Setting International Standards in Analyzing PatientReported Outcomes and Quality of Life Endpoints in Cancer Clinical Trials-Innovative Medicines Initiative (SISAQOL-IMI)

SISAQOL-IMI Consortium

DOI:
10.1016/S1470-2045(23)00157-2

Document Version
Peer reviewed version

Citation for published version (Harvard):
SISAQOL-IMI Consortium 2023, 'Setting International Standards in Analyzing PatientReported Outcomes and Quality of Life Endpoints in Cancer Clinical Trials-Innovative Medicines Initiative (SISAQOL-IMI): stakeholder views, objectives, and procedures', *The Lancet Oncology*, vol. 24, no. 6, pp. e270-e283. https://doi.org/10.1016/S1470-2045(23)00157-2

TITLE: Setting International Standards in Analysing Patient-Reported Outcomes and Quality of Life Endpoints in Cancer Clinical Trials - Innovative Medicines Initiative (SISAQOL-IMI): Stakeholder Views, Objectives, and Procedures

AUTHORS: Madeline Pe[1], Ahu Alanya[1], Ragnhild Sorum Falk[2], Cecilie Delphin Amdal[2,3], Kristin Bjordal[2,4], Jane Chang[5], Paul Cislo[5], Corneel Coens[1], Linda Dirven[6,7], Rebecca M. Speck[8], Kristina Fitzgerald[9], Jayne Galinsky[10], Johannes M. Giesinger[11], Bernhard Holzner[11], Saskia Le Cessie[12,13,14], Daniel O' Connor[15], Kathy Oliver[16], Vivek Pawar[17], Chantal Quinten[18], Michael Schlichting[19], Jinma Ren[5], Satrajit Roychoudhury[5], Martin J.B. Taphoorn[6,7], Galina Velikova[20], Lisa M. Wintner[11], Ingolf Griebsch[21], and Andrew Bottomley[1] *on behalf of SISAQOL-IMI Consortium*

AFFILIATIONS:

[1] European Organisation for Research and Treatment of Cancer (EORTC) Headquarters, Brussels, Belgium (M Pe PhD, A Alanya PhD, C Coens MSc, A Bottomley PhD)

[2]Research Support Services, Oslo University Hospital, Norway (R S Falk PhD, C D Amdal MD PhD, Prof K Bjordal MD PhD)

[3]Department of Oncology, Oslo University Hospital, Norway (C D Amdal MD PhD)

[4]Faculty of Medicine, University of Oslo, Oslo, Norway (Prof K Bjordal MD PhD)

[5]Pfizer Inc, US (J Chang MPH, P Cislo PhD, J Ren MD PhD, S Roychoudhury BSc)

[6]Department of Neurology, Leiden University Medical Center, the Netherlands. (L Dirven PhD, Prof M Taphoorn MD PhD)

[7]Department of Neurology, Haaglanden Medical Center, the Hague, the Netherlands. (L Dirven PhD, Prof M Taphoorn MD PhD)

[8]Critical Path Institute (C-Path), AZ, USA (R M Speck PhD)

[9]AbbVie, IL, USA (K Fitzgerald MPH)

[10]Myeloma Patients Europe, Brussels, Belgium (J Galinsky PhD)

[11]University Hospital of Psychiatry II, Medical University of Innsbruck, Austria (J M Giesinger PhD, Prof B Holzner PhD, L M Wintner PhD)

[12]Department of Clinical Epidemiology, Leiden University Medical Center, the Netherlands. (Prof S le Cessie PhD)

[13]Department of Biomedical Data Sciences, Leiden University Medical Center, the Netherlands. (D Thomassen MSc, Prof S le Cessie PhD)

[14]Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Belgium. (Prof S le Cessie PhD)

[15]Medicines and Healthcare products Regulatory Agency, London, UK (D O'Connor MBChB PhD)

[16]International Brain Tumour Alliance, Surrey, UK (K Oliver BA)

[17]EMD Serono, USA (V Pawar PhD)

[18]European Medicines Agency, Amsterdam, the Netherlands (C Quinten PhD)

[19]Merck, Healthcare KGaA, Germany (M Schlichting MSc)

[20]Leeds Institute of Cancer and Pathology, University of Leeds, St James's Hospital, Leeds, UK (Prof G Velikova PhD)

[21]Boehringer Ingelheim, Germany (I Griebsch PhD)

Corresponding Author

Madeline Pe, Ph.D., Quality of Life Department, European Organization for Research and Treatment of Cancer, 83/11 Avenue E. Mounier,1200 Brussels, Belgium; Tel: +32 (0) 2 774 16 61; madeline.pe@eortc.org

Total number of words: 4397

Search strategy and selection criteria

References for this Review were identified through searches of PubMed with the search terms *("patient reported outcome analysis") OR ("quality of life analysis") AND "cancer" AND "clinical trials"*. No date restrictions were included. Articles were also identified through searches of the authors' own files and recommendations by the SISAQOL-IMI Consortium. Only papers published in English were reviewed. The search was conducted on July 9, 2021. The final reference list was generated based on originality and relevance to the broad scope of this Review.

Abstract (150 words unstructured summary)
Patient-reported outcomes (PROs), such as symptoms, functioning and other health-related quality of life concepts are gaining a more prominent role in the benefit/risk assessment of cancer therapies. However, varying ways of analysing, presenting and interpreting PRO data may lead to erroneous and inconsistent decisions on the part of stakeholders, adversely impacting patient care and outcomes. The *Setting International Standards in Analysing Patient-Reported Outcomes and Quality of Life Endpoints in Cancer Clinical Trials - Innovative Medicines Initiative* (SISAQOL-IMI) Consortium builds on the existing SISAQOL work to establish recommendations on design, analysis, presentation, and interpretation for PRO data in cancer clinical trials. This paper presents an expanded set of topics and international stakeholder views on the need for SISAQOL-IMI, the agreed upon prioritized set of PRO objectives to focus on, and the roadmap to ensure that international consensus recommendations will be achieved.

Funding
The SISAQOL-IMI project has received funding from the Innovative Medicines Initiative (IMI) 2 Joint Undertaking under grant agreement No 945052. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and the European Federation of Pharmaceutical Industries and Associations (EFPIA).

Introduction

Patient-reported outcomes (PROs) are any outcome evaluated directly by the patient and based on the patient's perception of a disease and its treatment(s).[1] PROs is an umbrella term covering both single- and multi-dimensional measures of symptoms, functioning, and other health-related quality of life concepts. These PRO concepts are gaining a more prominent role in the benefit/risk assessment and relative effectiveness assessments of cancer therapies. Although various stakeholders are increasingly adopting the use of PROs in their decision-making,[2] evidence from systematic reviews has consistently shown the lack of standards and guidance on how PRO data are collected, analysed, presented and interpreted in cancer randomized clinical trials (RCTs).[3-8] Inconsistent and, at times, inappropriate PRO design, data collection, analysis and interpretation puts into question the reliability and robustness of PRO data, which in turn reduces the ability of PRO data to inform the overall risk or benefit assessment of cancer treatments.

Recommendations for handling PROs now exist for protocols (Standard Protocol Items: Recommendations for Interventional Trials-PRO extension; SPIRIT-PRO),[9,10] publications to improve reporting of PROs (Consolidated Standards of Reporting Trials Statement-PRO extension; CONSORT-PRO)[11] and for graphically displaying PRO data.[12] These and other methodologic guidance documents, as well as resources to aid in their use, are available at the PROTEUS Consortium (www.TheProteusConsortium.org). These reporting guidelines are essential since they provide key information to allow evaluation of the design and analysis used to inform PRO.[13] In addition to reporting guidelines, it is equally important that evidence-based and harmonized methodological standards are set so that PRO data from cancer clinical trials are analysed, presented and interpreted appropriately, PRO results are reproducible, and to ensure that PRO data can inform patient safety, treatment choices and policy decisions in a meaningful and reliable way.[13]

The need to improve the methodological quality of PRO design, collection, analysis and interpretation in cancer clinical trials was initially recognized by the prior "Setting International Standards in Analysing Patient-Reported Outcomes and Quality of Life Endpoints Data" (SISAQOL) initiative, which ran from 2016 to 2020. The first "edition" of SISAQOL addressed the challenges in the analysis and interpretation of

PROs in cancer randomised controlled trials (RCTs).[14] Key to this initiative was ensuring that the recommendations remained relevant across different types of PRO measures and incorporated the perspectives from various international stakeholders, including regulators, health technology assessment (HTA) bodies, industry and academic representatives, clinicians, methodological and applied statisticians, PRO experts, and patient representatives. The first SISAQOL Consortium completed their initial work and published international consensus recommendations for PRO data analysis in RCTs in 2020.[15] These recommendations are also in line with the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) E9 (R1) estimand framework of defining clear clinical study objectives to align with the study design, endpoint, and analysis.[16]

SISAQOL-IMI

The demand for better standards on the design and analysis of cancer clinical trials with PRO data has been further highlighted by a recent call of the Innovative Medicines Initiative (IMI).[17] The IMI Joint Undertaking (JU) is a public-private partnership (PPP) between the European Union (EU), represented by the European Commission (EC), and the European Federation of Pharmaceutical Industries and Associations (EFPIA).[18,19] IMI's aim is to address unmet needs in drug development that can be efficiently resolved within a pre-competitive space through multidisciplinary, multi-stakeholder collaborations. This specific IMI call emphasized the need to establish international standards and extend coverage in the analysis of PRO data, such as symptom and functional outcomes, in cancer clinical trials. IMI envisions that such standards will not only support the use of PRO data for optimal drug development and device approval by regulators and HTA bodies, but also help better communicate PRO results between clinicians and patients for the purposes of shared decision-making.[17] It is worth noting that SISAQOL-IMI is not about the development and validation of patient-reported outcomes, which are available in other sources.[20-24]

Recommendations from the first SISAQOL Consortium provided the initial step towards generating international consensus-based standards for the design and analysis of PRO data in RCTs. Based on the first SISAQOL Consortium, the original

participants joined new collaborators to respond to the IMI call in 2019. Knowledge and experience gained from prior work was leveraged with the goal to continue to advance international consensus recommendations on PRO data analyses. The new Consortium is referred to as SISAQOL-IMI (see Appendix 3 Figure 1 on page 25 for the development from SISAQOL to SISAQOL-IMI).

The aim of SISAQOL-IMI is to develop recommendations on the design, analysis, presentation and interpretation of PRO data considering the estimand framework[25] that address the needs of the entire spectrum of stakeholders. This is achieved by bringing together an extensive collection of international experts to collaborate and agree on a set of PRO data analysis recommendations (see Table 1 for a full list of organizations involved in SISAQOL-IMI). These stakeholder groups include academics, industry, non-profit/cancer organisations, small to mid-size enterprises or contract research organisations (CRO), regulators, HTA bodies, and patients' representatives.

Moreover, SISAQOL-IMI builds on the first SISAQOL Consortium effort by harmonising and updating available recommendations based on stakeholder needs and recent developments in the methodological literature. SISAQOL-IMI will also broaden its scope by not only developing recommendations for the design and analysis of PRO data within RCTs, but also exploring the feasibility of recommendations for non-RCTs (with a specific focus on single-arm studies), improving presentation of PRO results and producing guidance on how to define clinically meaningful differences. Using case studies, validation or testing work will also be done for these newly developed recommendations, specifically in assessing whether these recommendations are understandable and feasible to implement in protocols and statistical analysis plans in cancer clinical trials.

To ensure that these goals are achieved, the project is organised into five scientific work packages (WPs), and three cross-cutting WPs (see Appendix 3 Figure 2 on page 25 for the project structure and interaction between the WPs). The scientific WPs will focus on developing recommendations that are of high methodological quality; whereas the cross-cutting WPs will ensure that the output of the various SISAQOL-IMI scientific WPs remain cohesive and harmonised, that the consensus

process is transparent, and the final SISAQOL-IMI consensus-based recommendations address the needs of the various stakeholder groups.

Although individual stakeholders have broad guidelines on the use of PROs in cancer clinical trials,[26-31] a harmonised set of standards that are methodologically rigorous, practical and feasible to implement is needed. This can improve confidence that conclusions based on PRO data from cancer clinical trials are reliable, replicable, robust, interpretable, and clinically meaningful. Similar to the first SISAQOL Consortium, SISAQOL-IMI recommendations are intended to be generalisable to all validated PRO measures. Figure 1 in Appendix 3 on page 25 presents the workflow towards developing the final SISAQOL-IMI recommendations.

Need for SISAQOL-IMI: Views from stakeholder groups

During the SISAQOL-IMI kick-off meeting in March 2021, representatives from various SISAQOL-IMI stakeholder groups (clinicians, patients, academic, industry, regulatory and HTA bodies) presented their views on the current use of PROs and their expectations from the SISAQOL-IMI recommendations. Although stakeholders presented different views on the use of PROs for treatment decision-making, participants agreed that PRO data can be as important as other common clinical endpoints (e.g., overall survival, progression free survival) and they also provide complementary information to clinician-reported outcomes. However, standards and guidance for PRO data analysis in cancer clinical trials are needed to facilitate its use in stakeholder decision-making.

From *the clinicians' and patients' perspective*, it is essential that PRO data are collected in cancer clinical trials and analyzed optimally so that they can have information, directly from patients themselves, on the impact of cancer treatments on how patients feel and function. The analysis, presentation and interpretation of results should clearly highlight not only statistical significance, but also patient benefit and risk since these data may be used for the assessment of the overall clinical benefit of cancer treatments (e.g., see European Society for Medical Oncology-Magnitude of Clinical Benefit Scale[32]) and will have real-world impact on clinician-patient communication regarding treatment decisions.[26]

Academic and industry representatives discussed the many challenges when including PROs in the design of cancer clinical trials. The complexity of clinical trial

development increases when PRO endpoints are added. This includes aligning PROs with other (primary) clinical endpoints, designing the timing of PRO data collection, selecting appropriate analyses, handling intercurrent events (e.g., death and treatment discontinuation[16,25]) and missing data, and ensuring accurate communication and interpretation of PRO findings. These challenges have hindered the optimal use of PROs in clinical trials. SISAQOL-IMI is expected to help address these hurdles by providing recommendations for standardizing each of the aforementioned aspects of rigorously incorporating PROs in a clinical trial setting. Having such recommendations can help to focus on stakeholder needs, reduce patient burden and "research waste" by ensuring efficiency of PRO data collection and producing results based on meaningful analyses that will be useful for patient and other stakeholder decision-making.

Finally, r*egulators and HTA* bodies touched upon the issues they face when evaluating PRO data submitted by sponsors (trialists). For example, submissions often lack a clear PRO research objective, PROs are often positioned as exploratory endpoints, and there can be large amounts of missing PRO data, putting into question the robustness and reliability of the PRO data to inform the benefit and risks of cancer therapies. For this stakeholder group, the need for SISAQOL-IMI is to improve the standards of assessing PROs in cancer clinical trials by informing design, data collection and analysis practices so that PRO data can be fully considered in regulatory and HTA decision-making.

Work scope of SISAQOL-IMI: Setting PRO research objective priorities

Defining clear PRO objectives for cancer clinical trials has been a challenging task. PRO objectives in cancer clinical trials tend to be vague (e.g., to demonstrate that health-related quality of life is better with Treatment A than with Treatment B), leading to the use of varied analysis methods and producing seemingly conflicting PRO findings.[14] Therefore, as a starting point, the SISAQOL-IMI Consortium agreed on the importance of creating a priority set of PRO research objectives for which to evaluate and develop design and analysis recommendations. The taxonomy of PRO research objectives developed by the first SISAQOL Consortium was used to achieve this goal.

Prior to the SISAQOL-IMI kick-off meeting, each of the 41 organizations within the SISAQOL-IMI Consortium was presented with a series of PRO objectives previously identified by the first SISAQOL Consortium in a survey (see Appendix 1 for the survey questions). Each organization was encouraged to consult with various internal experts to produce a single response. To ensure that the patient voice is well-represented in the Consortium, the Workgroup of European Cancer Patient Advocacy Networks (WECAN) is one of the organizations in SISAQOL-IMI. WECAN is an umbrella organization representing 23 Pan-European cancer patient organisations. The views and responses of these patient networks are coordinated by Myeloma Patients Europe (MPE) and are sent in as a single response to the survey.

A PRO research objective was identified as high priority if at least two-thirds of the organization representatives responded “yes”. An objective was identified as low priority if less than half of the representatives responded “yes” on the PRO objective. A statement was “for discussion” if it did not meet the high or low priority criteria. Organization representatives who responded “don’t know” for a specific objective were not included in the denominator when calculating percent agreement.

An overall summary of the agreed high priority broad PRO objectives and endpoints is presented in the next sections below and a plain language version summary can be found on the project website as PowerPoint Presentation with voice-over narration, https://www.sisaqol-imi.org (see Appendix 4 for the slides). However, it should be noted that when developing a consensus position, reaching 100% agreement among the Consortium members is not always possible. Even if there is high level of agreement and a statement is accepted by two-thirds majority, there might be stakeholders with substantive concerns or different views which need to be considered. Additionally, the number of organizations representing each stakeholder group differ and stakeholder groups which are less represented may be underrepresented in the voting results.

To address these concerns, a diverging views document was drafted and agreed upon by all Consortium members (see Appendix 2). In addition to the overall results of the priority setting, more detailed results on the level of agreement by stakeholder group are also provided. Tables 2 and 3 presents the overall results of the priority

setting of PRO research objectives and initial stakeholder concerns and views on these objectives for RCTs and single arm studies, respectively. Tables 4 and 5 describe in more detail the level of agreement by stakeholder group for each PRO research objective and endpoint.

[INSERT TABLE 2, 3, 4 and 5 HERE]

Descriptive/exploratory PRO objectives. For both RCTs (32/39, 82% of survey participants that responded to this question) and single-arm studies (36/41, 88%), it was considered high priority to develop standards for descriptive PRO objectives (i.e., describing PRO data without drawing confirmatory conclusions; no hypothesis testing is conducted). Currently, many trials include PROs as descriptive / exploratory objectives, in addition to other primary or secondary endpoints. Given the lack of recommendations, the quality of the collected data tends to be substandard (e.g., high rates of missing data) and there is a risk of selective reporting. Other non-PRO trial data is often presented descriptively (e.g., CTCAE safety data, dose modifications, etc.), and the goal of SISAQOL-IMI is to improve the design, analysis, presentation, and interpretation of all PRO data, including descriptive objectives. This will ensure that the gathered data are not wasted; instead, the data can be used to reliably describe the patient perspective regarding treatment and inform the decisions of various stakeholders.

Confirmatory PRO objectives (superiority and equivalence/non-inferiority)*.* For RCTs, it was considered high priority to develop recommendations about the design and analysis of confirmatory PRO objectives, in which PRO data can be used to draw conclusions about treatment efficacy/clinical benefit. Conclusions about treatment efficacy/clinical benefit can be achieved either by demonstrating that the treatment arm is superior (40/41, 98%) or equivalent/non-inferior (38/40, 95%) to the control arm. The clear distinction between superiority and equivalent/non-inferiority as a PRO objective is critical. This will help avoid drawing conclusions about equivalence/non-inferiority from a statistical test addressing a PRO superiority objective. Instead, PRO non-inferiority/equivalence objectives require pre-specification of meaningful non-inferiority/equivalence margins.[166] The goal of SISAQOL-IMI is to ensure that, when PROs are included as confirmatory objectives (i.e., the aim is to either conclude superiority or equivalence/non-inferiority of the

treatment arm relative to the control arm), the standards for the design, analysis, presentation and interpretation of PRO data are on the same level as other clinical endpoints.

PRO endpoint: Magnitude of change at specific time point(s) and response patterns/profiles over a specified time frame. The aim of these two endpoints is to assess the level of change at a specific time point (time as discrete) or time frame (time as continuous) for a specific PRO domain. These endpoints require pre-specifying clinically relevant thresholds at the group level. Magnitude of change and response patterns/profiles (whether measured as change scores or as severity levels) were considered as high priority PRO endpoints for RCTs (magnitude of change: 37/40, 93%; response patterns/profiles: 31/34, 91%) and for single-arm studies (magnitude of change: 33/38, 87%; response patterns/profiles: 28/36, 78%). Divergence in recommendations is expected for the time element; that is, whether time is considered discrete or continuous will have an impact on the design and analysis of these endpoints. For magnitude of change at specific time point(s), initial stakeholder views highlighted the importance of the choice of the relevant time points and defining a clinically relevant change/difference for both superiority and non-inferiority objectives. For response patterns/profiles over a specified time frame, concerns were raised regarding the feasibility of developing a pre-defined hypothesis to implement this endpoint for a confirmatory objective, and that this endpoint may be more useful in a descriptive setting.

Time to improvement and Time to worsening*.* The aim of these two PRO endpoints is to evaluate the time it takes before a clinically relevant improvement (or worsening) is observed. For PRO domains or items that tend to be more susceptible to change, additional information on sustained improvement (or worsening) will be relevant to describe the change (e.g., duration of improvement or worsening). These two endpoints require pre-specifying a clinically meaningful improvement (or worsening) at the patient level. Time to improvement and time to worsening were considered as high priority for both RCTs (time to improvement: 32/37, 86%; time to worsening: 37/40, 93%) and single-arm studies (time to improvement: 30/39, 77%; time to worsening: 31/38, 82%). Although these two PRO endpoints have differing

within-treatment assumptions (i.e., whether an improvement or worsening is expected among the patients), when formulating recommendations, they will be evaluated together because both rely on time to event design and analysis assumptions. Initial stakeholder views indicated concerns regarding the implementation of this endpoint, including defining a clinically relevant PRO event (improvement or worsening), the need for prolonged relatively high frequency of assessments, lack of standards on how intercurrent events such as death would be addressed, and the difficulty in interpreting the resulting treatment estimates. Additionally, this endpoint assumes that if all patients are followed up long enough, they will eventually experience an improvement (time to improvement) or worsening (time to worsening) on that specific PRO domain.

Responder improvement and responder worsening at specific time point(s). The aim of these two PRO endpoints is to identify the number of patients with an improvement (or worsening) at a specific time point for a specific PRO domain. These endpoints require pre-specifying a meaningful improvement (or worsening) at the patient level. Responder improvement and responder worsening were considered as high priority for both RCTs (responder improvement: 34/38, 89%; responder worsening: 31/37, 84%) and single-arm studies (responder improvement: 30/38, 79%; responder worsening: 26/36, 72%). Similar to the time to event outcomes, when formulating recommendations for responder analyses, these two PRO endpoints will be evaluated together since both rely on similar design and analysis assumptions. Initial stakeholder views highlighted the importance of the choice of relevant time points and defining a clinically meaningful responder (improvement/worsening). Whether the responder will be defined based on an absolute value or change scores needs to be considered.

Overall average/median over a specified time frame and area under the curve over a specified time frame. The aim of these two PRO endpoints is to summarize all available scores for a specific PRO measure over a pre-specified time frame into a single data point per patient (either the average/median or area under the curve). Pre-defined clinically relevant thresholds at the group level are needed to aid interpretation of these endpoints. Overall average/median over a specified time frame and area under the curve over a specified time frame were considered as high priority for RCTs (overall average/median: 27/34, 79%; area under the curve: 32/35,

91%), but not for single-arm studies (overall average/median: 22/33, 67%; area under the curve: 21/35, 60%). When formulating recommendations, these two endpoints will be evaluated together since they are both summary measures that rely on similar design and analysis assumptions. Initial stakeholder views indicated concerns regarding the interpretation of these two endpoints. That is, different patterns of observed PRO assessments may lead to similar AUC or average scores, which may make the interpretation of these endpoints challenging. An added concern if it is used in single-arm trials, it becomes even more difficult to interpret in the absence of a control group.

Procedure of SISAQOL-IMI: Developing consensus recommendations

A critical part of SISAQOL-IMI will be to ensure recommendations for the prioritized objectives and endpoints are based on consensus and address the needs of relevant, key stakeholder groups. Since the goal of SISAQOL-IMI is to improve standards, it is critical that the consensus recommendations balance high methodological quality and feasibility.

Many researchers agree that no design and statistical method exists that can address all concerns. The choice of design and statistical methods is always a balance between feasibility, usefulness, and robustness, and is highly dependent on the study aims. Therefore, SISAQOL-IMI will ensure that the strengths and limitations of each recommendation will be specified based on evidence from the methodological literature, and that deviations from recommendations are acceptable with justification.

Based on literature reviews and expert discussions, the WPs will generate lists of recommendation statements that consortium members will review and provide feedback on via surveys. Five consensus meetings will be held where results will be presented and discussed with all SISAQOL-IMI members. During the planned consensus meetings, SISAQOL-IMI members will agree on recommendation statements through a defined consensus process. According to the rules agreed on by the SISAQOL-IMI Consortium, a proposed statement is accepted if at least two-thirds of the voters agreed on the statement. A statement is rejected if less than half

of the voters agreed on the statement. A statement is postponed or noted for discussion if it does not meet the agreement or rejection criteria, or if it is agreed by the consortium that more discussion is needed. Even if there is high level of agreement and a statement is accepted by two-thirds majority, there might be stakeholders with substantive concerns or different views. To address these various views whilst maintaining the notion of consensus, the SISAQOL-IMI Consortium will note where readers should take into consideration that individual organizations might have different views on specific recommendations given their institutional or stakeholder standpoint. To this purpose, for each *accepted* recommendation, together with the percentage agreement, the other diverging views will be included under “considerations”. In addition, a table showing percentage agreement by stakeholder group will be presented to inform readers about which stakeholder group might have a different position on a given recommendation statement. Finally, SISAQOL-IMI will provide concrete reasons when recommendation statements do not reach consensus for standards of methodological quality. For more details on this, see the diverging views document in Appendix 2.

To ensure that the consensus recommendation statements have external validity, three independent processes will be implemented by the SISAQOL-IMI Consortium. A *scientific work package* aims to independently validate the recommendation statements by assigning “blinded members” (i.e., SISAQOL-IMI organizations will assign this task to colleagues who are not involved in the development of the recommendations) to test the feasibility of implementing the recommendations and to provide feedback on the formulation of the recommendation statements. Blinded members will be given tasks with respect to writing a study protocol, the statistical analysis plan (SAP), and the visualization and presentation of results. They will be asked to complete these tasks making use of the recommendations. An *independent scientific advisory board* has been set-up to provide independent and critical review of the scientific quality of the recommendations. Finally, based on discussions with the European Medicines Agency (EMA), submission of SISAQOL-IMI outputs or recommendations for *qualification advice or opinion for novel methodologies* at EMA are being pursued. All these independent processes occur during the lifetime of the SISAQOL-IMI project to ensure that feedback from external experts and stakeholders are considered in the final SISAQOL-IMI recommendation statements.

Critical to the work of SISAQOL-IMI is the involvement of patients, caregivers and patient representatives in the development of these recommendations. This stakeholder group has a deep understanding and experience with the disease and treatment. They provide valuable insight into the research questions to be asked (research objectives), study design (timing and frequency of assessments) and interpretation of PRO findings (meaning to patients and their families). Their contribution will help make the SISAQOL-IMI recommendations relevant and meaningful to the patient experience.

Since all recommendations from SISAQOL-IMI are based on a multi-stakeholder consensus process, the proposed recommendation statements for the consensus meetings and the final recommendations should be interpretable by stakeholders with statistical and non-statistical backgrounds. To achieve this objective, different experts (statisticians, methodologists, PRO experts, clinicians, patients, and caregivers and patient representatives) are included in the process of developing these recommendations. To facilitate discussion among the stakeholders and to ensure harmonised terminology across WPs, an extensive glossary with both scientific and plain-language terminology and definitions will be developed alongside the consensus statements.

Two final consensus recommendation documents will be developed: (1) a technical document with an integrated technical/statistical section for statisticians and similar stakeholders who will need to execute or review the design, collection, analysis, and written/visual data interpretations, and (2) a plain-language version that can be used by the patient, caregiver, and clinical communities to facilitate interpretation and communication of the technical document and support the communication between users with different levels of statistical/technical expertise. This will allow parallel discussions at different statistical/technical levels and safeguard the importance of having recommendations that are understandable and meaningful to stakeholders irrespective of their statistical knowledge and methodological background.

Conclusion

The aim of SISAQOL-IMI is to improve how PROs are used in cancer clinical trials by developing a consensus-based set of best practice recommendations for the

design, collection, analysis, presentation and interpretation of PRO endpoints. The SISAQOL-IMI kick-off meeting was attended by all 41 SISAQOL-IMI organizations, representing various international stakeholder groups. Views from this meeting demonstrated the shared interest and commitment of the different organizations in improving standards for PRO endpoints in cancer clinical trials. A set of priority PRO objectives was agreed upon, for which SISAQOL-IMI will develop recommendations by the end of 2024. By the time that this manuscript has been completed, SISAQOL-IMI already had a second consensus meeting and agreed on their first set of recommendations. This (virtual) meeting was attended by all 41 organizations, showing a strong commitment to complete this work. A third consensus meeting is already planned for 2023. A separate manuscript will be drafted to present these recommendations by the end of 2024.

Continuing from the initial achievement of the first SISAQOL Consortium, these recommendations will contribute to the optimal use and understanding of the role of PRO measures in academic research, drug development, and approval of therapies by regulators and reimbursement decisions by HTA bodies. Having standards set for the use of PROs in cancer clinical trials will address the need to have more robust evidence on the impact of cancer treatments on patients' symptoms, functioning and general health related quality of life. This will also subsequently facilitate communication on benefits and risks of various cancer therapies by expanding existing information. Finally, it is also worth noting that some of the recommendations of SISAQOL-IMI are likely to applicable to other therapeutic areas, beyond oncology.

Contributors
All authors are members of the SISAQOL-IMI Steering Committee and were involved in the conceptualisation of this manuscript. M.Pe, A. Alanya, R. S.Falk, C.D. Amdal, K. Bjordal collected and analysed the data. M.Pe led in the drafting of the first version of the manuscript, with support from A. Alanya, R. S.Falk, C.D. Amdal, K. Bjordal A. Bottomley and I. Griebsch. All authors interpreted and reviewed the manuscript. All 41 organizations involved in SISAQOL-IMI (see member list in the on behalf of SISAQOL-IMI) reviewed and approved the final version of this manuscript.

Declaration of interests
This publication reflects the views of the individual authors and should not be construed to represent official views or policies of the European Medicines Agency (EMA), the US Food and Drug Administration (FDA), US National Cancer Institute (NCI), Medicines and Healthcare products Regulatory Agency (MHRA), Institute for Quality and Efficiency in Health Care (IQWiG), Health Canada, the Norwegian Medicines Agency (NOMA), the American Society of Clinical Oncology (ASCO) or the European Society for Medical Oncology (ESMO) or any other institution, organization, or entity. This publication reflects the authors' view and that neither IMI nor the EU, EFPIA are responsible for any use that may be made of the information contained therein.

SR is a current employee of Pfizer Inc and ex-employee of Novartis Pharma; JC, PC, and JR are current employees of Pfizer. All other authors declare no competing interests. VP is a current employee of EMD Serono. MS is a current employee of Merk, IG is a current employee of EMD Serono. GV had received consulting fees, payment from or were related to the following organizations Pfizer, Eisai, Roche, Novartis, Astra Zeneca, Sanofi, Seattle Genetics, EORTC QoL Group, EORTC Board. KO's organization has received sponsorship funding/grants for various of our annual programmes/activities from the following companies: Bristol-Myers Squibb, Novocure, Pfizer, Bayer, Novartis, Northwest Biotherapeutics, Karyopharm, MagForce, Medac, Photonamic, Apogenix, Elekta, GW Pharmaceuticals/Jazz Pharmaceuticals, consulting fees from Bristol-Myers Squibb and Novartis, honoraria from Sanofi, Sharing Progress in Cancer Care and Seagen. KO participated in an advisory board for Novartis, Novocure, Seagen, Eisai, BMS, Sanofi and undertook leadership roles in a number of organizations.

This study received no funding from the US National Institutes of Health (NIH). No other authors were fully or partly NIH funded, employed by NIH, or are in receipt of an NIH grant for this work.

References

1 European Medicines Agency OWP. European Medicines Agency Reflection Paper on the use of patient reported outcome (PRO) measures in oncology studies [Draft]. 2014. https://www.ema.europa.eu/en/documents/scientific-guideline/draft-reflection-paper-use-patient-reported-outcome-pro-measures-oncology-studies_en.pdf

2 Kluetz PG, O'Connor DJ, Soltys K. Incorporating the patient experience into regulatory decision making in the USA, Europe, and Canada. *Lancet Oncol* 2018 May; **19(5)**: e267–74. DOI: 10.1016/S1470-2045(18)30097-4

3 Hamel J-F, Saulnier P, Pe M, Efstathios Z, Musoro J, Coens C *et al.* A systematic review of the quality of statistical methods employed for analysing quality of life data in cancer randomised controlled trials. *Eur J Cancer* 2017 Sep; **83**: 166–76. DOI:https://doi.org/10.1016/j.ejca.2017.06.025

4 Fiteni F, Anota A, Westeel V, Bonnetain F. Methodology of health-related quality of life analysis in phase III advanced non-small-cell lung cancer clinical trials: a critical review. *BMC Cancer* 2016 Feb 18; **16**: 122. DOI: 10.1186/s12885-016-2152-1

5 Pe M, Dorme L, Coens C, Basch E, Calvert M, Campbell A, *et al.* Statistical analysis of patient-reported outcome data in randomised controlled trials of locally advanced and metastatic breast cancer: a systematic review. *Lancet Oncol* 2018 Sep; **19(9)**: e459–69. DOI:https://doi.org/10.1016/S1470-2045(18)30418-2

6 Fiero MH, Roydhouse JK, Vallejo J, King-Kallimanis BL, Kluetz PG, Sridhara R. US Food and Drug Administration review of statistical analysis of patient-reported outcomes in lung cancer clinical trials approved between January, 2008, and December, 2017. *Lancet Oncol* 2019 Oct; **20(10)**: e582–9. DOI:https://doi.org/10.1016/S1470-2045(19)30335-3

7 Fernandes LL, Zhou J, Kanapuru B, Horodniceanu E, Gwise T, Kluetz PG *et al.* Review of patient-reported outcomes in multiple myeloma registrational trials: highlighting areas for improvement. *Blood Cancer J* 2021 Aug; **11**: 148. DOI:10.1038/s41408-021-00543-y

8 Safa H, Tamil M, Spiess PE, Manley B, Pow-Sang J, Gilbert SM *et al.* Patient-Reported Outcomes in Clinical Trials Leading to Cancer Immunotherapy Drug Approvals From 2011 to 2018: A Systematic Review. *J Natl Cancer Inst* 2021 May; **113(5)**: 532–42. DOI: 10.1093/jnci/djaa174

9 Calvert M, Kyte D, Mercieca-Bebber R, Slade A, Chan AW, King MT *et al.* Guidelines for Inclusion of Patient-Reported Outcomes in Clinical Trial Protocols. *JAMA* 2018 Feb; **319(5)**: 483. DOI: 10.1001/jama.2017.21903

10 Calvert M, King M, Mercieca-Bebber R, Aiyegbusi O, Kyte D, Slade A *et al.* SPIRIT-PRO Extension explanation and elaboration: guidelines for inclusion of

patient-reported outcomes in protocols of clinical trials. *BMJ Open* 2021 Jun; **11(6)**: e045105. DOI: 10.1136/bmjopen-2020-045105

11 Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD. Reporting of patient-reported outcomes in randomized trials: The CONSORT PRO extension. *Jama* 2013 Feb; **309(8)**: 814–22. DOI: 10.1001/jama.2013.879

12 Snyder C, Smith K, Holzner B, Rivera YM, Bantug E, Brundage M. Making a picture worth a thousand numbers: recommendations for graphically displaying patient-reported outcomes data. *Qual Life Res* 2019 Feb; **28(2)**: 345–56. DOI: 10.1007/s11136-018-2020-3

13 Altman DG, Sauerbrei W, McShane LM. Importance of the distinction between quality of methodology and quality of reporting. *HPB* 2017 Jul; **19(7)**: 649–50. DOI: 10.1016/j.hpb.2017.02.444

14 Bottomley A, Pe M, Sloan J, Basch E, Bonnetain F, Calvert M *et al.* Analysing data from patient-reported outcome and quality of life endpoints for cancer clinical trials: a start in setting international standards. *Lancet Oncol* 2016 Nov; **17(11)**: e510–4. DOI: 10.1016/S1470-2045(16)30510-1

15 Coens C, Pe M, Dueck AC, Sloan J, Basch E, Calvert M *et al.* International standards for the analysis of quality-of-life and patient-reported outcome endpoints in cancer randomised controlled trials: recommendations of the SISAQOL Consortium. *Lancet Oncol* 2020 Feb; **21(2)**: e83–96. DOI: 10.1016/S1470-2045(19)30790-9

16 ICH Harmonised Tripartite Guideline. Statistical principles for clinical trials. International Conference on Harmonisation E9 Expert Working Group. *Stat Med* 1999 Aug; **18(15)**: 1905–42. Available from: https://www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials-scientific-guideline

17 Innovative Medicines Initiative (IMI), IMI2 - Call 18. 2019. https://www.imi.europa.eu/apply-funding/closed-calls/imi2-call-18 (accessed Nov 11, 2021).

18 Goldman M, Seigneuret N, Eichler H-G. The Innovative Medicines Initiative: an engine for regulatory science. *Nat Rev Drug Discov* 2015 Jan; **14(1)**: 1–2. DOI: 10.1038/nrd4520

19 Laverty H, Meulien P. The Innovative Medicines Initiative −10 Years of Public-Private Collaboration. *Front Med* 2019 Dec; **6**. DOI:10.3389/fmed.2019.00275.

20 Bushmakin AG, Cappelleri JC. A Practical Approach to Quantitative Validation of Patient-Reported Outcomes: A Simulation-Based Guide Using SAS. Hoboken, New Jersey: John Wiley & Sons. 2022.

21 Cappelleri JC, Zou KH, Bushmakin AG, Alvir JMJ, Alemayehu D, Symonds T. Patient-Reported Outcomes: Measurement, Implementation and Interpretation. Boca Raton, Florida: Chapman & Hall/CRC Press. 2013.

22 de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine. Cambridge University Press, 2011 DOI:10.1017/CBO9780511996214.

23 Peter M. Fayers DM. Quality of Life: The Assessment, Analysis and Reporting of Patient-reported Outcomes. 2016.

24 Streiner DL, Norman GR, Cairney J. Health Measurement Scales. Oxford University Press, 2015 DOI:10.1093/med/9780199685219.001.0001.

25 ICH Expert Working Group. International council for harmonisation of technical requirements for pharmaceuticals for human use: addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials E9(R1). 2019. https://www.ich.org/page/efficacy-guidelines#9-2 (accessed Sept 19, 2022).

26 European Society for Medical Oncology. ESMO- Magnitude of Clinical Benefit Scale (ESMO-MCBS). https://www.esmo.org/guidelines/esmo-mcbs (accessed Nov 11, 2021).

27 Fiero MH, Pe M, Weinstock C, King-Kallimanis B, Komo S, Kelpin HD *et al.* Demystifying the estimand framework: a case study using patient-reported outcomes in oncology. *Lancet Oncol* 2020 Oct; **21(10)**: e488–94. DOI:https://doi.org/10.1016/S1470-2045(20)30319-3

28 European Medicines Agency. Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man - the use of patient-reported outcome (PRO) measures in oncology studies - Scientific guideline. 2016. https://www.ema.europa.eu/en/appendix-2-guideline-evaluation-anticancer-medicinal-products-man-use-patient-reported-outcome-pro.

29 US Food and Drug Administration. Patient-reported outcome measures: use in medical product development to support labelling claims. 2009. https://www.fda.gov/media/77832/download (accessed Nov 23, 2021).

30 European Network for Health Technology Assessment. Methodological guideline for REA of pharmaceuticals: Health-related quality of life. 2013. https://www.eunethta.eu/methodological-guideline-for-rea-of-pharmaceuticals-health-related-quality-of-life/ (accessed Nov 23, 2021).

31 Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. General Methods version 6.0. 2020. https://www.iqwig.de/methoden/general-methods_version-6-0.pdf (accessed Nov 23, 2021).

32 Oosting, S.F., Barriuso, J., Bottomley, A., Galotti, M., Gyawali, B., Kiesewetter, B., Latino, N.J., Martinelli, F., Pe, M., Pentheroudakis, G. and Roitberg, F., Methodological and reporting standards for quality-of-life data eligible for European Society for Medical Oncology-Magnitude of Clinical Benefit Scale (ESMO-MCBS) credit. *Annals of Oncology* 2022 Dec. DOI: 10.1016/j.annonc.2022.12.004

TABLES

Table 1. SISAQOL-IMI Consortium members

Stakeholder Group	Organisations (alphabetical order)	Country
Academic (n=17)		
	Amsterdam UMC	NL
	Clinical Hospital Center Rijeka	HR
	Complejo Hospitalario de Navarra	ES
	Duke University School of Medicine	US
	Institut Hospital del Mar d'Investigacions Mèdiques	ES
	Johns Hopkins University (representing PROTEUS and PRO Data Presentation Stakeholder Group)	US
	Katholieke Universiteit Leuven	BE
	Leiden University Medical Center	NL
	Medical University of Innsbruck	AT
	Oslo University Hospital	NO
	Region Hovedstaden	DK
	The Symptoms Tool Executive Committee of the University of Texas MD Anderson Cancer Center (The Texas Group)	US
	University Health Network	CA
	University of Birmingham	UK
	University of Freiburg	DE
	University of Ghent	BE
	University of Leeds	UK
Industry (n=5)		
	AbbVie	US/DE

	Bayer	US/DE
	Boehringer Ingelheim	DE
	Merck Healthcare KGaA/EMD Serono	DE/US
	Pfizer	US/UK
Non-profit/Cancer organizations (n=8)		
	American Society of Clinical Oncology	US
	Critical Path Institute	US
	European Organisation for Research and Treatment of Cancer	BE
	European Society for Medical Oncology	CH
	National Cancer Center Hospital (Japan Clinical Oncology Group)	JP
	National Cancer Institute	US
	Queen's University at Kingston (CCTG)	CA
	University of Sydney (Sydney Quality of Life Office)	AU
Small to mid-size enterprise (SME)/Contract research organization (CRO) (n=4)		
	Adelphi Values	UK
	Evaluation Software Development	AT
	Modus Outcomes	FR
	Patient Relevant Evidence	US
Regulatory (n=4)		
	European Medicines Agency	EU (NL)
	Health Canada	CA
	Medicines and Healthcare products Regulatory Agency	UK
	US Food and Drug Administration	US
Health technology assessment (HTA) (n=2)		
	Institute for Quality and Efficiency in Health Care	DE
	Norwegian Medicines Agency	NO

Patients' representative (n=1)		
	Myeloma Patients Europe (on behalf of the Workgroup of European Cancer Advocacy Networks – WECAN consisting of 23 Pan-European cancer patient organisations.)	BE

Table 2.

Results from Priority Setting of PRO Objectives/Endpoints for RCTs (n = 41)

	Randomized controlled trial			Initial views from different SISAQOL-IMI members
"Would you or your organisation consider using…"	Yes, n	No, n	Yes, %*	
Clinical benefit/treatment efficacy objective (confirmatory: superiority) *Definition:* to demonstrate that based on a PRO domain, the treatment group is superior to (or better than) the reference group by a clinically relevant treatment effect size	40	1	98	• Superiority PRO objective can be used for a primary or secondary endpoint but its assessment and interpretation in conjunction with other clinical endpoints should be considered. Justification should be provided. For example: PROs can be used to differentiate treatments (demonstrate superiority) in a non-inferiority RCT where the primary endpoint is progression free survival. • No significant difference in a superiority objective does not imply equivalence or non-inferiority. • Standards for PRO design and analyses should be treated in the same standard as other survival or response endpoints (e.g., should indicate potential sources of biases and how this is minimized, define what a validated endpoint is, pre-planned in the Statistical Analysis Plan [hierarchical testing], be objectively measured [validated PRO measure for the patient population]). • Scores and differences from the PRO endpoints should be interpretable (e.g., what does a 10 point difference mean?). Both statistical significance and clinically meaningful difference is critical for drawing comparative conclusions on PROs (control for type 1 error should be considered, implications for sample size calculation, what is a clinically meaningful difference?).
Clinical benefit/treatment efficacy objective (confirmatory: equivalence/non-inferiority) *Definition:* to demonstrate that based on a PRO domain, the treatment group is similar (equivalent) or not worse (non-inferior) than the reference group by a pre-specified clinically relevant margin	38	2	95	• Non-inferiority/equivalence PRO objective can be used for a primary or secondary PRO endpoint but its assessment and interpretation in conjunction with other clinical endpoints should be considered. Justification should be provided. • Ensure that the PRO measure is valid and reliable for the target population and is responsive to change. • Ensure that non-inferiority / equivalence margins are pre-specified.

Descriptive objective (exploratory/descriptive) *Definition:* to present PRO findings but no comparative conclusions between treatment arms will be drawn	32	7	82	• PROs are often used as descriptive objectives with the idea that this can complement primary and secondary clinical objectives. • Descriptive / exploratory objectives lack adequate rigor and does not allow for comparison of PRO results between groups. There are concerns about the robustness of the data generated, lack of pre-defined hypothesis (including validated instruments) and analysis of PRO endpoints. • Improving standards on how PROs are analysed even for descriptive / exploratory is needed to allow better use of this data (e.g., data must be of high quality).
Time to improvement *Definition:* the time it takes before a clinically relevant improvement from a PRO domain is observed within a pre-specified timeframe.	32	5	86	• This objective assumes that if all patients are followed up long enough, they will experience an improvement. Therefore, this can be used if the expected assumption is that patients will improve in that PRO domain (e.g., patients undergoing primary curative therapy with acute toxicity or trials where the goal is to alleviate patients' symptoms). This objective may not be appropriate for early diagnosis with minimal symptoms (where no improvement in symptoms is expected) or for patients with poor prognosis (where an improvement is not expected). • Additional information on sustained improvement will be relevant to describe this endpoint (e.g., duration of the improvement) since PRO concepts and symptoms may be more volatile relative to other clinical endpoints (e.g., death or PFS). The relative proportion of responders would also be important to report for each arm (similar to time to response for other clinical endpoints). • Defining a clinically meaningful event (improvement) is important. • General design and statistical issues for time to event objectives need to be addressed. This endpoint may require relatively frequent assessments over a potentially long period of time.
Time to worsening *Definition:* the time it takes before a clinically relevant worsening from a PRO domain is observed within a pre-specified timeframe	37	3	93	• Time to worsening is a common endpoint in cancer clinical trials since the expectation is that patients tend to worsen over time for various PRO domains and symptoms (e.g., due to toxicity). It is highly relevant for cancers with poor prognosis where maintaining functioning will be the main goal. This endpoint can capture the start of experiencing a (symptom/domain) worsening. • This may also be a more relevant objective because worsening signals the end of the period of sufficient favourable effects [of a medicine] for a patient. It is therefore closely related to important favourable effects that determine the magnitude of benefits of a medicine (e.g., duration of response). • Additional information on sustained worsening will be relevant to describe this endpoint (e.g., duration of the worsening) since PRO domains and symptoms may be more volatile relative to other clinical endpoints (e.g., death or PFS). • Defining a clinically meaningful event (worsening) is important.

				• General design and statistical issues for time to event objectives need to be addressed. This endpoint may require relatively frequent assessments over a potentially long period of time.
Time to stable state *Definition:* After observing a clinically relevant change from baseline (either worsening or improvement) from a PRO domain, the time it takes before the PRO domain returns to its baseline value (i.e., no change from baseline or as change from baseline within the predefined baseline margin)	21	11	66	• Time to stable state is not often seen as an endpoint in cancer clinical trials. However, this may be relevant in contexts where patients may experience a deterioration from their baseline state, but it will be temporary (e.g., if patients have good functioning at baseline, or minimal symptoms; and the expectation is that patients will experience a temporary deterioration, and they will go back to baseline after the deterioration). • Currently the definition will measure the stable state after patients have improved or worsened, which will be hard to interpret. • If time to stable state is used, this may no longer be a comparison of randomized groups because this endpoint assumes that patients have to experience a change (e.g., worsen) before one can measure the event, "time to stable state". Non responders (those who do not change) will be excluded from analyses. • Defining a clinically meaningful event (stable state) is important. • General design and statistical issues for time to event objectives need to be addressed. This endpoint may require relatively frequent assessments over a potentially long period of time.
Time to end of stable state *Definition:* The time it takes until the stable state ends or time until a clinically relevant improvement or worsening from a PRO domain is observed.	14	12	54	• Currently, the definition implies ending the stable state by both improvement and worsening, which will be hard to interpret. • This endpoint may be difficult to implement in cancers with poor prognosis because of the short disease duration (less frequent assessments). • Time to improvement and time to worsening could also address this objective. • Additional information on sustained "end of stable state" will be relevant to describe this endpoint (e.g., duration of the worsening) since PRO concepts and symptoms may be more volatile relative to other clinical endpoints (e.g., death or PFS). • Defining a clinically meaningful event (end of stable state) is important. • General design and statistical issues for time to event objectives need to be addressed. This endpoint may require relatively frequent assessments over a potentially long period of time.
Magnitude of change at specific time point(s) *Definition:* The actual value or change from baseline value for a PRO domain at pre-defined time points	37	3	93	• Defining the relevant timepoint would imply making an assumption on when an improvement or worsening will happen across patients. • Defining a clinically relevant change/difference for both superiority and non-inferiority objectives will be critical (e.g., MID) for this endpoint. • Whether the absolute value or change scores will be used as data needs to be taken into account.

Responder with improvement at specific time point(s) *Definition:* Whether the value (or change from baseline value) from a PRO domain at a specific time point reaches a pre-defined improvement threshold or not	34	4	89	• Defining the relevant timepoint would imply making an assumption on when an improvement or worsening will happen across patients. • Defining a clinically meaningful responder (improvement) is important. • This can be used for supportive information for a clinical finding and support interpretation for other PRO endpoints (e.g., magnitude of change or time to event endpoints). • Is there a possibility to include stable state in the definition of improvement (i.e., patients who did not worsen)?
Responder with worsening at specific time point(s) *Definition:* Whether the value (or change from baseline value) from a PRO domain at a specific time point reaches a pre-defined worsening threshold or not.	31	6	84	• Defining the relevant timepoint would imply making an assumption on when an improvement or worsening will happen across patients. • Defining a clinically meaningful responder (worsening) is important. • This can be used for supportive information for a clinical finding and support interpretation for other PRO endpoints (e.g., magnitude of change or time to event endpoints).
Responder with stable state at specific time point(s) *Definition:* Whether the value (or change from baseline value) from a PRO domain at a specific time point remains within a pre-defined baseline margin or not.	19	12	61	• This may be relevant for trials where the goal is "maintenance of a specific PRO domain". • Defining the relevant timepoint would imply making an assumption on when stable state will happen across patients. • Defining a meaningful responder (stable state) is important. • This can be used for supportive information for a clinical finding and support interpretation for other PRO endpoints (e.g., magnitude of change or time to event endpoints). • When interpreting results for stable state alone (responder) this will imply merging patients who improved and worsened into one category (non-responders). • It may be more meaningful to report descriptively improved/stable/worsened by time point (rather than just stable state). • The focus of interpretation is usually the number of patients who improved or worsened, rather than those who remained stable. • Is there a possibility to include stable state in the definition of improvement (i.e., patients who did not worsen)?
Overall average or median over a specified timeframe *Definition:* The average or median score of all available	27	7	79	• This definition needs to be clarified as overall average or median over time for an individual (and not group). • The advantage of this endpoint is that it allows the assessment of PROs over an entire time frame, making full use of the information that was collected.

scores from a PRO domain over a pre-specified timeframe.				• The concern for this endpoint lies on the interpretation since averages over time wash out changes at specific time points. It then puts into question how the resulting estimate can be understandable and interpretable by patients. • This has been suggested as a useful approach if the planned assessments differ between trial participants (e.g., stopped assessments due to death and progression). However, there are concerns that this may not be an appropriate approach to handle these "missing data" since they are usually missing not at random.
Area under the curve over a specified timeframe *Definition:* The area under the curve value of all available scores from a PRO domain over a pre-specified timeframe.	32	3	91	• This endpoint requires the timeframe specified to be meaningful and comparable between the two arms. • There are concerns over interpretation of an area under the curve result (i.e., different patterns of observed assessment that reflects different clinical realities may lead to similar AUC). • This endpoint has strong missing data assumptions.
Best score over a specified time frame *Definition:* The best score of all available scores from a PRO domain over a pre-specified timeframe	12	17	41	• Best / worst score over a time frame is prone to bias (and measurement error). For example, one assessment point can have a high score and rest is low, and the interpretation of the results will be biased towards showing benefit (high score). • Timing of best score will also differ for each patient. • One assessment of a best score and not knowing the duration of that (e.g., if temporary or sustained) is not useful in terms of clinical relevance. • There is a need to determine whether the best score is clinically meaningful. • There is a need to find a context where this endpoint will be relevant.
Worst score over a specified time frame *Definition:* The worst score of all available scores from a PRO domain over a pre-specified timeframe	15	18	45	• Best / worst score over a time frame is prone to bias (and measurement error). For example, one assessment point can have a low score and rest is high, and the interpretation of the results will be biased towards showing risks/harm (low score). • This is similar to how CTCAE / safety data is reported and makes the PRO reporting similar to the clinician reporting. • There is a need to determine whether the worst score is clinically meaningful.
Response patterns/profiles over a specified time frame *Definition:* The longitudinal pattern of all available scores from a PRO domain over a pre-specified timeframe	31	3	91	• It will be difficult to develop a pre-defined hypothesis for comparability and efficacy for this PRO endpoint. • This is more useful for descriptive rather than confirmatory objectives (superiority / non-inferiority).

Note: Definitions for PRO objectives are based on the SISAQOL recommendations[15].

For each PRO objective, organization representatives were asked whether they or their organization would consider using the specified objective in cancer randomized controlled trials (RCTs) or single-arm studies. Organizations could respond “yes”, “no” or “don’t know” to each item. If organizations replied no to a PRO objective, they were encouraged to state their reasons. An open-ended question was also included to capture additional PRO objectives.
A PRO research objective was identified as high priority if at least two-thirds of the organization representatives responded “yes”. An objective was identified as low priority if less than half of the representatives responded “yes” on the PRO objective. A statement was “for discussion” if it did not meet the high or low priority criteria. Organization representatives who responded “don’t know” for a specific objective were not included in the total number of responses. Initial views from different SISAQOL-IMI members are summarized by qualitative comments from the survey, which can guide further discussions for each objective or endpoint.
The cells with white background show the PRO objectives that reached the high priority threshold (>2/3). Yes (%) is calculated as the number of “yes” votes divided by the total number of “yes” and “no” votes (don’t know or missing is excluded).
Reaching 100% agreement among the Consortium members was not always possible and individual organizations might have different views on specific recommendations given their institutional or stakeholder standpoint. Therefore, we encourage readers to read this table together with Tables 4 and 5 to consider the positions of relevant stakeholder groups.

Table 3.

Results from Priority Setting of PRO Objectives/Endpoints for Single-arm studies (n = 41)

	Single-arm studies			Initial views from different SISAQOL-IMI members
“Would you or your organisation consider using…”	Yes, n	No, n	Yes, %*	
Clinical benefit/treatment efficacy objective (confirmatory: superiority) *Definition:* to demonstrate that based on a PRO domain, the treatment group is superior to (or better than) the reference group by a clinically relevant treatment effect size	23	13	64	• The importance of using PROs in single-arm trials can support the assessment of PROs in rare cancers where RCTs would not be feasible. This is a key area to explore given the number of single-arm trials that are being done in oncology. • There are significant concerns in concluding superiority in single-arm trials based on PRO data. In general, single-arm trials are sensitive to selection bias. Without a comparator arm, the impact of treatment on PROs cannot be disentangled from baseline prognostic factors, making it difficult to contextualize the benefit or impact of treatment for the patient population of interest. • The lack of comparator arm can be addressed by the use of external control groups, historical data (reference data). However, there are questions on the quality of the reference data, and whether the patient population and treatment are comparable between the trial data and the reference data. There is a need to better define the reference population and

				ensure that the external control group/historical data are true comparators for the specific trial. • Even if good reference data is used for a single-arm trial, a confirmatory study using an RCT would still need to be explored.
Clinical benefit/treatment efficacy objective (confirmatory: equivalence/non-inferiority) *Definition:* to demonstrate that based on a PRO domain, the treatment group is similar (equivalent) or not worse (non-inferior) than the reference group by a pre-specified clinically relevant margin	19	19	50	• All comments on superiority objective in single-arm studies apply. • An added concern is the usefulness of a non-inferiority / equivalence objective in a single-arm trial, even if the endpoint was not a PRO.
Descriptive objective (exploratory/descriptive) *Definition:* to present PRO findings but no comparative conclusions between treatment arms will be drawn	36	5	88	• Currently, the use of PROs in single-arm trials would usually be limited to descriptive/exploratory objectives. This can be due to the limited number of patients. The results from a single-arm trial (e.g., phase II) can be used to inform the PRO hypothesis at a later phase. • Standards for descriptive objective need to be better defined, including the use of a validated instrument and PRO data being collected reliably so that this objective can be useful as supportive information.
Time to improvement *Definition:* the time it takes before a clinically relevant improvement from a PRO domain is observed within a pre-specified timeframe.	30	9	77	• This objective can be used if the expected assumption is that patients will improve in that PRO domain (e.g., patients undergoing primary curative therapy with acute toxicity or trials where the goal is to alleviate patients' symptoms). This objective may not be appropriate for early diagnosis with minimal symptoms (where no improvement in symptoms is expected) or for patients with poor prognosis (where an improvement is not expected). • This endpoint can be relevant as a descriptive objective (since reference data are not available). This can also be used in Phase II to guide the analysis planned for Phase III. • Defining a clinically meaningful event (improvement) is important. • General design and statistical issues for time to event objectives need to be addressed. This endpoint may require relatively frequent assessments over a potentially long period of time.
Time to worsening *Definition:* the time it takes before a clinically relevant	31	7	82	• This objective may be appropriate for patients with poor prognosis (where a worsening is expected). • This endpoint can be relevant as a descriptive objective (since reference data are not available). This can also be used in Phase II to guide the analysis planned for Phase III.

worsening from a PRO domain is observed within a pre-specified timeframe				• Defining a clinically meaningful event (worsening) is important. • General design and statistical issues for time to event objectives need to be addressed. This endpoint may require relatively frequent assessments over a potentially long period of time.
Time to stable state *Definition:* After observing a clinically relevant change from baseline (either worsening or improvement) from a PRO domain, the time it takes before the PRO domain returns to its baseline value (i.e., no change from baseline or as change from baseline within the predefined baseline margin)	15	19	44	• Time to stable state is not often seen as an endpoint in cancer clinical trials. However, this may be relevant in contexts where patients may experience a deterioration from their baseline state, but it will be temporary (e.g., if patients have good functioning at baseline, or minimal symptoms; and the expectation is that patients will experience a temporary deterioration, and they will go back to baseline after the deterioration). • Currently the definition implies measurement of stable state after patients have improved or worsened. Non responders (those who do not improve or worsen) will be excluded from analyses. This will further aggravate the issue of small sample size from single-arm trials, which may prevent the robust measurement of this endpoint. • Defining a clinically meaningful event (stable state) is important. • General design and statistical issues for time to event objectives need to be addressed. This endpoint may require relatively frequent assessments over a potentially long period of time.
Time to end of stable state *Definition:* The time it takes until the stable state ends or time until a clinically relevant improvement or worsening from a PRO domain is observed.	15	18	45	• Currently the definition implies ending the stable state by both improvement and worsening, which will be hard to interpret. • This endpoint may be difficult to implement in cancers with poor prognosis because of the short disease duration (less frequent assessments). • Time to improvement and time to worsening could also address this objective. • Additional information on sustained "end of stable state" will be relevant to describe this endpoint (e.g., duration of the worsening) since PRO concepts and symptoms may be more volatile relative to other clinical endpoints (e.g., death or PFS). • Defining a clinically meaningful event (end of stable state) is important. • General design and statistical issues for time to event objectives need to be addressed. This endpoint may require relatively frequent assessments over a potentially long period of time.
Magnitude of change at specific time point(s) *Definition:* The actual value or change from baseline value for a PRO domain at pre-defined time points	33	5	87	• Defining the relevant timepoint would imply making an assumption on when an improvement or worsening will happen across patients. • Defining a clinically relevant change/difference for both superiority and non-inferiority objectives will be critical (e.g., MID) for this endpoint. Will these MIDs take into account the baseline scores of the patient or will they differ depending on where the patient starts on the scale? • Whether the absolute value or change scores will be used as data needs to be taken into account.

Responder with improvement at specific time point(s) *Definition:* Whether the value (or change from baseline value) from a PRO domain at a specific time point reaches a pre-defined improvement threshold or not	30	8	79	• Defining the relevant timepoint would imply making an assumption on when an improvement or worsening will happen across patients. • Defining a clinically meaningful responder (improvement) is important. • This can be used for supportive information for a clinical finding and support interpretation for other PRO endpoints (e.g., magnitude of change or time to event endpoints). • Is there a possibility to include stable state in the definition of improvement (i.e., patients who did not worsen)?
Responder with worsening at specific time point(s) *Definition:* Whether the value (or change from baseline value) from a PRO domain at a specific time point reaches a pre-defined worsening threshold or not	26	10	72	• This is a relevant endpoint to descriptively report toxicity or symptoms. • Defining the relevant timepoint would imply making an assumption on when an improvement or worsening will happen across patients. • Defining a clinically meaningful responder (worsening) is important. • This can be used for supportive information for a clinical finding and support interpretation for other PRO endpoints (e.g., magnitude of change or time to event endpoints).
Responder with stable state at specific time point(s) *Definition:* Whether the value (or change from baseline value) from a PRO domain at a specific time point remains within a pre-defined baseline margin or not.	19	17	53	• This may be relevant for trials where the goal is "maintenance of a specific PRO domain". • Defining the relevant timepoint would imply making an assumption on when stable state will happen across patients. • Defining a meaningful responder (stable state) is important. • This can be used for supportive information for a clinical finding and support interpretation for other PRO endpoints (e.g., magnitude of change or time to event endpoints). • When interpreting results for stable state alone (responder) this will imply merging patients who improved and worsened into one category (non-responders). • It may be more meaningful to report descriptively. improved/stable/worsened by time point (rather than just stable state). • The focus of interpretation is usually the number of patients who improved or worsened, rather than those who remained stable. • Is there a possibility to include stable state in the definition of improvement (i.e., patients who did not worsen)?
Overall average or median over a specified timeframe	22	11	67	• This definition needs to be clarified as overall average or median over time for an individual (and not group).

Definition: The average or median score of all available scores from a PRO domain over a pre-specified timeframe.				• The advantage of this endpoint is that it allows the assessment of PROs over an entire time frame, making full use of the information that was collected. • The concern for this endpoint lies on the interpretation since averages over time wash out changes at specific time points. It then puts into question how the resulting estimate can be understandable and interpretable by patients. • This has been a suggested approach if the planned assessments differ between trial participants (e.g., stopped assessments due to death and progression). However, there are concerns that this may not be an appropriate approach to handle these "missing data" since they are usually missing not at random. An added concern is if it is used in single-arm trials, the findings become even more difficult to interpret in the absence of a control group.
Area under the curve over a specified timeframe *Definition:* The area under the curve value of all available scores from a PRO domain over a pre-specified timeframe.	21	14	60	• This endpoint requires the timeframe to be specified to be meaningful. • There are concerns over interpretation of an area under the curve result (i.e., different patterns of observed assessment that reflects different clinical realities may lead to similar AUC). • This endpoint has strong missing data assumptions. • An added concern is if it is used in single-arm trials, it becomes even more difficult to interpret in the absence of a control group.
Best score over a specified time frame *Definition:* The best score of all available scores from a PRO domain over a pre-specified timeframe	12	17	41	• Best / worst score over a time frame is prone to bias (and measurement error). For example, one assessment point can have a high score and rest is low, and the interpretation of the results will be biased towards showing benefit (high score). • Timing of best score will also differ for each patient. • One assessment of a best score and not knowing the duration of that (e.g., if temporary or sustained) is not useful in terms of clinical relevance. • There is a need to determine whether the best score is clinically meaningful. • There is a need to find a context where this endpoint will be relevant.
Worst score over a specified time frame *Definition:* The worst score of all available scores from a PRO domain over a pre-specified timeframe	13	18	42	• Best / worst score over a time frame is prone to bias (and measurement error). For example, one assessment point can have a low score and rest is high, and the interpretation of the results will be biased towards showing risks/harm (low score). • This is similar to how CTCAE / safety data is reported and makes the PRO reporting similar to the clinician reporting. • There is a need to determine whether the worst score is clinically meaningful.
Response patterns/profiles over a specified time frame *Definition:* The longitudinal pattern of all available scores	28	8	78	• It will be difficult to develop a pre-defined hypothesis for comparability and efficacy for this PRO endpoint. • This is more useful for descriptive rather than confirmatory objective (superiority / non-inferiority).

from a PRO domain over a pre-specified timeframe				

Note: Definitions for PRO objectives are based on the SISAQOL recommendations[15]

For each PRO objective, organization representatives were asked whether they or their organization would consider using the specified objective in cancer randomized controlled trials (RCTs) or single-arm studies. Organizations could respond "yes", "no" or "don't know" to each item. If organizations replied no to a PRO objective, they were encouraged to state their reasons. An open-ended question was also included to capture additional PRO objectives. A PRO research objective was identified as high priority if at least two-thirds of the organization representatives responded "yes". An objective was identified as low priority if less than half of the representatives responded "yes" on the PRO objective. A statement was "for discussion" if it did not meet the high or low priority criteria. Organization representatives who responded "don't know" for a specific objective were not included in the total number of responses. Initial views from different SISAQOL-IMI members are summarized by qualitative comments from the survey, which can guide further discussions for each objective or endpoint.

The cells with white background show the PRO objectives that reached the high priority threshold (>2/3). Yes (%) is calculated as the number of "yes" votes divided by the total number of "yes" and "no" votes (don't know or missing is excluded).

Reaching 100% agreement among the Consortium members was not always possible and individual organizations might have different views on specific recommendations given their institutional or stakeholder standpoint. Therefore, we encourage readers to read this table together with Tables 4 and 5 to consider the positions of relevant stakeholder groups.

Table 4: Results from Priority Setting of PRO Objectives/Endpoints for RCTs by stakeholder group

"Would you or your organisation consider using…"	**Academic (N=17)**	**Industry (N=5)**	**Non-profit/ Cancer org. (N=8)**	**SME/ CRO (N=4)**	**Regulatory (N=4)**	**HTA (N=2)**	**Patient repr. (N=1)**	**Total (N=41)**
	#agreed/n (%)	#agreed/n (%)	#agreed/n (%)	#agreed/n (%)	#agreed/n (%)	#agreed/n (%)	#agreed/n (%)	#agreed/n (%)
Clinical benefit/treatment efficacy objective (confirmatory: superiority)	17/17 (100)	5/5 (100)	7/8 (88)	4/4 (100)	4/4 (100)	2/2 (100)	1/1 (100)	40/41 (98)
Clinical benefit/treatment efficacy objective (confirmatory: equivalence/non-inferiority)	17/17 (100)	5/5 (100)	7/8 (88)	4/4 (100)	2/3 (67)	2/2 (100)	1/1 (100)	38/40 (95)
Descriptive objective (exploratory/descriptive)	14/16 (88)	5/5 (100)	7/8 (88)	2/3 (67)	2/4 (50	1/2 (50)	1/1 (100)	32/39 (82)
Time to improvement	15/17 (88)	5/5 (100)	6/7 (86)	3/4 (75)	1/2 (50)	1/1 (100)	1/1 (100)	32/37 (86)
Time to worsening	16/17 (94)	5/5 (100)	8/8 (100)	4/4 (100)	2/3 (67)	1/2 (50)	1/1 (100)	37/40 (93)
Time to stable state	11/14 (79)	3/4 (75)	3/8 (38)	2/2 (100	0/2 (0)	1/1 (100)	1/1 (100)	21/32 (66)
Time to end of stable state	9/13 (69)	2/4 (50)	2/6 (33)	0/1 (0)	0/1 (0)	0/0[1] (-)	1/1 (100)	14/26 (54)
Magnitude of change at specific time point(s)	15/16 (94)	5/5 (100)	7/8 (88)	4/4 (100)	4/4 (100)	1/2 (50)	1/1 (100)	37/40 (93)
Responder with improvement at specific time point(s)	14/16 (88)	5/5 (100)	6/6 (100)	3/4 (75)	4/4 (100)	1/2 (50)	1/1 (100)	34/38 (89)
Responder with worsening at specific time point(s)	15/16 (94)	3/4 (75)	6/7 (86)	3/4 (75)	2/3 (67)	1/2 (50)	1/1 (100)	31/37 (84)
Responder with stable state at specific time point(s)	9/13 (69)	3/3 (100)	3/8 (38)	1/3 (33)	1/1 (100)	1/2 (50)	1/1 (100)	19/31 (61)

Overall average or median over a specified timeframe	11/13 (85)	5/5 (100)	4/7 (57)	3/3 (100)	2/3 (67)	1/2 (50)	1/1 (100)	27/34 (79)
Area under the curve over a specified timeframe	16/16 (100)	5/5 (100)	7/7 (100)	1/3 (33)	2/3 (67)	1/1 (100)	0/0[2] (-)	32/35 (91)
Best score over a specified time frame	8/13 (62)	3/5 (60)	0/6 (0)	0/1 (0)	0/1 (0)	0/2 (0)	1/1 (100)	12/29 (41)
Worst score over a specified time frame	9/14 (64)	2/5 (40)	1/7 (14)	1/2 (50)	1/2 (50)	0/2 (0)	1/1 (100)	15/33 (45)
Response patterns or profiles over a specified time frame	13/14 (93)	4/5 (80)	5/6 (83)	3/3 (100)	3/3 (100)	2/2 (100)	1/1 (100)	31/34 (91)

[1] Both HTAs responded don't know to this item

[2] The patient representatives responded don't know to this item

When calculating the proportion of agreement, "don't know", "not applicable" and "missing" were omitted from the denominator (n).

The cells with white background show the PRO objectives that reached the high priority threshold (>2/3 majority).

The patients are represented by WECAN which is an umbrella organisation representing 23 Pan-European cancer patient organisations, and patient input is coordinated by the Myeloma Patients Europe.

Table 5: Results from Priority Setting of PRO Objectives/Endpoints for Single-arm studies by stakeholder group

"Would you or your organisation consider using…"	**Academic (N=17)**	**Industry (N=5)**	**Non-profit/ Cancer org. (N=8)**	**SME/CRO (N=4)**	**Regulatory (N=4)**	**HTA (N=2)**	**Patient repr. (N=1)**	**Total (N=41)**
	#agreed/n (%)	#agreed/n (%)	#agreed/n (%)	#agreed/n (%)	#agreed/n (%)	#agreed/n (%)	#agreed/n (%)	#agreed/n (%)
Clinical benefit/treatment efficacy objective (confirmatory: superiority)	13/17 (76)	4/4 (100)	2/5 (40)	3/4 (75)	0/3 (0)	0/2 (0)	1/1 (100)	23/36 (64)
Clinical benefit/treatment efficacy objective (confirmatory: equivalence/non-inferiority)	11/16 (69)	3/5 (60)	1/7 (14)	3/4 (75)	0/3 (0)	0/2 (0)	1/1 (100)	19/38 (50)
Descriptive objective (exploratory/descriptive)	17/17 (100)	5/5 (100)	6/8 (75)	4/4 (100)	3/4 (75)	0/2 (0)	1/1 (100)	36/41 (88)
Time to improvement	15/17 (88)	4/5 (80)	6/8 (75)	4/4 (100)	0/2 (0)	0/2 (0)	1/1 (100)	30/39 (77)
Time to worsening	16/17 (94)	3/4 (75)	6/7 (86)	4/4 (100)	1/3 (33)	0/2 (0)	1/1 (100)	31/38 (82)
Time to stable state	9/14 (64)	1/4 (25)	1/7 (14)	3/3 (100)	0/3 (0)	0/2 (0)	1/1 (100)	15/34 (44)
Time to end of stable state	11/14 (79)	1/5 (20)	1/7 (14)	1/2 (50)	0/2 (0)	0/2 (0)	1/1 (100)	15/33 (45)
Magnitude of change at specific time point(s)	16/16 (100)	5/5 (100)	6/8 (75)	4/4 (100)	1/2 (50)	0/2 (0)	1/1 (100)	33/38 (87)
Responder with improvement at specific time point(s)	14/16 (88)	5/5 (100)	6/8 (75)	3/4 (75)	1/2 (50)	0/2 (0)	1/1 (100)	30/38 (79)
Responder with worsening at specific time point(s)	12/14 (86)	3/4 (75)	6/8 (75)	3/4 (75)	1/3 (33)	0/2 (0)	1/1 (100)	26/36 (72)
Responder with stable state at specific time point(s)	10/14 (71)	3/5 (60)	2/7 (29)	2/3 (67)	1/4 (25)	0/2 (0)	1/1 (100)	19/36 (53)
Overall average or median over a specified timeframe	10/12 (83)	4/5 (80)	3/8 (38)	4/4 (100)	0/1 (0)	0/2 (0)	1/1 (100)	22/33 (67)

Area under the curve over a specified timeframe	12/15 (80)	4/5 (80)	4/7 (57)	0/3 (0)	0/2 (0)	0/2 (0)	1/1 (100)	21/35 (60)
Best score over a specified time frame	8/13 (62)	2/4 (50)	1/6 (17)	0/2 (0)	0/1 (0)	0/2 (0)	1/1 (100)	12/29 (41)
Worst score over a specified time frame	7/13 (54)	2/4 (50)	1/7 (14)	1/2 (50)	1/2 (50)	0/2 (0)	1/1 (100)	13/31 (42)
Response patterns or profiles over a specified time frame	13/15 (87)	5/5 (100)	4/7 (57)	3/3 (100)	2/3 (67)	0/2 (0)	1/1 (100)	28/36 (78)

When calculating the proportion of agreement, “don’t know”, “not applicable” and “missing" were omitted from the denominator (n).

The cells with white background show the PRO objectives that reached the high priority threshold (>2/3 majority).

The patients are represented by WECAN which is an umbrella organisation representing 23 Pan-European cancer patient organisations, and patient input is coordinated by the Myeloma Patients Europe.

Figure 1: Development from SISAQOL to SISAQOL-IMI, and workflow towards developing the final SISAQOL-IMI recommendations
The recommendations are developed through 5 consensus meetings. The consensus reports after each consensus meeting document the process by which the consensus is reached and present the final version of the recommendation statements. The SISAQOL, Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data, refers to the former Consortium that published the first set of recommendation statements. The SISAQOL-IMI, Setting International Standards in Analysing Patient-Reported Outcomes and Quality of Life Endpoints in Cancer Clinical Trials - Innovative Medicines Initiative, refers the current Consortium that is working towards extending the work of SISAQOL.

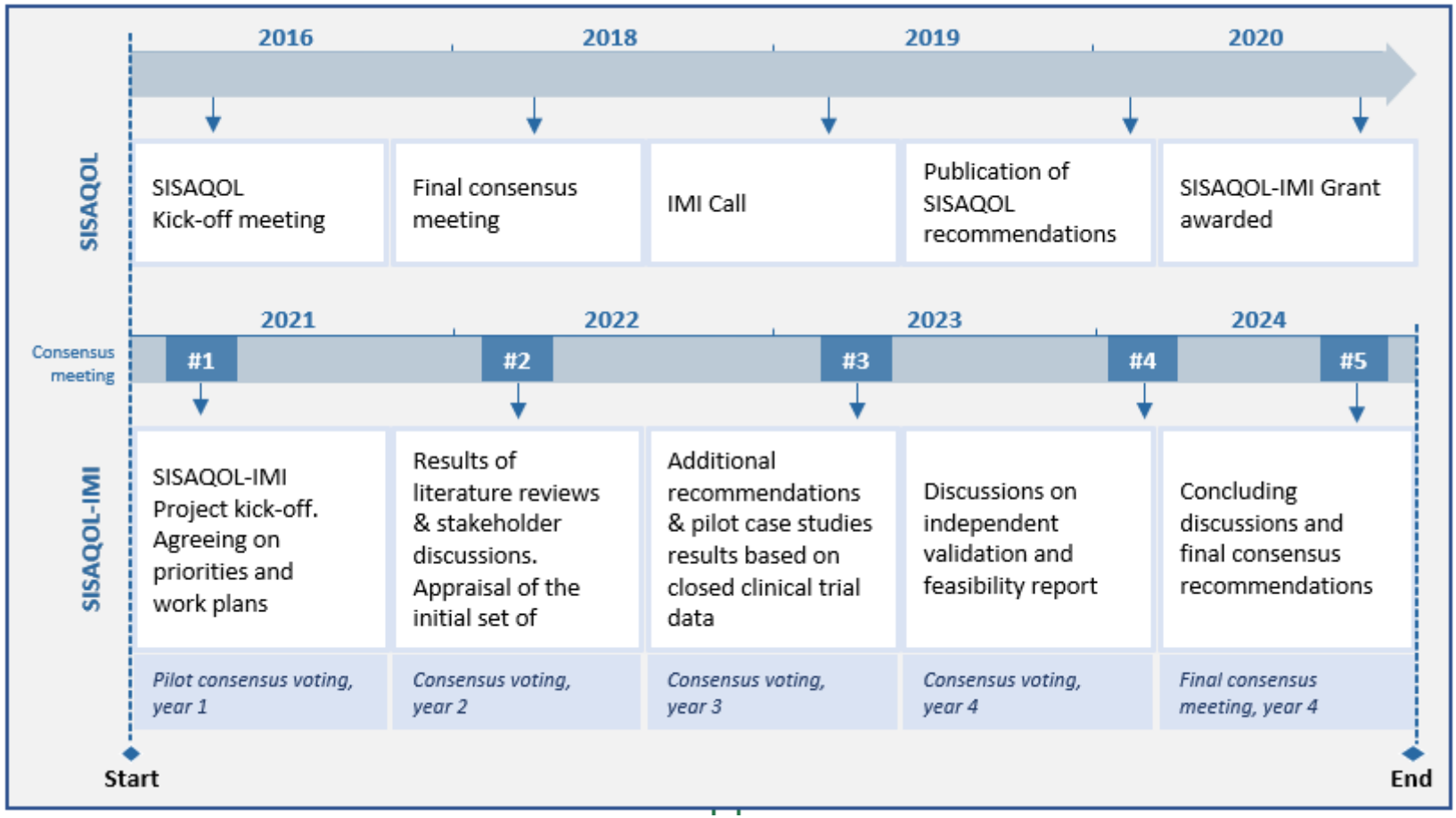

Figure 2: Project structure and interaction between the WPs
SISAQOL-IMI is organised into five scientific work packages (WP 2, 3, 4, 5, 6), and 3 cross-cutting work packages (WP 1, 7, 8).