phase 4. The psychometric analyses in Phase 3 are more exploratory in nature, exploring the scale structure of the module, while Phase 4 is confirmatory, seeking to confirm the scale structure hypothesized based on Phase 3 analyses.

In Phase 3, reliability of hypothesized scales may be tested using Cronbach's alpha coefficient and correlation-based methods (e.g. multi-trait analysis and item-scale correlations or scale-scale and item-item correlations). It is recommended that these simple methods are supplemented with more complex methods like (exploratory) factor analysis for more in-depth investigation of the hypothesized scale structure. If the hypothesized scale structure is not confirmed, items should be treated as single items, not deleted, in order to maintain the content validity.

Also, score distributions (i.e. skewness, floor and ceiling effects) of the multi-item scales and single items must be examined. It is advisable to exclude scales where the majority of the patients have the same outcome value between patients and within patients over time.

Depending on numbers, some form of validity testing (e.g. known-group comparisons) can be carried out (e.g. patients on and off treatment). Evaluations of differential item functioning with regard to central groups (e.g. the groups used to construct the applied sample matrix and translations) may also be carried out.

At this stage, overlap with the QLQ-C30 items should be investigated. Where appropriate, recommendations may be made that responses to the QLQ-C30 items are integrated into the module scoring structure.

4.5.5 Output from Phase 3a and 3b

The output from Phase 3 is a module (completed Phase 3 module) that will be tested in a Phase 4 field study. Proposals for scoring the module should be specified, with a description of the single- and multi-item scales and observations regarding any scales that are likely to be highly related/correlated. The provisional scoring should take into account the distinction between symptom and functional scales and the handling of missing item responses and conditional questions.

4.5.6 Phase 3 report

A Phase 3a report should be submitted and approved before Phase 3b is initiated. The main purpose of Phase 3b is the preliminary formulation and testing of the hypothesized scales to identify any necessary modifications of the multi-item scales before Phase 4. Hence there should be no substantial changes to item formulations following Phase 3b. If there are any changes to items after Phase 3b, these should be approved before Phase 4 is initiated. If there are no changes to items in Phase 3b, then Phase 3b does not need approval from the PMDC and the Phase 3b report can be submitted as part of the Phase 4 report.

4.5.7 Naming modules

Modules that have completed Phase 3 should be referred to in a standard way.

The module name will be 'EORTC QLQ-' followed by two or three capital letters that will denote the relevant tumour site (e.g. BR for breast cancer, OES for oesophageal cancer), treatment modality (e.g. RT for radiotherapy, CT for chemotherapy), population (e.g. ELD for elderly) or QoL dimension (e.g. BI for body image or SX for sexuality) followed by one or two integers that denote the number of items included (e.g. EORTC QLQ-BR23, EORTC QLQ-OES18).

Exceptions are permitted if justified.

For the update of existing modules, the rule described above is usually followed; the fact that this is a new version of an existing instrument will be specified both on the EORTC QLG website and in the EORTC QLG Item Library. The old and new instrument might have the same number of items; if that is the case please contact the PMDC to agree upon a name that allows distinction between the two instruments.

4.6 Phase 4: Field-Testing

Phase 4 studies may be conducted through the EORTC headquarters or in prospective field studies coordinated outside EORTC HQ. In rare instances, it is possible that data collected from within EORTC or other clinical trials can be used as data for Phase 4 validation. However, if such data are used to validate the module, they must fulfil the same standards and quality as expected from a prospective Phase 4 study and the acceptability, reliability, validity, responsiveness and cross-cultural applicability of the module must be evidenced.

4.6.1 Conditions required to proceed

1) Approval of Phase 3 report

The Phase 3a report must be approved by the PMDC following peer review. If any changes are made to the items in the preliminary Phase 3 module following Phase 3b, a Phase 3b report must also be approved. Where a paper in addition to a report is to be submitted for publication, this must be approved by the EORTC QLG EC prior to submission.

2) Phase 4 protocol

Once a module has completed Phase 3a, a detailed protocol is required to specify the study aims and procedures for field-testing in Phase 4.

A protocol, case report forms and all patient measures should be submitted together. If the study uses retrospective data analysis, a description of the data source(s) must be provided with the protocol. The protocol is peer reviewed either during the process of grant applications if applicable, or by the PMDC and by the EORTC Protocol Review Committee (PRC) when the study is conducted through EORTC HQ.

A standard protocol template is available and should be used to prepare the Phase 4 study protocol. A standard CRF is also available. All of this can be downloaded from the EORTC QLG website.

3) Financial support

In most circumstances, Phase 4 studies are funded by the EORTC QLG and are usually eligible for a fast-track application process. Further guidance can be obtained from pmdc@eortc.org.

4.6.2 Aim & design

The module and its scale structure should be field-tested in a large, international group of patients in order to determine its acceptability, reliability, validity, responsiveness and cross-cultural applicability. Given the international nature of this work, the latter is key.

It is necessary to field-test the module because:

1) The hypothesized scale structure based on the Phase 3b analyses and other psychometric properties needs to be confirmed in a large, independent sample;

2) Completion of the module in Phase 3a is carried out in the presence of a researcher, and the module may perform differently when completed without such supervision;

3) Items should ideally be evaluated in a culturally and linguistically more diverse sample than those typically included in the initial development (i.e. in Phases 1 and 3).

Field-testing consists of:

1) Administering the EORTC QLQ-C30 and the completed Phase 3 module to patients belonging to the target population, but who were not involved in Phases 1 or 3

2) Measuring the time needed to complete the module

3) Completion of a debriefing questionnaire, to determine the **acceptability** of the module, by each patient after completion of the module (see Appendix 7 Debriefing questionnaire for Phase 4 for debriefing questions)

4) Inviting a clinically stable subset of participants to complete the EORTC QLQ-C30 and the completed Phase 3 module again, usually one to two weeks later, to evaluate test-retest reliability. The effect of recollection needs to be considered when selecting the time interval

5) Inviting a subset of participants who have changed clinically to complete the EORTC QLQ-C30 and the completed Phase 3 module again to evaluate responsiveness.

4.6.3 Sample

Participants should represent all groups in the target population for which the module is being devised. A sample matrix should be drawn up to include all relevant treatments (e.g. surgery, chemotherapy, radiotherapy) and patient groups. The module developers may choose to group patients by relevant strata as appropriate for the module. The sample matrix may be similar to that used in Phase 3 (Appendix 1 Matrix for assisting in targeting patient recruitment), but module developers will need to take account of planned known-group comparisons, accessibility of patients in different treatment groups or stages, and the subject matter of the module when planning Phase 4 recruitment.

Phase 3b and Phase 4 should be based on two completely independent samples. This is the optimal and recommended approach. In very special cases, e.g. in rare diseases, participants who took part in Phase 3b could be invited to take part in Phase 4, as long as new data were collected. This is not the preferred approach and it would require careful justification in the grant proposal; or, if this issue is only discovered during the conduct of the module development, it should be discussed with the PMDC as soon as possible.

All enrolled patients must be accounted for in the final report. For patients who do not complete the required forms, reasons for the non-compliance must be recorded to ensure that patient burden is not underestimated. In populations where compliance is expected to be problematic (e.g. poor prognosis), the protocol must identify barriers and proposed solutions. If needed, the targeted sample size should be inflated or adapted to account for missing data.

Sample size

The sample size required will depend not only on the number of items, the number of scales and the magnitude of the correlations, but also on the heterogeneity of the sample. If there are conditional items, this will affect the required sample size: information from literature and the previous phases on the prevalence should provide information on the expected completion rates for these items.

As per Fayers and Machin (2016), it would usually be reasonable to aim to recruit a minimum of 10 patients per item in the module. If IRT is included in the analysis plan, at least 400 patients will be needed.

It is crucially important that the patients sampled are representative of the full range of the intended population: a large sample in which nearly all patients make more or less the same responses is clearly uninformative despite its size.

As cross-cultural applicability of the module is crucial, it is important to have large enough samples from each country to allow for cross-country comparisons, e.g. DIF analysis. It is recommended that at least 50 patients are recruited from each linguistic group if this is feasible and practicable; exceptions for special situations are allowed but should be discussed.

Taking these recommendations together, it is rarely justifiable to include fewer than 300 patients in Phase 4, and it will often be advisable to aim for at least 400 patients, reasonably distributed across groups, countries and the range of the outcomes. Module developers should obtain expert statistical advice before finalizing their sample size.

Cross-cultural consistency

Phase 4 is an international field-test/study. As many countries as practical should be involved (including at least all those participating in Phase 3). No more than 30% of patients should come from any one country.

Translations into languages not tested in Phase 3 must be developed and pilot-tested before they can be used in Phase 4. For more details see chapter 7.

4.6.4 Method

Patients are asked to fill in the EORTC QLQ-C30, the completed Phase 3 module and the debriefing questionnaire.

For the purpose of external validation of the module, additional information should be collected. Dependent on the QoL dimensions assessed, this information should include socio-demographic and clinical data, but also additional instruments assessing relevant QoL dimensions. Since the module will contain items specific to certain groups of patients and/or QoL dimensions, external validation criteria should be specific to the patient groups concerned (e.g. breast conserving therapy versus mastectomy to validate a body image scale included in a breast cancer module). The relevant patient groups and the corresponding comparisons should be identified before starting so that the required data can be collected and an adequate analysis plan can be set up.

Scale structure and reliability

Score distributions (i.e. skewness, floor and ceiling effects) of the multi-item scales and single items must be examined. It is advisable to exclude scales where the majority of the patients have the same outcome value between patients and within patients over time.

A range of analyses can be conducted to test the module's hypothesized scale structure empirically and to establish scale reliability. Multi-trait scaling analysis and confirmatory methods, particularly confirmatory factor analysis, are recommended in Phase 4 in order to test and (hopefully) confirm the scale structure's hypothesized multi-item scales (Fayers and Machin, 2016) based on the Phase 3b analyses. It is recommended that factor analysis methods suitable for ordinal data are used as traditional factor analysis may overestimate the number of factors (Bjorner et al., 2003; Muthen and Muthen, 2010).

The *internal consistency* of the multi-item scales can be assessed by Cronbach's alpha coefficient. Reliability of a magnitude of 0.70 or greater is desirable for group level data (Fayers and Machin, 2016).

More recent approaches to scale construction could also be adopted, including those based on item-response theory (Embretson and Reise, 2000; Fayers and Machin, 2016) and differential item functioning or item bias analysis (Groenvold et al., 1995; Groenvold and Petersen, 2005; Fayers and Machin, 2016) (see *Statistical considerations for validation of the module* below).

To evaluate the module's test-retest reliability or stability, intra-class correlation coefficients are generally used. Thresholds are controversial, but for comparing groups of patients, many investigators regard correlations of at least 0.70 as "acceptable", and those that exceed 0.80 as "good".

Sample size determines the certainty of the estimates, and this determines the confidence intervals.

Prior to conducting the test-retest analysis, check for outliers.

Validity

A range of analyses is available to evaluate the validity and responsiveness of the multi-item scales and single items. For example, known-groups comparison (Fayers and Machin, 2016) can be used to evaluate the extent to which the module is able to discriminate between subgroups of patients with different disease stages, current symptoms, performance status and/or other relevant variables. For multi-item scales, analysis of variance (or a simple t-test) may be used to test for the statistical significance of group differences while for single items, non-parametric or logistic methods may be more appropriate.

For each scale of interest, the known-groups comparison must be pre-specified in the protocol. The known-groups construction must be organized so that the resulting subgroups are distinct, feasible (i.e. sufficient patients expected in each subgroup) and hypothesized to overlap in content with that scale. Power calculations may be provided to show that relevant differences (rule of thumb: 5–10 points) can be reliably found.

The responsiveness of the module can be evaluated by examining differences in scores at different times during the course of the disease or treatment; for example, comparing scores before and during chemotherapy. Changes in scores over time may also be examined in relationship to changes in a criterion parameter such as performance status. For each scale of interest, the responsiveness comparison must be pre-specified in the protocol. The response criteria must be organized so that the resulting change over time is unambiguous, feasible (i.e. sufficient patients expected in each response group) and hypothesized to overlap in content with that scale.

Apart from statistical significance, attention should also be paid to magnitude and precision of the constructed differences. These should be reported via the estimated score differences and their respective confidence intervals, and effect sizes. For clinical significance, see below.

Statistical considerations for validation of the module

Item Response Theory (IRT)

Using IRT is not mandatory but the module developers may wish to consider this as part of the validation stage.

IRT is a useful tool in the selection of items for inclusion or exclusion during Phase 4. IRT is particularly suitable for reducing the number of items to be included, for example if it is desirable to produce a shorter module, or when merging two similar modules (see chapter 5). IRT and item characteristic curves also provide a powerful method for identifying poorly performing items, even when using IRT in its most basic form. At the other extreme, if aiming to develop item banks

for CAT measurement or customized short forms, IRT is prerequisite. IRT analyses should normally be preceded by multi-trait and/or factor analyses, and focus on the analysis of multi-item scales, which can be regarded as essentially unidimensional. IRT requires substantial numbers of patients, typically at least 400. For further details on IRT including Rasch models see Appendix 8 Item Response Theory (IRT) for scale structure and selection of items in Phase 4.

Differential item functioning (DIF)

Differential item functioning (DIF) analysis examines whether items in a scale perform similarly across various subgroups. If one or more items in a scale exhibit DIF, comparisons across groups may not be 'fair' in the sense that patients at the same level of the construct being measured but from different subgroups may get different scores on the scale. It is particularly relevant to examine for DIF in groups that may be compared using the module. This could include different disease stages, different treatments and different translations. Several methods for evaluating DIF exist, including methods based on contingency tables, logistic regression and IRT. Logistic regression has the advantage of being flexible yet relatively simple, and may be used regardless of the scale being scored using sum or IRT scoring. For further details see e.g. Fayers and Machin (2016).

Establishing thresholds and MIDs for clinical importance for EORTC QLG modules

Thresholds for clinical importance (TCIs) have been recommended in the literature (Wintner et al., 2016; Snyder et al., 2019) for facilitating the interpretation of PRO scores. Such cut-off scores can be used to make PRO measures usable for symptom screening in daily practice or for calculating prevalence rates in clinical trials.

Phase 4 may include the development of TCIs for the health domains covered by an EORTC QLG module relying on the methodology applied previously for establishing TCIs for the EORTC QLQ-C30 (Giesinger et al., 2020b) and the EORTC CAT Core measures (Giesinger et al., 2020a). This anchor-based approach relies on three criteria for clinical importance (limitations, need for help, and worries) that were derived from a mixed methods study (Giesinger et al., 2018) in cancer patients and healthcare professionals and on a classification rule to summarize the individual criteria in a binary variable.

The following three anchor items (with domain-specific wording) are recommended for establishing TCIs for EORTC QLG modules:

- Anchor item on limitations: "Has your SYMPTOM/PROBLEM limited your daily life?"
- Anchor item on need for help: "Have you needed any help or care because of your SYMPTOM/PROBLEM?"

- Anchor item on worries: “Has your SYMPTOM/PROBLEM caused you or your family/partner to worry?”

All three anchor items use the response categories: Not at all – A little – Quite a bit – Very much.

A patient is categorized as a case, i.e. as having a clinically important problem/symptom, if (s)he selects “quite a bit” or “very much” on at least one of the anchor items.

For multi-item scales in EORTC QLG modules covering different types of symptoms (e.g. clinical scales such as the QLQ-PR25 hormonal treatment-related symptoms scale) it may be useful to establish TCIs not only at scale-level but also for the individual health issues covered by this scale (i.e. at item-level).

Statistical analysis for establishing TCIs should rely on receiver operating characteristic (ROC) analysis, using the scales as predictors and the binary criterion of clinical importance for definition of cases/non-cases.

To improve the interpretability of scores derived from a new module, the definition of minimal important differences (MIDs) between groups or minimal important changes (MICs) over time may be included in Phase 4. For more information and guidance, feel free to send an email to pmdc@eortc.org.

Item reduction

The completed Phase 3 module to be field-tested could contain more items than is desirable. This problem may be avoided if adequate numbers are recruited in Phase 3 and module developers apply appropriate thresholds for inclusion of items. Nevertheless, on the basis of the data collected in Phase 4, elimination of some items may be warranted on psychometric grounds, e.g. an item may be redundant to other items in a scale or it may not fit well in a hypothesized scale. Clinically important items should be kept even if they do not fit in a multi-item scale. Results from the debriefing questionnaire should also be used to guide item reduction. The time required to complete the questionnaire should be kept practical. If some patients require substantially more time than anticipated, possible explanatory characteristics (e.g. old age, cognitive problems) should be investigated. Items that are reported as upsetting or confusing must be considered for deletion but must be weighed against their relevance as assessed in the previous phases.

It is important to provide the rationale for dropping an item.

4.6.5 Output from Phase 4

A module will be considered as **validated** when the EORTC QLG EC has approved the Phase 4 report.

A validated module should consist of a single version of the module itself and a scoring/interpretation guideline.

The module development process will be considered completed and be uploaded to the EORTC QLG website once the PMDC is informed that it has been accepted (in press or online) for publication.

Publications

All publications originating from a module update will be prepared on behalf of the EORTC QLG. It is advisable to discuss authorship before project initiation: a publication/authorship policy is recommended. Please see recommendations on authorship in section 9.1.3.

5 UPDATE OF AN EXISTING EORTC QLG MODULE

5.1 Introduction

In oncology, there are ongoing changes and advances being made in treatment and its evaluation. The introduction of novel systemic agents (including immunotherapy, biological and targeted agents), and new radiotherapy and surgical techniques means that EORTC QLG modules may become partially obsolete or require additional items to cover side effects or benefits associated with novel approaches. In addition, in the course of widespread use of an EORTC QLG module in clinical trials and other research settings, psychometrically weak items or scales in existing modules may be identified. Consequently, a module may need an update to ensure that it addresses key QoL issues relevant to current treatments and/or to update scales or items in the original module with weak clinical or psychometric properties. The EORTC QLG proposes that the following methodology is employed when updating existing EORTC QLG modules.

Rationale for module update

Before initiating an EORTC QLG module update, the rationale for why an update is needed, as well as the target population for which the module will be updated, should be established.

- The rationale can be based on evidence suggesting that important clinical domains and symptoms are not included (in whole or partially) in the original module because treatments or substances with substantial new side effects or negative impact on functioning have been introduced since the original module development.

- Another reason for a module update can be methodological issues reported from the original module's use (e.g. reported issues on not sufficiently covering side effects of current treatments, psychometric issues, cross-cultural use, compliance by patients in terms of missing values, compliance by researchers in terms of scales used and reported).

- Furthermore, the rationale for a module update could also include reference to a changing cultural landscape. The update may also improve the body of evidence for cultural suitability of the module and to improve its ecological validity (i.e. the module is relevant in a broad range of cultural contexts).

It is preferable to keep the items and scales from the original module if appropriate, in order to be able to compare new data with historical data.

Application procedure and decision

At the beginning of the application process for an EORTC QLG module update, the project needs to be discussed with the EORTC QLG (e.g. in the PMDC session, at the biannual QLG meeting), the EORTC, and the relevant DOG first.

Those who contributed to the original module should be contacted whenever this is possible, and offered the possibility of contributing to the new project (see section 9.1.3). Establishing contact can be facilitated by members of the PMDC (contact pmdc@eortc.org for further queries).

The decision on whether to proceed with the proposed module update will be made on a case-by-case basis based on the rationale presented at the project pitch. Applicants are advised to conduct a thorough and comprehensive literature review, focussing on potential methodological problems reported concerning the use of the original module or a clear need for updating the module (i.e. due to considerable changes in the treatment landscape). If the evidence presented is not sufficient for a decision, the applicants will be asked to conduct a literature review (pre-Phase 1) which might be built upon in Phase 1.

Phases of module update

The process of updating an EORTC QLG module mirrors the development process and consists of four phases:

Phase 1: generation of new issues related to the new treatment and identification of problematic items and/or scales in the original module

Phase 2: creation of a revised module by conversion of new issues into items and changing the wording of problematic items

Phase 3: pre-testing the new module

Phase 4: international validation field-testing

Module developers should refer to chapter 4 of these guidelines for more detailed guidance on the different phases of module development.

5.2 Phase 1

Phase 1 is aimed at compiling an exhaustive list of QoL issues relevant for the patient group under investigation and covering the new treatments identified in the research question. In the process of compiling the list, the following sources are used:

1) Literature, including a) qualitative research on the patient perspective, b) 'grey literature' from the pharmaceutical industry such as investigator brochures on new drugs to obtain potential pharmaceutical toxicities, and c) existing questionnaires. Special considerations should be taken to review all available literature on the original version of the module

2) Interviews with healthcare professionals

3) Interviews with patients from the target population.

In addition, any problems that have arisen in the use of the original module, which may require modification, should be identified (via e.g. literature review, experience by developers of original module).

5.2.1 Methods

Literature review

When updating an existing QoL module, the purpose of the review is two-fold:

1) Updating the list of issues that cover 'new' treatments, etc. that have emerged since the original module was developed, giving different QoL implications for patients

2) Summarizing the use of the EORTC QLG module in trials/ studies and issues that emerge, e.g. in relation to its content validity or other methodological issues.

Consequently, two separate literature searches should be conducted to update the literature review from the original Phase 1 development.

Literature review 1 – review to identify new concepts: The first literature review is aimed at identifying all qualitative studies that report the potential QoL issues associated with new treatments and to provide a list of additional new issues. Recent literature reporting on the patient perspective should be reviewed and considered. This would serve to collate the current understanding of the disease area from the patient's perspective prior to conducting further interviews. With the focus on updating an existing EORTC QLG module, the purpose of this review is not to create a list of QoL issues from scratch but to check whether issues within the module are still relevant and important, as well as to capture new QoL issues as a result of changes to treatment and management which are not covered by the current version of the module. The literature search should also encompass the search for questionnaires on the respective disease. Questionnaires should be scrutinized with regard to symptoms and quality of life issues relevant for patients of the respective population.

Literature review 2 – review how the original module has been used and has been performing to identify how the module may be improved: The second literature review should identify all

studies and clinical trials that have used the original EORTC QLG module. Tables should be created to summarize the studies and potential methodological problems with the module. It is desirable to tabulate data concerning score distribution, psychometric properties of scales and any qualitative information reported. Information about missing data from particular scales and items, and details of how long the modules took to complete, and patients' responses to them, may be useful. The tables may also include information from the publications about reported problems experienced by users of the module. This information will be useful for later decisions about which scales need to be changed.

Module developers who wish to publish their review should refer to section 4.1.4.

From the literature review, a list of new issues will emerge, containing additional issues thought to be potentially relevant for patients and additional issues not included in the existing module.

Interviews

The above-mentioned searches will result in an extensive list of issues. Interviews should be undertaken to discuss and consider the potential new issues suggested from the literature and to discuss potential changes to existing scales and items. As the advice for interviews is much the same as that given for developing a new module, module developers should refer to section 4.1.4 of these guidelines for more detailed guidance on conducting interviews with patients and healthcare professionals. The guidance below is specific to updating modules.

Interviews with healthcare professionals

The recruitment of healthcare professionals should follow the same guidelines set out in section 4.1.3.

The list of QoL issues, made up of the original module and any new issues, may be administered in the form of a structured, personal interview in which basically two questions are asked: (a) are issues included which the specialists consider irrelevant for this patient group, and if so, why do they consider these issues irrelevant?; and (b) are there issues missing from this list that the specialists consider relevant, and if so, why do they consider these issues relevant?

To establish the relative importance of the QoL issues, the specialists should be asked to identify a subset (e.g. a quarter) of issues that, in their opinion, affect patients' QoL most profoundly and which should definitely be included in the final module. Appendix 2 Generation of relevant QoL issues in Phase 1: provides an example of a detailed interview protocol. New issues identified by healthcare professionals during these interviews should be added to the issue list but at this stage researchers should not remove any issues, in particular those from the original module, before patients have been consulted.

Patient interviews

Interviews should be conducted with at least 20 patients who were not involved in the development of the first version of the module. Sample size estimation is based on the patient matrix and stratification. It is recommended that 5–10 patients should be interviewed from each cell in the matrix. Investigators are advised to aim for a well-balanced and representative sample with regard to the new treatment or condition. Interviews should continue until no new issues arise, i.e. data saturation is achieved.

Patients may be interviewed before, during or after treatment. Investigators should ensure that patients receiving new treatment strategies are well represented in the patient sample. Similar numbers of patients should be recruited from each country participating in Phase 1 with no more than 30% from a single country. See 4.1 for more details.

Procedure

Healthcare professionals and patients will receive a list of issues, combining the two issue lists, the 'old' module and the 'new' issues. The list of new or modified QoL issues may be administered in the form of a semi-structured face-to-face interview. If possible, interviews should be recorded for the purpose of reporting accuracy and audit. When this is not possible, careful field notes should be taken at the time of the interview.

5.3 Phase 2

The new issues are converted into items as described for Phase 2 of development of new modules (see section 4.2). For the sake of consistency and whenever possible, items from the existing EORTC QLG module will be used or any new items will be taken from the EORTC QLG Item Library.

Considerations in this phase need to be based on a) how extensively the original module needs to be changed, and b) how many new items relevant to the new treatment or the changing cultural landscape need to be added. Based on these considerations, investigators need to decide if:

1) Scales or items in the original module will be updated (i.e. an updated module with new and old items combined, some old items removed or changed); and/or

2) Additional items relevant to the new treatment from the EORTC QLG Item Library are to be added (i.e. retain the old module but include additional items or add a module relevant to a new treatment).

Module developers are encouraged to discuss decisions with the PMDC. When it is decided to include relevant symptom items or to add a module from the EORTC QLG Item Library, guidelines on its use[7] need to be complied with. The module developers are advised to contact the EORTC QLG Item Library team for further assistance (itemlibrary@eortc.org). Please refer to section 4.2.3 for further information on the EORTC QLG Item Library.

For some of the original modules, the introductory text and the wording of the original items may also need to be modified. It is advisable to consult the QLD Translation Unit at this stage about possible improvements to original wording.

After this stage, the procedure will be formally peer reviewed by the PMDC. After approval, the provisional list of items is ready for Phase 3.

5.4 Phases 3 and 4

The pre-testing and validation of the updated module will follow the standard guidance as given for new module development (see sections 4.3 and 4.4). In addition, it may be appropriate to compare compliance and acceptability of the new version with the previous version.

When publishing the new module, it is important to clarify which scales of the original module are unchanged and which items have been added to make it easier for users to compare studies using the original and the new module. For publishing additional items chosen from the EORTC QLG Item Library, please refer to the EORTC QLG Item Library User Guidelines[7].

Publications

All publications originating from a module update will be prepared on behalf of the EORTC QLG. It is advisable to discuss authorship before project initiation. Please see recommendations on authorship in section 9.1.3.

[7] https://www.eortc.org/app/uploads/sites/2/2018/09/IL-manual-20180305.pdf

6 COMBINING/MERGING TWO MODULES

6.1 Introduction

The EORTC QLG modules provide organ- or treatment-specific assessment tools for QoL in a wide range of tumour sites. Occasionally it may be appropriate to combine two existing modules.

Examples of existing combinations of modules are the oesophago-gastric module EORTC QLQ -OG25 (from a combination of the oesophagus and stomach modules, EORTC QLQ-OES18 and EORTC QLQ-STO22) and the cholangiocarcinoma and gallbladder cancer module EORTC QLQ-BIL21 derived from the pancreas and hepatocellular carcinoma modules (EORTC QLQ-PAN26 and EORTC QLQ-HCC18).

Arguments for combining two or more modules should be provided, and should be based on empirical data including clinical and psychometric results from previous use of the modules in question. Factors to consider include: the degree of overlap of symptoms and side effects, and thereby the items in the modules of the two or more tumour sites; the extent of similarity of progression of the diseases; and the potential side effects of treatments offered. Treatment may have changed over time and patients represented by two or more modules may be included in the same trials. In each case, it is important to consider whether it makes clinical sense to try to combine modules or whether it would be best to start a completely new module.

Combining modules is not necessarily easier than starting from scratch but may have the advantage of using questions that have been tested and studied using psychometric tests as part of a Phase 3 or Phase 4 study. A frequently used module may be more important to retain than a module with little existing data.

Approval from the EC should be obtained before starting the work. It is necessary to demonstrate the need for a combined module, and that it is appropriate for the two or more modules concerned. The target population of interest should be clearly defined. The original developers of the candidates for module merging should be consulted, and preferably be involved in the process.

6.2 Phase 1

A literature search should be performed with two different aims:

1) To collect information on the clinical and psychometric validity of the existing modules

2) To collect information, using all relevant terms, relating to the new diagnosis/organ. All issues arising should be listed as described for Phase 1 development.

The two existing modules should be reviewed and all the questions combined into a single set of logical, clinically sensible groupings (probably, but not necessarily, corresponding to scales of the existing modules). Some existing scales will be combined in this process. In addition, any new issues arising from the literature search should be included in these groupings of items. This may result in item groupings with a combination of 'issues' and 'questions' which may be difficult to work with. Some patients and healthcare professionals may be confused by the variation between items (questions) and issues (described features of QoL) or may have a preference for one format over the other which could bias responses. Therefore, the issues should be converted into questions/items at this stage, if possible using items from the EORTC QLG Item Library that have been used previously in validated modules, and as such will have been translated. This approach differs from the procedures for developing new modules and for updating an old module.

This module should be evaluated by patients and healthcare professionals as described for Phase 1 of new module development (see section 4.1).

6.3 Phase 2

Because the majority of items are derived from existing questionnaires, and additional issues are already framed as items, Phase 2 is relatively straightforward. Decision rules for inclusion and exclusion of items should be agreed, and a final list of items derived. Special care should be given to explore critical comments in the literature on each of the existing modules.

After removing unwanted items, the original item groupings may be used as 'hypothesized scales' or they may be rearranged into clinically meaningful scales with additional individual items if necessary. At this point the module developers should decide whether a new module is needed at all, or whether using one of the original modules would suffice. If the old module/items are equally relevant as new items, the old version should be retained in order to be able to compare results with previous publications. For any new issues, the EORTC QLG Item Library should be referenced for any existing items and multi-item scales to simplify the process.

It is recommended that this decision should be discussed with the PMDC. A report for Phases 1 and 2 must be approved before the work can progress to Phase 3. The final list of items should be sent to the QLD Translation Unit for translatability assessment (if any changes are necessary) and formatting.

6.4 Phases 3a and 3b

The provisional module should then be tested in a further sample of patients of all relevant stages of the disease and from different countries (see chapter 4). Standard psychometric tests may be applied to the results to check correlation of questions and internal validity of the module. A more limited approach may be used if changes from old modules are small.

6.5 Phase 4

A field study should be carried out as per the recommendations outlined for new module development (see section 4.6). Special care must be taken to adequately represent the intended target population. Consistency of the results within the populations as defined in the original modules is strongly advised.

Publications

All publications originating from a module update will be prepared on behalf of the EORTC QLG. It is advisable to discuss authorship before project initiation. Please see recommendations on authorship in section 9.1.3.

7 LANGUAGE AND TRANSLATION

All translations must be discussed with and coordinated by the QLD Translation Unit.

Module development should be conducted simultaneously in several languages and cultural groups. This enables the cultural and linguistic validity of the newly developed measure and its translations to be achieved.

7.1 Language

The working language for module development is British (UK) English. For presentation to patients, issues and items will be translated into the patient's own language, following the EORTC QLG Translation Procedure.

The English version of the module should be considered the source version for all modules even though the PI might come from another language group. All translations should originate from the English version (except for cultural adaptations, which are carried out from the basis of an existing translation developed from English). Publications or presentations of the module should use the English language labels for the items and/or scales whenever appropriate.

The language groupings used by the QLD Translation Unit are:

a) English-speaking countries (UK, US, Canada, Australia, etc.)

b) West-Germanic languages: German (Germany, Austria), Dutch (the Netherlands, Belgium)

c) Scandinavian languages: Danish, Swedish, Norwegian, Icelandic

d) Slavic languages: Polish, Czech, Slovak, Russian, Belarusian, Ukrainian, Slovene, Serbian, Croatian, Macedonian, Bulgarian

e) Romance languages: French (France, Belgium), Italian, Spanish (Spain), Portuguese (Portugal), Romanian

f) Other European languages: Greek, Hungarian, Finnish, Lithuanian, Latvian, Estonian, Welsh etc.

g) Non-European languages: Turkish, Hebrew, Chinese (China, Taiwan, Singapore, Malaysia, Hong Kong), Spanish (Spanish-speaking Latin American countries), Portuguese (Brazil), etc.

7.2 Translation procedure

The modules should undergo a rigorous translation process, based on an iterative forward-backward procedure. The process is described in detail in the EORTC QLG Translation Manual. For help with any questions or problems, developers can contact the QLD Translation Unit.

The aim of translation is to produce modules which are clear, expressed in language of common use and conceptually equivalent to the original module.

Module developers should consult the QLD Translation Unit before starting any translation work.

Phase 1 (Creating a list of issues)

For the collection of issues, developers must initially consult the EORTC QLG Item Library. Developers should use existing items if available. There are around 1,000 items in the EORTC QLG Item Library and it is important to make sure that duplication is avoided. Developers must select the recommended wording for each item if multiple options are available for a single issue/item, as highlighted in the EORTC QLG Item Library.

If translations of issues are required, these are prepared by the developers and collaborators – there is no involvement of the QLD Translation Unit in the translation process, except for general advice.

Phase 2 (Transforming issues into items)

For the phrasing of issues into items, developers should consult the EORTC QLG Item Library. There they will find not only suitable formulations of items in English, but also translations in a number of languages.

Once the final list of items is ready, it should be sent to the QLD Translation Unit for translatability assessment. The goal of the assessment is twofold:

1) To make sure that the English items are worded in a correct, consistent way, using standard, tested structures that have been proven to perform well in other languages

2) To avoid translation problems related to difficult syntax, terms, cultural issues, etc. A non-exhaustive list of words and structures to be avoided is available in Appendix 9 Examples of item structure and word selection.

All suggestions are then sent to the module developers. Feedback on the items should be sought from everyone involved in the module's development, including a patient representative if

possible, and other colleagues. Further discussion may be required with the QLD Translation Unit. The approved items are then formatted in the standard template for translation.

The QLD Translation Unit coordinates the translations into all languages needed for Phase 3. The standard translation procedure will be applied as specified in the EORTC QLG Translation Manual. The translations are developed by a translation agency specialized in PRO translation and coordinated by the QLD Translation Unit. The costs of the translation process are covered from the general EORTC QLG budget for translations and therefore do not have to be included in the specific study budget. However, the translation process takes at least 8 weeks and should be included in the timelines of the module's development.

Phase 3 (Testing the new module in patients; interviewing patients regarding critical items)

The preliminary module that has been generated in Phase 2 is pilot-tested on patients in Phase 3a and 3b.

In Phase 3a, patients complete the module, and are then interviewed in depth, to determine items that are difficult to understand, upsetting, or not relevant to or reflective of their experience of their disease or treatment. Thus, Phase 3a can be regarded as the pilot-test that is required as an integral part of the EORTC QLG translation process.

The results of Phase 3a should be summarized by country and language and sent to the QLD Translation Unit for review and archiving. All changes resulting from the interviews should be discussed with the QLD Translation Unit before including them in Phase 3b. This is to reduce the likelihood of any further changes to the items or translation after Phase 3a. Changes made to items depend on the comments of patients. These can be classified as:

1) General comments on the wording, requiring changes to the English source and all translations

2) Language-specific comments, requiring changes to a particular translation.

If necessary, the QLD Translation Unit will prepare the new version of the module following Phase 3a. In Phase 3b, a preliminary evaluation of the psychometric properties of the provisional module is carried out so it is unlikely that further changes will be required to the wording of the items. However, if there are changes required after Phase 3b, these must be discussed and classified as above. An approved Phase 3b report will then be required before Phase 4 can proceed. Once the Phase 3b report has been approved, the QLD Translation Unit will prepare the new version of the module for Phase 4. The extent of changes required determines the time needed to prepare the module for Phase 4. This should be discussed with the QLD Translation Unit and included in the timelines of module development.

Translation and preparation of translations for Phase 4

Once the Phase 3 report has been approved and the list of collaborators for Phase 4 finalized, the module developers should contact the QLD Translation Unit to discuss the preparation of translations. The QLD Translation Unit coordinates the changes approved as a result of Phase 3 and prepares the translations tested in Phase 3 for use in the Phase 4 study and for external users (as modules under development). These translations are considered to be linguistically validated, since they followed the EORTC QLG translation process and were pilot-tested on patients in Phase 3.

Translations needed for Phase 4 but not included in Phase 3 have to be developed from scratch. These translations should be included in the Phase 4 study budget. In order to receive a cost estimate, the module development group should contact the QLD Translation Unit.

The Phase 4 translations are developed by a translation agency specialized in PRO translation and coordinated by the QLD Translation Unit. Before they can be used in the Phase 4 study, they should be pilot-tested by the module development collaborator(s) on 5–10 patients in order to check the comprehensibility of the new language version. The pilot-testing results must be sent to the QLD Translation Unit for review.

Module developers should consult the QLD Translation Unit before starting any translation work.

8 PREPARATION OF DOCUMENTS

The lead PI is responsible for ensuring strict adherence to ethical guidelines, data protection regulation, research governance, quality assurance, data management and statistical analysis procedures and the rigorous documentation of these. Documentary evidence of these aspects should be submitted to the PMDC with the Phase 1/2 report, the Phase 3a report and Phase 3b report if needed, and finally the Phase 4 report, and this will be archived at the QLD in case of future regulatory inspection. Templates are available for all phases and their use is mandatory. They can be requested via email from pmdc@eortc.org.

The availability of detailed documentation relating to module development serves two purposes:

1) To inform all interested members of the EORTC QLG

2) To provide a record of independent peer review.

Reports or papers need to include information about the sample (inclusion and exclusion criteria, recruitment procedures), data collection procedure and results (e.g. scale structure, internal consistency reliability, stability, clinical validity, and responsiveness).

The documentation of the entire module development process for each module will accumulate during the development process and will include the following documents each submitted at the appropriate time:

- A proposal of the planned module including its objectives and the multi-disciplinary and multi-cultural involvement of contributors
- A brief written report for each biannual EORTC QLG meeting describing the progress since the last meeting
- Ethics approval, protocol, informed consent form, and list of PIs and sites
- The preliminary module after completion of Phase 2, with a table of issues and resulting items produced
- Two reports on the module development process in Phases 1 and 2, and Phase 3 (the latter may be a paper for publication combined with the appendices needed to provide full details)
- Reports on the translation and pilot-testing of the module in each language separately
- A report on the procedures and results of large-scale field-testing (Phase 4; usually as a scientific paper combined with the appendices needed to provide full details)

- Reviewers' comments on each of the three module development reports and translation processes, and the PI's replies
- Copy of the cleaned anonymized minimum dataset (final database and locked).

All phases of the module development process are described in a flow chart in Appendix 10.

Module construction is a sequential, step-wise process in which a new phase cannot be entered into unless the previous phase has been successfully completed (this does not apply to Phases 1 and 2, which are usually conducted together). Permission to proceed to the next phase is based on the approval of the previous phases, for which several documents need to be prepared.

All documents should be written in English.

A copy of all final reports should be provided to the PMDC and will be stored at the QLD for archiving, research governance and data sharing.

8.1 Proposal

Before initiation of a module development project, the proposal should be discussed at the PMDC session of the EORTC QLG meeting to ascertain that the module is within the remit of the EORTC QLG and it does not overlap with other module development projects. If there is overlap, the work needs to be coordinated to avoid duplication of effort.

The multi-national, cross-cultural, and multidisciplinary composition of the EORTC QLG and the experience in module development accumulated by its group members enables important scientific and cultural input into the development of new modules.

Investigators who wish to develop a module should:

- Send an email to pmdc@eortc.org and ask for the short outline of new projects
- Fill out the short outline of new projects and send it back by the provided deadline
- Present at one EORTC QLG meeting
- Receive endorsement from the EORTC QLG EC.

8.2 Phase 1 to 3 Protocol

The writing of the Phase 1 to 3 protocol has already been extensively discussed in the previous chapters. The EORTC Clinical Operations Manager will contact the PI(s) to ask for the following documents: ethics approval, protocol, informed consent form and list of PIs and sites.

8.3 Phase 1/2 Report

Usually, no report is required after Phase 1 alone. However, a Phase 1 report may be required by the PMDC and the EORTC QLG EC on a case-by-case basis (for example, when providing support on the need to update or merge a module).

After completion of Phase 2, a Phase 1/2 report must be submitted to the PMDC for review. This will ensure that the development process has been conducted satisfactorily and that identical wording is used in newly proposed modules for those items that are similar in content.

The Phase 1/2 report will contain detailed information on literature searches, qualitative interviews, and the rationale for selection of the draft list of issues for presentation to patients in Phase 3. Furthermore, it should include a table of issues with the resulting list of items, before the translatability assessment is carried out by the QLD Translation Unit and the final formatted module. A template with sub-headings and directions on the required information to report on is available and should be followed when writing the Phase 1/2 report.

The report will be reviewed by the PMDC, at least two other members of the EORTC QLG, and staff from the QLD and/or the Statistical Support Group (SSG) if needed. Approval of the report is required before progress to Phase 3. If the submitted report is a manuscript for publication combined with the appendices needed, an additional EC member will also review the report. Review of draft publications should preferably be completed within four weeks. Developers should not submit their paper for publication until it has been approved by the EC.

Publications should include in the author list *'on behalf of the EORTC Quality of Life Group'*.

8.4 Phase 3a and 3b Reports

After completion of the first three Phases (generation of QoL issues, creation of a provisional module and pre-testing) Phase 3a and 3b reports are required. A Phase 3a report should be submitted and approved before Phase 3b is initiated. If there are any changes to items after the pre-testing of the hypothesized scales in Phase 3b, these should be approved by the PMDC before

Phase 4 is initiated. If there are no changes to items in Phase 3b, then Phase 3b does not need approval from the PMDC and the Phase 3b report can be submitted as part of the Phase 4 report.

Phase 3a and 3b will describe the results of pre-testing and will outline the issues covered in the draft module. All changes made to the wording as a result of the testing should be described. Deviations from the guidelines and the reasons for deviations should also be reported.

The report will be reviewed by the PMDC, at least two other members of the EORTC QLG, and staff from the QLD and/or the SSG if needed. Review of draft publications should be completed within four weeks. Developers should not submit their paper for publication until it has been approved by the EC.

Publications should include in the author list '*on behalf of the EORTC Quality of Life Group*'.

8.5 Phase 4 Protocol

A protocol needs to be prepared for Phase 4, using the standard protocol template which can be obtained by sending an email to the PMDC.

A standard CRF is available as well and can be obtained the same way.

8.6 Phase 4 Report

The final international field-testing (Phase 4) should be written up in a report for the PMDC combined with the appendices needed. The variations possible in a field study, and in the evaluation of scale structure, make it difficult to be prescriptive about the requirements for a Phase 4 report.

Module developers must submit the report to the PMDC. The report will be reviewed by the PMDC, QLD/SSG staff if needed, and all EC members. The review should be carried out before submission for publication, in order to benefit from rapid constructive comments from the EORTC QLG.

Publications should include in the author list '*on behalf of the EORTC Quality of Life Group*'.

8.7 After Validation

Inclusion of data from the final module within the data repository.

Lead PIs are asked to provide the PMDC staff within the QLD with the documents which are listed below plus a copy of the cleaned anonymous minimum dataset (final and locked).

Lead PIs should ensure data quality. Shared data should be anonymized, where anonymized data are defined as data from which the participant cannot be identified by the recipient of the information.

Within three months of the module development process being concluded (i.e. three months from when the Phase 4 report is approved, unless a different agreement is in place) the lead PI is responsible for sending to the person responsible for the data repository:

- A copy of the final Phase 4 protocol
- A blank copy of the Patient Information Sheet
- A blank copy of the Clinical Research Form
- A copy of the cleaned anonymous minimum dataset (final and locked).

8.8 Progress Reports

For each biannual meeting of the EORTC QLG, a brief follow-up report of module development is required, describing the progress since the last meeting. This should include any milestones that have been reached, information on data collection, the expected timescales, and actions being taken in case of delays in progress. Possible deviations from the guidelines and problems that may have been encountered should be raised, and any planned or published publications – either papers or conference abstracts – and new translations that have become available since the last EORTC QLG meeting should be indicated.

The PMDC must be notified of all published papers for addition to the list of EORTC QLG publications on the EORTC QLG website.

The PIs will be contacted by the PMDC staff at the QLD with the request to send the update on their module(s) in development prior to each EORTC QLG meeting. The reported progress will be reviewed by the PMDC and may be discussed at the PMDC meeting or briefly presented in the plenary meeting of the EORTC QLG. A report on the progress of all the modules in development will be made available to EORTC QLG members on the EORTC QLG website and a summary of the

discussions will be circulated in the minutes of the EORTC QLG meetings. The information on each module provided on the EORTC QLG website will be updated after each meeting.

If a module does not progress to Phase 4, the reasons for this will be documented on the EORTC QLG website. The EORTC QLG reserve the right to commission other module developers to complete a module where an original module developer has confirmed they do not wish to take the module development forward and/or does not engage with the EORTC QLG for a significant period (> 3 years) of time.

Modules that are not developed within an agreed time period may be withdrawn from the portfolio.

9 PUBLICATION AND OWNERSHIP

9.1 Publications

9.1.1 General rules

All publications of EORTC QLG modules in development must be reviewed and approved by the EORTC QLG EC.

'On behalf of the EORTC QLG' should be included in the author list if:

- **The EORTC QLG has funded a project, or**
- **Investigators would like to use the claim 'on behalf of the EORTC QLG' in their publication.**

To use the label 'on behalf of the EORTC QLG', these articles MUST be reviewed and approved by the EORTC QLG EC before being submitted for publication.

The items from modules that have only completed Phase 2 or Phase 3 may not be published. Descriptions of the module development may be published, including a description of the issues contained in the module, but these publications should not contain the text of the questionnaire.

No restrictions are made with respect to the publication of the text with items and layout of Phase 4 modules. However, all publication of modules should carry the EORTC logo, and copyright must be asserted.

Authorship

In accordance with the International Committee of Medical Journal Editors, each author on an EORTC publication should have participated sufficiently in the work to take public responsibility for the content. All other contributors who do not meet sufficient criteria for authorship will be acknowledged in the publication (see EORTC publication policy, POL 009[8]).

In the case of module development, the authorship should include members of the module protocol writing committee, and also investigators who recruit > 10% of the eligible patients (i.e. fulfilling the criteria of the pre-defined sample matrix).

These recommendations can be adapted by each module author (e.g. in cases of needing many countries, investigators who recruit over >5% of eligible patients may be considered). It is

[8] https://www.eortc.org/policies-guidelines/

strongly recommended that these authorship rules are discussed and agreed at the start of the module and documented in the protocol.

All investigators who contributed patients to the study (i.e. clinicians) or contributed scientifically to the study (i.e. collaborators from those same institutions) are to be acknowledged in the publication. The acknowledgement list should include the name of all participating institutions and the name of the clinicians and other scientists involved with the study at that institution. Whenever a study participant has moved from one institution to another in the course of the study, that participant is listed with the institution to which he/she was affiliated at the time of starting his/her participation in the study, with the mention '(now at [new affiliation])' (EORTC POL 009).

Publications describing the development process of module development should include in the authors list '*on behalf of the EORTC QLG*' and should be approved by the EORTC QLG EC before submission for publication.

Publications reporting Phase 4 studies should include 'EORTC', 'Phase 4' and 'field study' in their title or key words to facilitate identification.

9.1.2 Publication rules when using Phase 3 modules

When researchers other than the module developers use Phase 3 modules, the following rules for publication of the research apply:

1) The module itself may not be published other than by its developers

2) The module developers should, in principle, have the right to publish their data first. However, if this is not possible, publications should be negotiated on a case-by-case basis

3) Collaboration between the PI(s) of the module and its users is advised with respect to the scoring and scale structure of the module

4) At least one developer of the module should be offered the opportunity to be involved and thereby become a co-author on publications that include information on the psychometric performance of the module

5) The module developers and other researchers should agree in advance on the required access of the module developer to the data derived from the module, and such socio-demographic/clinical data as would be necessary for the purpose of psychometric or clinical validation.

If a module is funded by multiple organizations (e.g. PI applies for general academic grant; developer receives funding from commercial party; local funding to support trial in single institution or country; patient interest group provides monetary or material support) all developers should provide a conflict of interest statement. If a developer has a potential conflict of interest during the module development (receives funding or obtains commercial position) he or she should declare it and this should be reviewed by the EORTC QLG EC.

The lead module developer has the duty to remain available to answer questions about the module for a period of at least three years after publication.

9.1.3 Update of an EORTC QLG module

Authorship

Developers who contributed to the original module should be contacted and offered the possibility to contribute to the new project. If they are interested in collaborating, they should be offered the possibility to contribute to the publications. If they prefer not to collaborate, they should simply be acknowledged. In this last case written confirmation that the original lead module developers do not wish to be included as co-authors is required. If obtaining this written confirmation is not possible, justification should be provided (this applies, for example, to the case where the original module developer is not active in the field anymore).

9.1.4 Combining two modules

Authorship

The same procedure should be followed as detailed in 9.1.3.

9.2 Ownership

The modules developed under the auspices of the EORTC QLG are the copyright of the Group in all stages of development. A users' agreement and copyright procedures will follow those drawn up for the core questionnaire.

The term '*EORTC QLG module'* is reserved for modules which fulfil the criteria detailed in these guidelines. Publication of the development process should include in the title reference to the phase of development being reported, for example: '*Phase 1 to 3 testing of an EORTC QLG Module… [specify purpose or tumour type]'* and the name of the module (e.g. EORTC QLQ-BR23), and should always end the author list with '*on behalf of the EORTC Quality of Life Group'.*

Researchers who develop modules to supplement the EORTC QLQ-C30 that do not meet these criteria are **not permitted** to use the term '*EORTC QLG Module'* and should **explicitly** state that the resulting module cannot be regarded as an official EORTC QLG module.

9.3 Using a Module in Research

Modules that have completed Phases 1 and 2 are not suitable for primary research and should not be used. Information about the development process may be published; unpublished material can be obtained directly from the module developers.

Modules that have completed Phase 3 are not freely available, but may be obtained from the QLD. Users of 'completed Phase 3' modules are advised to perform psychometric analysis of their data prior to undertaking the analysis of their main study data, for example, calculating Cronbach's alpha coefficients to ensure that the module is performing as expected.

The module developers should, in principle, have the right to publish data from psychometric analyses before any users of the module do. However, if this is not possible, publications should be negotiated on a case-by-case basis.

The EORTC QLG will hold the copyright of the provisional modules. The module developers shall have a royalty-free license to use and validate the provisional modules for non-commercial purposes until completion of Phase 4.

If researchers want to use Phase 3 modules, they may do so if:

1) They have received explicit permission from the QLD

2) They leave the module's integrity intact and do not revise items. However, if they want to add items as an addition to the module they may do so after consulting the EORTC QLG Item Library researcher at the QLD

3) They agree, if requested, to contribute a minimum clinical dataset for purposes of the psychometric/clinical validation of the module

4) They use the hypothesized scale structure as agreed by the module developer

5) They respect the publication rights, rules and regulations (e.g. see section 9.1.2).

Validated modules that have completed Phase 4 are the property of the EORTC QLG and can be downloaded from the EORTC QLG website after a users' agreement has been signed.

Revised modules

Revised modules that have completed Phase 3 are available for general use, with the same rules that apply for the other modules. However, users that wish to use a module for which two versions exist (the original validated version and the revised completed Phase 3 version) should be informed about the two possibilities and given the choice between them.

10 APPENDICES

Appendix 1 Matrix for assisting in targeting patient recruitment

The groups used to construct a sample matrix for different modules may vary, and should be decided by the module developers in advance of each phase. Usually, the matrix will include selected groups from two of three categories: disease stage, treatment type, or stage of treatment. Module developers may decide to combine cells, and to avoid recruitment in some cells as appropriate, relevant to the tumour type or condition being assessed. Two examples are shown below.

In addition to the characteristics used to define the recruitment matrix, it is important to ensure that other characteristics in the recruited sample, e.g. gender, age, and performance status, reflect the target population.

In Phase 1, recruitment should be spread evenly across the cells chosen for inclusion of patients, to ensure a representative sample of patients. In Phase 3, each designated cell should contain at least 15 patients.

Example 1

	Surgery	Chemo-therapy	Radio-therapy	Palliative Care
Localized Disease	X	X	X	-
Advanced Disease	X	X	X	X
Palliative Care	-	-	-	X

Example 2

	Pre Treatment	Mid Treatment	Post Treatment	Palliative Care
Localized Disease	X	X	X	-
Advanced Disease	X	X	X	-
Palliative Care	-	-	-	X

Appendix 2 Generation of relevant QoL issues in Phase 1: Example of healthcare professional interview

The researcher should begin the interview with some introductory remarks to explain its nature and purpose. For example:

We already have a questionnaire assessing the quality of life of cancer patients in general. Quality of life issues which are relevant to specific diagnostic patient groups are not included in this questionnaire. We are asking for your help in devising a questionnaire which will be used to assess the quality of life of patients who have (specific disease or treatment).

Place the issue list before the healthcare professional and continue as follows:

Here you can see a list with issues relevant to cancer patients with (specific disease or treatment).

Could you please indicate for each issue the extent to which you find it relevant for this patient group? "Relevance" refers to the frequency with which a specific complaint occurs and, when it occurs, the trouble it may cause. Thus the more frequently a complaint occurs and the more trouble it causes, the more relevant it will be for this patient group.

Response categories should range from (1) "not relevant" to (4) "very relevant".

After completion the researcher asks:

Could you please tell me for each issue for which you circled 1 (not relevant) or 2 (a little relevant) why you consider it not or only a little relevant?

The researcher should note reasons.

The researcher should then ask the healthcare professional to identify a subset of issues e.g. between 5 and 10, which should definitely be included in the final module and also any issues which they think should definitely not be included.

The list of issues (including any new issues which you have identified) is too long to be administered to patients. Therefore a subset of issues must be chosen. Please could you mark those items that, in your opinion, affect the quality of life of these patients most profoundly and that we should definitely include in the final questionnaire. You may choose (specify the exact number) issues that you consider to be most relevant and that you think should definitely be included. If there are items that you think should definitely be <u>excluded</u> please mark these also and say why you think they are not a priority.

To assess whether the list of issues covers all aspects of quality of life in the target patient group (including all possible subgroups of disease or treatment), the researcher should explore the breadth of the list of issues.

If appropriate, place the EORTC QLQ-C30 before the healthcare professional.

This is the existing questionnaire that assesses the quality of life of cancer patients in general. Could you please read these questions? You may have thought of other things that are not included in this questionnaire, nor in the previous list of issues you have just rated.

Please consider patients at all stages of disease, and patients undergoing any type of treatment for this condition. Can you think of anything else that may be of relevance to this patient group and is not included in these two questionnaires?

If yes: *Please name each of these issues so I can write them down.*

For each additional issue: *Could you tell me about this?*

Appendix 3 Generation of relevant QoL issues in Phase 1: Example of patient interview

Open-ended questions

The researcher should begin the interview with some introductory remarks to explain its nature and purpose. For example:

We are asking for your help in devising a questionnaire which will be used to monitor the experiences of patients who have (specific disease or treatment). I would like to ask you a few things about your health. Can you tell me about the experiences you may have had as a result of your disease (or treatment)?

Neutral probes, which should always be open and non-judgemental, should be used to elicit more information if necessary, e.g. *'Can you tell me more about that?', 'In what way?', 'Just how do you mean?', 'Can you give me an example?', 'Can you think of any additional experiences?'*.

It may be useful to prompt the patient to consider specific domains, especially if the literature review has suggested that these may be relevant to the patient group. Some examples are:

- *Do you have other problems with your physical functioning/health/changes in sleep patterns?*
- *What are you not able to do that you would formerly do before your illness, and why?*
- *Are you limited in normal daily activities (e.g. shopping) or self-care (e.g. washing/bathing) compared to before your illness? What is it that limits you?*
- *Are you undertaking fewer social activities (e.g. hobbies, meeting up with friends) and why?*
- *Have changes in relationships with family/friends occurred?*
- *Do you have financial problems or worries due to your illness?*
- *Have your personal feelings changed (e.g. satisfaction with life, spirituality)?*
- *Has your emotional wellbeing changed (e.g. feelings of anxiety or worrying)?*
- *Are there any other issues or comments you would like to make regarding your illness and treatment and your quality of life?*

Relevance and importance of EORTC QLQ-C30 and issue list

The EORTC QLQ-C30 and/or issue list may be shown to the patient after the initial open-ended questions.

Place the EORTC QLQ-C30 (and any relevant list of items or issues) before the patient and continue as follows:

Here you see a list of experiences related to (condition, treatment or additional QoL dimension) which a patient who is (relevant characteristics) may have.

Please could you indicate for each experience separately the extent to which you have had it during your illness and, if you have experienced it, how much of an impact it has had on/troubled you.

This is an example which could be used to determine the relevance of an issue in a more complex setting e.g. during the development of a module for spiritual wellbeing:

This is a list of thoughts and/or feelings which patients with cancer may experience. Could you please go through the list and, for each one, tell me how much it has been something which you have felt or thought about.

When possible, patients should be encouraged to use the "not at all, a little, quite a bit, very much" response scale to indicate the relevance and importance of each experience.

The EORTC QLQ-C30 and any other relevant list of items or issues can also serve as a prompt to stimulate further suggestions:

Do you have any other symptoms or problems not mentioned in the questionnaire/list which we haven't already talked about?

Inclusion and exclusion of issues

Finally, it may be helpful to ask patients to select which issues (from their own experiences and the existing issue list) they think are the most important ones to include in the final module and whether any should be excluded. To support this, the interviewer will need to summarize the issues raised during the interview and the patient will need sight of the issue list:

Please could you select the (5–15) issues which you think have troubled you the most/are most important to you.

Are there any issues which you think we should not ask patients about?

Appendix 4 Decision rules for selection of QoL issues in Phase 1

In principle, if one or more patients or healthcare professionals mentions an issue, it should be included, provided that the rationale is plausible. If a larger number of patients (>30) have contributed to the list of issues, the threshold for inclusion of an issue into the new module could be that it was mentioned by more than 5% of patients.

At this stage, one should feel reluctant to exclude issues. However, if the number of patients interviewed is large (>30) and the list of issues has been scored by patients or by healthcare providers, issues that have a low mean score (e.g. < 2) for relevance or importance may be considered for exclusion. If any patients and/or healthcare professionals express the view that an issue should definitely not be included, this should be carefully considered.

The module developers should review each issue in the context of the proposed scale structure (i.e. each scale considered in turn as a group of issues). It is necessary to consider the meaning of each issue, whether there is overlap or redundancy within the proposed new issues, and whether the issue is already assessed by the QLQ-C30. Some issues must be handled sensitively when creating a module. For example, issues about approaching death are clearly important to some patients, but may cause distress to others. Alternative phrasing (e.g. “approaching end of life”) may be more acceptable.

In some circumstances, a comparative approach could be used. For example, in selecting issues for inclusion in the QLQ-ELD15 (module for elderly patients), the percentage prevalence of each issue was determined in both the >70 years group and the 50–69 years control group to determine if it was a general concern of all cancer patients, or if it specifically applied to older cancer patients. Issues that were cited by patients at least 1.5 times older than younger (a ratio of 3:2) were considered for inclusion in the new module.

All decisions about inclusion should be reviewed by the entire module development group to ensure consensus on the inclusion or exclusion of issues. The module developers may agree to vary the criteria for particular issues, if there is a strong argument for doing so. This should be recorded in the Phase 1/2 report.

Appendix 5 Example of a patient interview in Phase 3a

Introduction

Pre-testing is designed to collect response data, to record evaluation of relevance and importance and to record the subjective impression of the patients after they have completed the EORTC QLQ-C30 and the new provisional module.

For short provisional modules, the interview should examine each item individually. For longer lists, the interviewer should ask the patient to identify particular aspects of the whole module, and discuss these in detail.

Administration of EORTC QLQ-C30 and the module

The patient is asked to complete the EORTC QLQ-C30 and the new provisional module.

We have two questionnaires that ask about you and your health and quality of life. I will ask you first to complete these questionnaires. After you have completed them, I will interview you to make sure we asked the right questions in the right way. We want to be sure that we cover the most important aspects of patients' experience of (disease/treatment/characteristics).

Place the EORTC QLQ-C30 before the patient who then completes it.

As a result of your (illness/treatment) you may have experiences in common with other patients who have the same problem. These particular experiences are not covered by this more general questionnaire. We would like to add some extra questions to take account of those things which may be important to you and other patients who have (disease/treatment/characteristics). We are now asking your help in devising these additional questions.

We think that this questionnaire may be more useful for patients who have (specific disease or treatment, or additional QoL dimension). As you complete these questions, please could you talk through what you think each item is about and why you are giving the answer you choose. Please let me know if you find any of the items confusing or difficult to answer, or if they are annoying or upsetting in any way.

Place the provisional module before the patient, who completes it. Keep encouraging the patient to talk through their thoughts whilst completing the items, e.g.:

- *Did you have difficulty replying to this question?*
- *Did you find this question annoying?*
- *Did you find this question confusing?*
- *Did you find this question upsetting?*
- *How would you have asked this question?*

Interview directed at the entire module

If modules contain a large number of items (e.g. over 30), the time involved in questioning about each individual item could be prohibitive. In those cases, the questions may be directed towards the entire module. For example:

- *Were there questions that you found difficult to answer?*
- *Were there questions that you found annoying?*
- *Were there questions that you found confusing?*
- *Were there questions that you found upsetting?*
- *Were there questions that you found intrusive?*
- *Do you have other comments about these questions?*

These general questions may then be supplemented by the further probing of selected module items, for example, questions that are expected to cause some difficulty and items that appear to be troublesome during the interview.

Relevance and importance

The patient is then asked to indicate the extent to which they have experienced each issue in the module during their illness (rather than just in the last week) using the "not at all, a little, quite a bit, very much" scale. For items referring to problems the patient has experienced, ask them to rate on the same scale how much the issue has troubled them or had an impact, as a measure of importance. Also ask:

- *Do you think that this problem is related to (disease or treatment)?*

Completion of the interview

- *Can you think of additional issues that are relevant for you but are not included in either of the questionnaires?*

Thank the patient for their contribution to the research.

Appendix 6 Decision rules for inclusion or exclusion of items in Phase 3

In Phase 3 it is necessary to reduce the (usually) long provisional list of items to a shorter (preferably no more than 20) list of items for the new module. Some selection must be applied to remove unnecessary items, balanced against the need to produce a module that adequately covers all the QoL concerns of the target patient group.

The viewpoint of the patient should be given the greatest weight in the selection of items.

General principles

Items should be rated both relevant (e.g. <25% scored 0 ("not at all")) and important by patients (e.g. >60% scored 3 or 4 ("quite a bit" or "very much")). Problems (e.g. symptoms) that relatively few patients describe, and abilities/activities that relatively few patients were limited in, may be of little relevance for inclusion in the final module. These are candidate items for deletion.

Parameters of each item to be considered include the mean score and the number of patients reporting the item (score 2, 3 or 4) divided by the total number that completed the item (prevalence ratio). A full range of responses is important: items that have limited variance should be excluded. In particular, floor and ceiling effects should be looked for in the distribution of responses to each item.

Negative items (e.g. symptoms) score more highly (3 or 4) if the symptom is greater, whereas positive items (e.g. functions) score highly if disability is less. For the purposes of these decision rules, responses can be standardized by inverting responses to the positive items to correspond with response categories ranging from 1 ('no problem') to 4 ('very problematic').

Decision rules

The following cut-off points provide a starting point for decision rules for selection of items for retention in the final module. Module developers may vary these criteria on a case-by-case basis, e.g. depending on the number of items pre-tested and the sample size. It is suggested that all items meet relevance and importance criteria (rules 1 and 2) and at least five of the other seven criteria. Module developers should agree decision rules before beginning the analysis.

1. Relevance: <25% scored 0
2. Importance: >60% scored 3 or 4
3. Mean score > 1.5
4. Prevalence ratio >30% *or* prevalence of scores 3 or 4 >50%
5. Range > 2 points
6. No floor or ceiling effect: responses in categories 3&4 or 1&2 >10%
7. No significant concerns expressed by patients (e.g. item is upsetting, ambiguous)
8. Consistency across languages/cultures
9. Compliance: at least 95% response to the item

The agreed decision rules may be modified if preliminary inspection shows that they would lead to exclusion of too many or too few items. Any post hoc variation to the decision rules should be explained in the Module Development Report.

Answers to the open interview questions may be used to support the removal of items which do fulfil the decision rules criteria (e.g. for the majority of patients, the issues are not related to the disease, or the question meant something different) or the retention of items which do not fulfil the decision rules criteria (e.g. when the importance was stressed in a considerable number of interviews).

Addition of new items

Additional issues (not included in the provisional module) may arise during Phase 3 which are felt sufficiently important to warrant consideration for inclusion. However, investigators should retain a high threshold for the addition of new items at this stage. Investigators may wish to agree a defined proportion of patients that report a missing issue before it could be added to the module at this stage (e.g. 30%) or they may wish to consider each suggestion on a case-by-case basis. Decisions should be documented in the report. If additional issues are developed into items, some preliminary testing of the new items is advised before proceeding to Phase 3b.

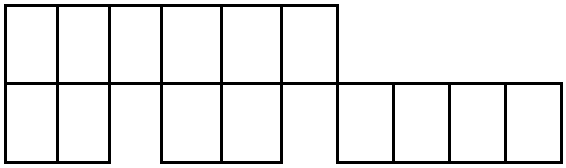
Rephrasing items

On the basis of the interviews, questions may be identified that troubled (some of) the patients. This information should be taken seriously. Even when a small number of patients had difficulty answering the questions, these should be rephrased, subdivided or substantially changed as appropriate and new patients should be consulted on the appropriateness of the revisions.

Appendix 7 Debriefing questionnaire for Phase 4

Patient Study ID

Date of Interview:



ABCXX Debriefing questionnaire

1. How long did it take you to complete the questionnaire?

☐
minutes

2. Did anyone help you to complete the questionnaire?

No **Yes** **If so, who:**

What kind of help was provided?

Example of a more detailed way to ask whether help was provided and which kind:

Did the patient require help to complete the QLQ-SWB36? If yes, please state whether the help was:

- *Practical (e.g. patient has poor eyesight, or is unable to hold a pen and write easily, or tires easily)*
- *Supportive (e.g. patient appreciated or felt reassured that someone was with them to discuss the measure, but they did not need clarification or further explanations)*
- *Required for understanding (e.g. patient was unable to respond without asking for more information or an alternative explanation).*

How much help was provided?

3. Were there questions that you found confusing or difficult to answer?

No **Yes**

If so, which ones?

4. Were there questions that you found upsetting?

No **Yes**

If so, which ones?

5. Please use the space below if you have other comments about the questionnaire.

Thank you!

Appendix 8 Item Response Theory (IRT) for scale structure and selection of items in Phase 4

For QLQ module development, IRT can be used as a psychometric development tool. If an early version of a module contains a lot of candidate items that are all believed to be measuring much the same thing, IRT provides an excellent means for identifying the most informative one or two items, and for quantifying how much extra information or precision would be gained from increasing the scale length by including additional items. The publications describing the QLQ-C15-PAL provide examples of shortening some scales of the QLQ-C30 (Petersen et al., 2006; Groenvold et al., 2006; Bjorner et al., 2004).

Whereas traditional psychometrics explores averages (means), standard deviations and correlations of the responses to questions, IRT is concerned with the *probability* that any particular patient will select one or another response option. Factor analysis and IRT are concerned with 'latent variables' that are not directly measurable, but which the scale-score is assumed to represent.

IRT has become increasingly widely used in module development, and – when applicable – possesses some major advantages over traditional methods. IRT is primarily useful when there are a number of items that all address a single homogeneous dimension. Like factor analysis, it is of no value for single items (although it may aid in selecting a single item from among a group of similar items). Unlike factor analysis, it is not suitable for multi-item scales that lack homogeneity – as might be the case if several items are deliberately chosen to extend the breadth of coverage of a concept (i.e. multi-item scales characterized by a low Cronbach's α).

IRT can be used (a) solely as an aid to developing a scale that is then to be scored using traditional methods (such as summation, as commonly used for most HRQoL scales), or (b) to develop an IRT-based scale that is also scored using IRT computer software. Examples of (b) would be primarily but not necessarily scales such as physical functioning, where a number of items might be chosen to target patients with varying levels of ability (e.g. "*Can you get out of bed?", "Can you run a marathon?"*). For such scales, IRT can provide a consistent scoring system.

To apply IRT, we would typically require data on at least 400 patients, and the sample should contain a fairly even spread of patients across the continuum of interest (i.e. it is unhelpful and uninformative to have a lot of patients responding with 'no problems'). A simple introduction to IRT may be found in Fayers and Machin (2007), while a more detailed exposition is provided by Embretson and Reise (2000).

Appendix 9 Examples of item structure and word selection

Words and structures – consistency and translation issues

Did you have trouble doing…?

→ Better: Have you had problems…?

- "Trouble" is difficult to translate
- Tense should match the standard time frame "in the past week"

To what extent/How much have you…?

→ Better: Have you…?

- Item format has to match the response scale of "Not at all – A little – Quite a bit – Very much" as closely as possible

Have you been worried…/ Have you worried… / Did you worry…

→ Better: Have you worried…?

- Consistency of similarly worded items should be ensured

Were you concerned about disruption of family life?

→ Better: Have you worried that your family life might become disrupted?

- For items on worries, it should be clear if the situation is at present or in the future in order to facilitate translation and avoid ambiguity

… in one or both legs?

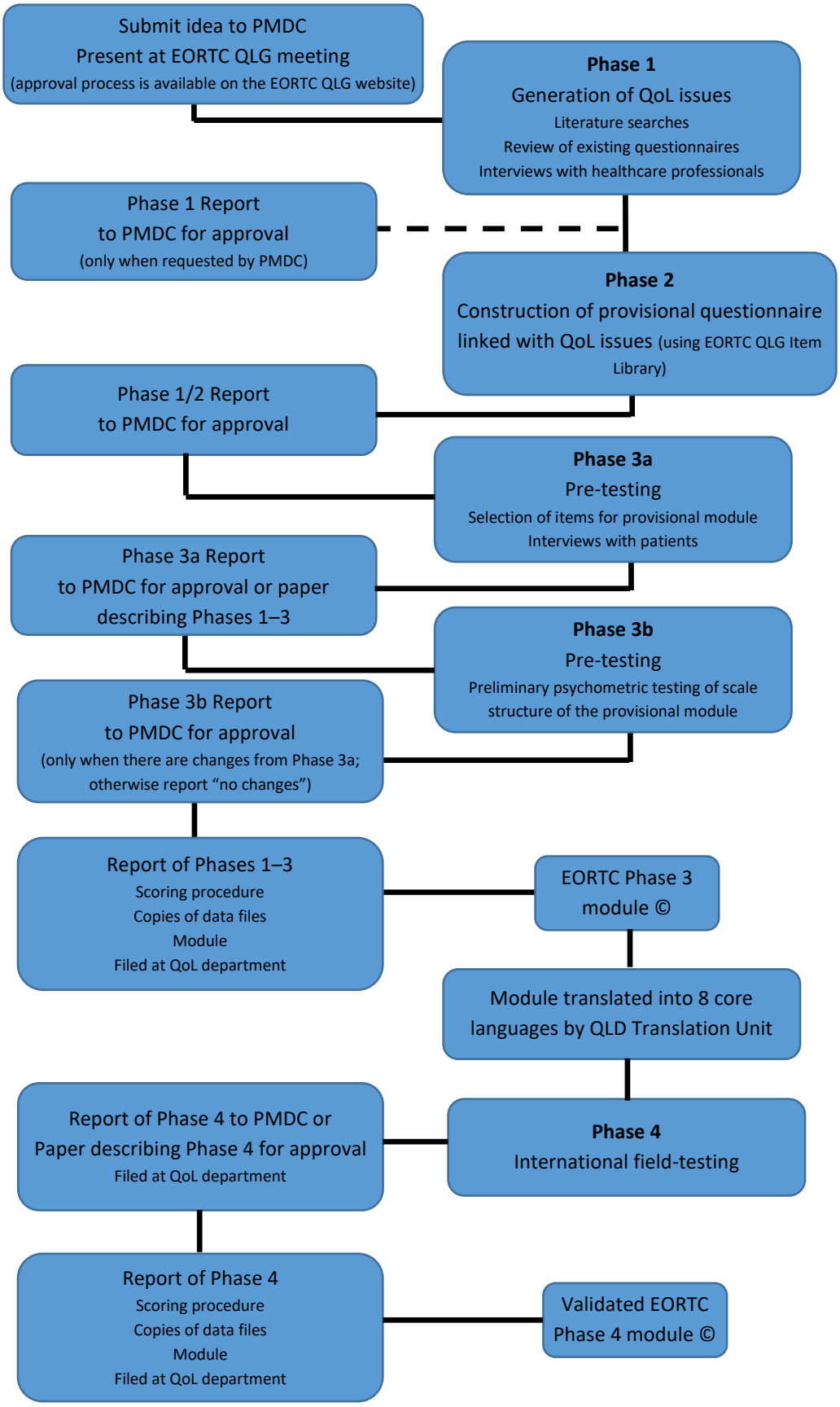
→ Better: in your legs?

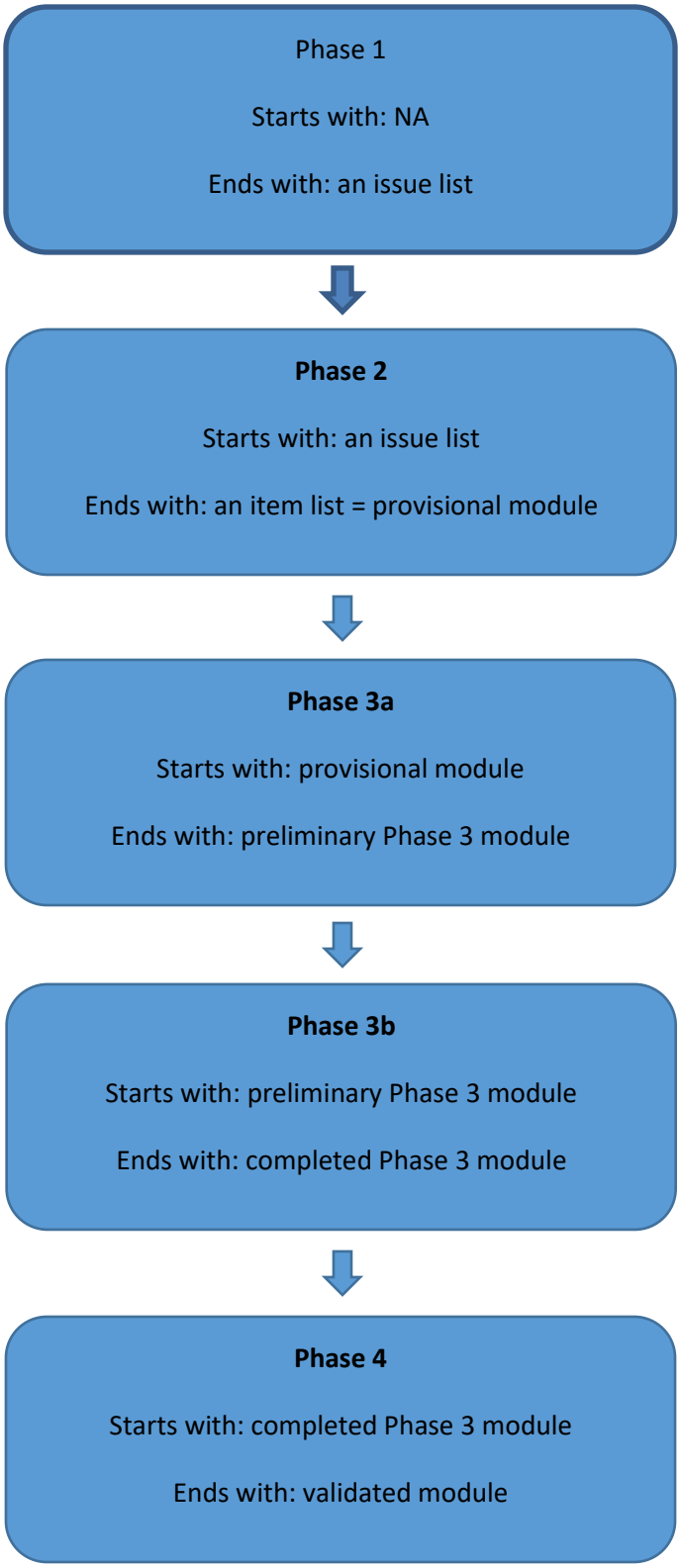
- In some languages, one leg and both legs have different grammatical genders and the translated item becomes unnecessarily complicated

Item wording – general recommendations

- Specify context to help patients focus on the intended aspect, e.g. by adding "… because of your disease or treatment?"
- Add examples to minimize vagueness of meaning in English and translations, e.g.: "(e.g. healthy eating, exercise)"
- Make sure the terms are specific: for example, terms referring to sexuality have to be chosen on purpose – in translation, it makes a difference whether the item refers to intimacy, sex, sexual activity, intercourse, etc.
- Avoid colloquialisms and phrasal verbs – they can be mistranslated and introduce lengthy structures in translation (for example, "do up" instead of "fasten")
- Consider educational level in selection of words (e.g. "leakage" (lower educational level) versus "incontinent")

Appendix 10 Flow chart




Phase 1
Starts with: NA
Ends with: an issue list
Phase 2
Starts with: an issue list
Ends with: an item list = provisional module
Phase 3a
Starts with: provisional module
Ends with: preliminary Phase 3 module
Phase 3b
Starts with: preliminary Phase 3 module
Ends with: completed Phase 3 module
Phase 4
Starts with: completed Phase 3 module
Ends with: validated module

Documents (protocols and reports)

Before starting with Phase 1:

- Phase 1 (or 1&2) grant application
- Phase 1 (or 1&2) protocol

At the end of Phase 1:

- Phase 1 report if the application was only for Phase 1

Before starting with Phase 2:

- Approved Phase 1 report if the application was only for Phase 1
- Phase 2 grant application if the application was only for Phase 1
- Phase 2 protocol if the application was only for Phase 1

At the end of Phase 2:

- Phase 2 (or 1&2) report

*** If the application is for Phase 1&2: one protocol for Phase 1&2 and one report for Phase 1&2 ***

Before starting with Phase 3:

- Phase 2 (or 1&2) report approved
- Phase 3a&3b grant application
- Phase 3a&3b protocol

At the end of Phase 3a:

- Report for Phase 3a

Before starting with Phase 3b:

- Phase 3a report approved

At the end of Phase 3b:

- Report for Phase 3b if applicable (i.e. if changes in Phase 3b)

Before starting with Phase 4:

- Phase 3b report approved if applicable (i.e. if changes in Phase 3b)
- Grant application for Phase 4
- Phase 4 protocol

At the end of Phase 4:

- Phase 3b&4 report (or only 4, if changes in Phase 3b and hence Phase 3b report already submitted and approved)

11 REFERENCES

Aaronson N.K., Bullinger M. & Ahmedzai S. (1988), 'A modular approach to quality of life assessment in cancer clinical trials'. In: Scheurlen H., Kay R., Baum M. (eds) *Cancer Clinical Trials. Recent Results in Cancer Research*, 111: 231–249. Heidelberg: Springer Berlin.

JB Bjorner, M Kosinski, JE Ware Jr (2003), The feasibility of applying item response theory to measures of migraine impact: A re-analysis of three clinical studies *Quality of Life Research,* 12 (8), 887-902

Bjorner, J.B., Petersen, M.Aa., Groenvold, M., Aaronson, N., Ahlner-Elmqvist, M., Arraras, J.I., Brédart, A., Fayers, P., Jordhøy, M., Sprangers, M., Watson, M. & Young, T. (2004), Use of item response theory to develop a shortened version of the EORTC QLQ-C30 emotional functioning scale. *Quality of Life Research,* 13(10): 1683–1697.

Brod, M., Tesler, L.E. & Christensen, T.L. (2009), Qualitative research and content validity: developing best practices based on science and experience. *Quality of Life Research,* 18(9): 1263–1278.

Converse, J. & Presser, S. (1986), *Survey Questions: Handcrafting the Standardized Questionnaire*. Sage University Paper series on Quantitative Applications in the Social Sciences, 63. Beverly Hills, CA: Sage Publications.

Cocks, K., King, M.T., Velikova, G., de Castro Jr, G., Martyn St-James, M., Fayers, P.M. & Brown, J.M. (2012), Evidence-based guidelines for interpreting change scores for the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. *European Journal of Cancer*, 48(11): 1713–1721.

Embretson, S.E. & Reise, S.P. (2000), *Item Response Theory for Psychologists*. Mahwah, NJ: Laurence Erlbaum Associates.

Fayers, P.M. & Hand, D.J. (1997), Factor analysis, causal indicators and quality of life. *Quality of Life Research* 6(2): 139–150.

Fayers, P.M., Hand, D.J., Bjordal, K. & Groenvold. M. (1997), Causal indicators in quality of life research. *Quality of Life Research,* 6(5): 393–406.

Fayers, P.M. & Hays, R.D. (2014), Don't middle your MIDs: Regression to the mean shrinks estimates of minimally important differences. *Quality of Life Research*, 23(1): 1–4.

Fayers, P.M. & Machin, D. (2016), *Quality of Life: The Assessment, Analysis and Interpretation of Patient-reported Outcomes,* 3rd edition. Chichester: John Wiley & Sons.

Fortune-Greeley, A.K., Flynn, K.E., Jeffery, D.D., Williams, M.S., Keefe, F.J., Reeve, B.B., Willis, G.B. & Weinfurt, K.P. (2009), Using cognitive interviews to evaluate items for measuring sexual functioning across cancer populations: improvements and remaining challenges. *Quality of Life Research,* 18(8): 1085–1093.

Francis, J. J., Johnston, M., Robertson, C., Glidewell, L., Entwistle, V., Eccles, M.P. and Grimshaw, J.M. (2010), What is an adequate sample size? Operationalising data saturation for theory-based interview studies. *Psychology & health* 25(10): 1229-1245.

Giesinger, J.M., Aaronson, N.K., Arraras, J.I., Efficace, F., Groenvold, M., Kieffer, J.M., Loth, F.L., Petersen, M.Aa., Ramage, J., Tomaszewski, K.A., Young, T. & Holzner, B. on behalf the EORTC Quality of Life Group (2018), A cross-cultural convergent parallel mixed methods study of what makes a cancer-related symptom or functional health problem clinically important. *Psychooncology,* 27(2): 548–555.

Giesinger, J.M., Loth, F.L.C., Aaronson, N.K., Arraras, J.I., Caocci, G., Efficace, F., Groenvold, M., van Leeuwen, M., Petersen, M.Aa., Ramage, J., Tomaszewski, K.A., Young, T. & Holzner, B. on behalf the EORTC Quality of Life Group (2020a), Thresholds for clinical importance were defined for the European Organisation for Research and Treatment of Cancer Computer Adaptive Testing Core – an adaptive measure of core quality of life domains in oncology clinical practice and research. *Journal of Clinical Epidemiology,* 117: 117–125.

Giesinger, J.M., Loth, F.L.C., Aaronson, N.K., Arraras, J.I., Caocci, G., Efficace, F., Groenvold, M., van Leeuwen, M., Petersen, M.Aa., Ramage, J., Tomaszewski, K.A., Young, T. & Holzner, B. on behalf the EORTC Quality of Life Group (2020b), Thresholds for clinical importance were established to improve interpretation of the EORTC QLQ-C30 in clinical practice and research. *Journal of Clinical Epidemiology,* 118: 1–8.

Groenvold, M., Bjorner, J.B., Klee, M.C. & Kreiner, S. (1995), Test for item bias in a quality of life questionnaire. *Journal of Clinical Epidemiology,* 48(6): 805–816.

Groenvold, M. (1997), Quality of life in breast cancer adjuvant therapy: validation and pilot testing of a combination of questionnaires. *The Breast,* 6(2): 97–107

Groenvold, M., Petersen, M.Aa. (2005), The role and use of differential item functioning (DIF) analysis of quality of life data in clinical trials. *Assessing Quality of Life in Clinical Trials* (2. Udgave)

Groenvold, M., Petersen, M.Aa., Aaronson, N.K., Arraras, J.I., Blazeby, J.M., Bottomley, A., Fayers, P.M., de Graeff, A., Hammerlid, E., Kaasa, S., Sprangers, M.A.G. & Bjorner, J.B. (2006), The development of the EORTC QLQ-C15-PAL: a shortened questionnaire for cancer patients in palliative care. *European Journal of Cancer,* 42(1): 55–64.

Kerr, C., Nixon, A. & Wild, D. (2010), Assessing and demonstrating data saturation in qualitative inquiry supporting patient-reported outcomes research. *Expert Review Pharmacoeconomics Outcomes Research,* 10(3): 269–281.

Koller, M., Warncke, S., Hjermstad, M.J., Arraras, J., Pompili, C., Harle, A., Johnson, C.D., Chie, W.C., Schulz, C., Zeman, F., van Meerbeeck, J.P., Kuliś, D. & Bottomley, A. on behalf of the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Group & the EORTC Lung Cancer Group (2015), Use of the lung cancer-specific Quality of Life Questionnaire EORTC QLQ-LC13 in clinical trials: A systematic review of the literature 20 years after its development. *Cancer*, 121(24): 4300–4323.

Koller, M., Hjermstad, M.J., Tomaszewski, K.A., Tomaszewska, I.M., Hornslien, K., Harle, A., Arraras, J.I., Morag, O., Pompili, C., Ioannidis, G., Georgiou, M., Navarra, C., Chie, W.C., Johnson, C.D., Himpel, A., Schulz, C., Bohrer, T., Janssens, A., Kuliś D. & Bottomley, A. (2017), An international study to revise the EORTC questionnaire for assessing quality of life in lung cancer patients. *Annals of Oncology*, 28(11): 2874–2881.

Kuliś, D., Bottomley, A., Velikova, G., Greimel, E. & Koller, M., on behalf of the EORTC Quality of Life Group (2017), *EORTC Quality of Life Group Translation Procedure,* 4th edition. Brussels: EORTC.

Krueger, R.A. & Casey, M.A. (2000), *Focus Groups: A Practical Guide for Applied Research,* 3rd edition. Thousand Oaks, CA: Sage Publications.

Kvam, A.K., Wisløff, F. and Fayers, P.M. (2010), Minimal important differences and response shift in health-related quality of life; a longitudinal study in patients with multiple myeloma. *Health and Quality of Life Outcomes*, 8(1), p.79.

Maringwa, J.T., Quinten, C., King, M., Ringash, J., Osoba, D., Coens, C., Martinelli, F., Vercauteren, J., Cleeland, C.S., Flechtner, H., Gotay, C., Greimel, E., Taphoorn, M.J., Reeve, B.B., Schmucker-Von Koch, J., Weis, J., Smit, E.F., van Meerbeeck, J.P., Bottomley, A., EORTC PROBE project & the Lung Cancer Group (2011), Minimal important differences for interpreting health-related quality of life scores from the EORTC QLQ-C30 in lung cancer patients participating in randomized controlled trials. *Support Care Cancer,* 19(11): 1753–1760.

Muthén, L.K. and Muthén, B.O. (2010), 1998–2010 Mplus user‘s guide. *Muthén and Muthén*, pp.39-49.

Nunnally, J.C. & Bernstein, I.H. (1994), *Psychometric Theory,* 3rd edition. New York: McGraw Hill.

Osoba, D., Rodrigues, G., Myles, J., Zee, B. & Pater, J. (1998), Interpreting the significance of changes in health-related quality-of-life scores. *Journal of Clinical Oncology,* 16(1): 139–144.

Petersen, M.Aa., Groenvold, M., Aaronson, N., Blazeby, J., Brandberg, Y., de Graeff, A., Fayers, P., Hammerlid, E., Sprangers, M., Velikova, G., & Bjorner, J.B. for the European Organisation for Research and Treatment of Cancer Quality of Life Group (2006), Item response theory was used to shorten EORTC QLQ-C30 scales for use in palliative care. *Journal of Clinical Epidemiology,* 59(1): 36–44.

Petersen, M.Aa., Groenvold, M., Aaronson, N.K., Chie, W.C., Conroy, T., Costantini, A., Fayers, P., Helbostad, J., Holzner, B., Kaasa, S., Singer, S., Velikova, G. & Young, T. (2011), Development of computerized adaptive testing (CAT) for the EORTC QLQ-C30 physical functioning dimension. *Quality of Life Research*, 20(4): 479–490.

Petersen, M.Aa., Groenvold, M., Aaronson, N., Brenne, E., Fayers, P., Damgaard-Nielsen, J., Sprangers, M. & Bjorner, J.B. for the European Organisation for Research and Treatment of Cancer Quality of Life Group (2005), Scoring based on item response theory did not alter the

measurement ability of the EORTC QLQ–C30 scales. *Journal of Clinical Epidemiology,* 58(9): 902–908.

Revicki, D., Hays, R.D., Cella, D. & Sloan, J. (2008), Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology,* 61(2): 102–109.

Rothman, M., Burke, L., Erickson, P., Leidy, N.K., Patrick, D.L. & Petrie, C.D. (2009), Use of existing patient-reported outcome (PRO) instruments and their modification: the ISPOR Good Research Practices for Evaluating and Documenting Content Validity for the Use of Existing Instruments and Their Modification PRO Task Force Report. *Value in Health,* 12(8): 1075–1083.Snyder, C., Brundage, M., Rivera, Y.M. & Wu, A.W. (2019), A PRO-cision Medicine Methods Toolkit to Address the Challenges of Personalizing Cancer Care Using Patient-Reported Outcomes: Introduction to the Supplement. *Medical Care,* 57(5 Suppl 1): S1–S7.

Sprangers, M.A.G., Cull, A., Bjordal, K., Groenvold M., Aaronson, N.K. & the EORTC Study Group on Quality of Life (1993), The European Organization for Research and Treatment of Cancer approach to quality of life assessment: guidelines for developing questionnaire modules. *Quality of Life Research,* 2(4): 287–295.

Streiner, D.L. & Norman, G.R. (2008), *Health Measurement Scales. A practical guide to their development and use*, 4th edition. Oxford: Oxford University Press.

Terluin, B., Eekhout, I., Terwee, C.B., de Vet, H.C.W. (2015), Minimal important change (MIC) based on a predictive modeling approach was more precise than MIC based on ROC analysis. *Journal of Clinical Epidemiology,* 68(12): 1388–1396.

Wintner, L.M., Sztankay, M., Aaronson, N., Bottomley, A., Giesinger, J.M., Groenvold, M., Petersen, M.Aa., van de Poll-Franse, L., Velikova, G., Verdonck-de Leeuw, I. & Holzner, B. on behalf of the EORTC Quality of Life Group (2016), The use of EORTC measures in daily clinical practice – A synopsis of a newly developed manual. *European Journal of Cancer,* 68: 73–81.

Wintner, L.M., Sztankay, M., Giesinger, J.M., Aaronson, N., Bottomley, A., Velikova, G., Verdonck de Leeuw, I., van de Poll-Franse, L., Groenvold, M., Petersen, M.Aa., Holzner, B., on behalf of the EORTC Quality of Life Group (2016), Manual for the Use of EORTC Measures in Daily Clinical Practice. Brussels: EORTC.

Websites:

https://www.eortc.org/ – the main EORTC website, from which is possible to access the websites of the DOGs

https://qol.eortc.org/ – the website of the EORTC QLG

https://qol.eortc.org/item-library/ – the website of the EORTC QLG Item Library

https://www.fda.gov/

http://www.prisma-statement.org/

D/2021/6136/01
ISBN 978-2-930064-48-2