15 October 2018

Submission of comments on FDA Standard Core Clinical Outcome Assessments and Endpoints Notice Number: NOT-FD-18-014

ISOQOL welcomes the FDA's desire to promote development of publicly available standard core sets of Clinical Outcome Assessment (COA) measures for specific disease indications. In response to FDA's request for information (RFI), ISOQOL polled our membership; a summary of responses is provided below for FDA's consideration.

General Comments:

A number of similar efforts are underway to establish minimum core sets of outcome measures for selected therapeutic areas and for various purposes (e.g., ICHOM, COMET, OMERACT, etc.). Ideally, a consortium would be established with all stakeholders represented, to align these efforts, avoid duplication of effort, increase efficiency and reduce patient burden. We would encourage FDA to identify and review existing efforts for examples of lessons learned and best practices. COMET has a searchable database of ongoing efforts, as well as guidelines, to develop core outcome sets.

Additionally, many of the questions in the RFI reference modifying existing COA tools for novel target populations. In all cases, it is important to consider input from the various key stakeholders to identify target concepts/endpoints and then carefully evaluate existing COAs in terms of how well their content matches the target concepts/endpoints, as well as the strength of the evidence supporting their content validity and measurement properties, and in what contexts. While we encourage the use/modification of existing measures, including item banks, not all existing measures have such supportive evidence, and in some cases, a *de novo* approach may be optimal to ensure treatment benefit can be captured in the target population.

RFI Question/ISOQOL Response

1. If currently available measures do not fit the topics, concepts, or wording patients identified for a given disease area based on recent patient input—for example the words or time frame used to describe fatigue or pain may be different than wording in available measures—how would you efficiently address this?

   When possible, consistent measurement of concepts across disease states may ultimately support efficient COA development as well as facilitate better understanding of treatment benefit. However, if modifications to existing measures are deemed necessary based on patient input, wording would be revised and tested via cognitive interviewing with a small sample of patients representing the target population. Ideally, changes would be minimized, and the instrument developer would approve the modifications. The extent of modification would then dictate the need for additional quantitative validation of the COA. If extensive modifications are deemed necessary, a decision would need to be made as to whether modification is the best approach versus developing a *de novo* measure.

2. What if the already established concept is more general and does not specifically tie to a tool or specific endpoint targeting the impact that a new drug would potentially have on this concept—how would you address this?

555 E. Wells Street #1100
Milwaukee, WI 53202
+1-414-918-9797

This scenario suggests that no existing tool is available that would capture the general target concept, in which case a new instrument would be needed. However, there may be individual items that could be pulled from item banks or existing measures that, together, could capture the more general concept (with developer approval as necessary). In any case, a measure would need to be adapted/developed to match the more general concept and potential impacts of the condition and its treatment. Cognitive interviewing and further assessment of measurement properties would be warranted.

3. What if a measure or tool is available for an identified concept, but it is more general and not specifically targeting the impact that a new drug would potentially have on this concept—how would you address this?

   Similar to above, this scenario suggests that no existing tool is available that would be specific enough to capture treatment benefit, in which case a new instrument would be needed. However, there may be specific individual items that could be pulled from item banks or existing measures to create the new instrument (with developer approval as necessary) or added as a supplementary assessment to the more generic instrument (modular approach). In any case, a measure would need to be adapted/developed to match the more specific concept. Cognitive interviewing and further assessment of measurement properties would be warranted.

4. What if a tool is available but has not been tested in the disease population under study—how would you address this?

   If the measure captures the target concept and has been successfully used in a number of diseases, then it would be reasonable to believe that it would work in other conditions as well. However, to ensure acceptance by various stakeholders, the measures, tools and endpoints would need to have sufficient evidence of content validity and other measurement properties for the anticipated patient population, and confirmation of the consistency of psychometric properties in the new target population would be recommended. Co-design of the study with patient partners may assist in the selection of appropriate measures which have face validity and are acceptable to the study population. It is also worth considering that there may be circumstances where the measure doesn't need to be vetted with every new population. Perhaps criteria could be established to identify when this may be the case, thereby removing the need to perform additional validation prior to use.

5. For a given disease area, what approaches can be taken to engage multiple authorities (e.g. international regulatory bodies, HTAs) and other decision makers (e.g., professional societies, research organizations, clinicians, regulated medical product industry) to gain needed input and ultimate acceptance of the same standard core set—while minimizes overall reporting burden for patients living with that disease?

   A collaboration could be created with each disease state that would be coordinated by an independent research body similar to the C-Path PRO Consortium. Once within a collaborative model, there will need to be a prioritization process and development of working groups within disease areas. Additional collaboration with ongoing efforts such as OMERACT and ICHOM will be important to avoid duplicate effort and conflicting recommendations for core COA recommendations.

6. What if publicly available measures and tools are not as good of a fit compared to a proprietary measure set that is licensed to those able to pay a more-than-nominal fee and accept the terms specified by the license holder—how would you approach this?

   The publicly available tool would need to be evaluated to determine if it is "good enough" for its intended use. Alternatively, perhaps the owners of the proprietary tool would be willing to negotiate an appropriate fee structure (nominal or free for certain uses) and acceptable terms, though the main barrier would be public availability rather than cost, given that COA costs make up a very small portion of the

overall trial budget. Otherwise, new public, non-proprietary measures would need to be developed and validated.

7. What would you do to determine the applicability of a measure or measure set and scoring developed in one region of the world to patient (likely clinical trial) populations in other regions (e.g., across all regions now represented by ICH regulatory members and observers)? What types of issues might arise that suggest limited applicability or operability? What would you do to address such issues that indicate limited applicability or operability?

   Rigorously developed COA measures should have included a robust linguistic validation/cultural adaptation process to ensure that the measure is applicable across cultures and languages. As part of an evaluation of existing COAs, evaluation of the cultural adaptation process used in development should be a key attribute. We would propose to begin to develop and support repositories of psychometric properties of linguistically validated translations. Evidence of content and quantitative adequacy across multiple languages and cultures will support wider use without the need for additional psychometric analyses in each new language/culture. If evidence of cultural relevance/acceptability is not available, a translatability assessment could be performed, involving a review by multiple linguists to evaluate the instrument in terms of cultural relevance and appropriateness. In terms of issues, aspects that are related to the "environment", for example differences in the standard of care or health care system, may impact the use of an instrument across global regions. In addition, treatment effect could vary by language, and translation issues would be one factor to explore.

8. For most diseases, the patients who seek treatment may range in age. They may also range in level of literacy, levels of cognition, disease severity, and other dimensions that may affect their ability to report or otherwise reliably participate in data collection. The measures, tools, and endpoints that are developed and used need to recognize and address these important variations in the patient population. For a given context of disease and patient subpopulation:

8a. How would you identify the limitations of existing measures and tools?

   This must be a part of the qualitative and quantitative assessment of measures. Discussions with clinical experts and patient representatives will help identify sub-populations that would need to be considered (different age groups, severity levels, disease types, cognitive ability, etc.) when determining the adequacy of a measure. These sub-populations would then need to be considered and included in the evaluation of the measures. For example, patients' ability to self-complete could be examined qualitatively, and ceiling & floor effects could be examined quantitatively.

   Not all of these problems can be overcome, thus using multiple assessments (e.g. PRO, ClinRO, ObsRO, PerfO) may be required.

8b. What sort of limitations would you anticipate may be identified? Please offer some examples based on your previous experience.

   The instrument not being adequate for all severity levels – floor and ceiling effects.

   The need for different assessments for different sub-populations and then how to address that in the analysis – PRO for older children, different PRO for adults, ObsRO for younger children.

   Different countries conceptualize and report outcomes differently e.g., the construct of 'blocks walked' in function. Terminology varies, e.g., UK use of rucksack. How fatigue is described and quantified varies across countries.

8c. For the various types of issues identified, what would you do to address these limitations? What has been your experience with the effectiveness of these strategies in the past?

Revise items or response scales to capture broader range of impacts.

For children, consider ObsRO for all ages and then compare to PRO. If the results are close enough then consider ObsRO across the age spectrum - with the full understanding that this has its limitations. Alternatively, attempt to develop a measure that is applicable for wide age range – e.g., age 8 through adults, using commonly understand language, assuming target concepts are the same across ages.

Perform translatability assessment and linguistic validation in the target population.

8d. Considering experiences with studies with pediatric patients—for a given disease or condition and patient age group-- what has been your approach to reporting: who does the reporting for the patient (and why), how is the reporting collected? When?

Patient where possible but observer assessment where patient is unable to self-report. Generally, this means the use of PRO for children aged 8 and older, and ObsRO for children under 11 (thus some overlap for the 8-10 age group), but this depends on various factors, including the nature of the concept being measured, the recall period, etc. The perspectives of caregivers and children may or may not align, however both are important to take into consideration.

"When" depends on the study design but it is important to be consistent - i.e. always in clinic or always at home, as there can be systematic differences between the two.

9. How much time does it take (elapsed time of the project from start to finish) and how much does it cost (estimates of both the US dollar amount and level of effort would be helpful) to test and modify an existing measure set or tool?

6 to 12 month process and a minimum cost of $100,000, but if extensive modifications needed, costs and timeline could be comparable to *de novo* development

Factors influencing time and budget include: target population, recruitment method, mode of administration, extent of modifications.

10. How much time does it take (elapsed time of the project from start to finish) and how much does it cost (estimates of both the US dollar amount and level of effort would be helpful) to do “from the ground-up” development and testing of a new measure set or tool?

12 months to up to 3 years (but if through consortium, could be longer) and costs ranging from $500,000 to $800,000 to develop a measure. This is for single language. Translation/linguistic validation and eCOA costs would also need to be considered.

Factors influencing time and budget include: target population, recruitment method, mode of administration, scope of measure.

The effort generally includes COA methodologists, clinical experts, patient representatives and collaboration with a specialized COA research vendor with expertise in instrument development, including both qualitative and quantitative methods.

11. What are the expertise and operation management skill sets needed for the work described in this RFI?

    For collaborative projects, there is a need for strong project management, facilitation skills, and legal expertise (contracting). Scientific expertise includes consensus methods, study design for both qualitative research and psychometric evaluation, interviewing, qualitative and quantitative data analysis, item generation, linguistic/cultural validation, and eCOA. Additionally, experience in recruiting, data management, GCP, and IRB submission/review process is important. Skills around setting up library/archives online and managing over time, including access/rights are also useful.

12. What should the maintenance lifecycle for measures, tools, and endpoints be? Describe example recommended maintenance plans.

    Varies a lot by disease, changes in treatment, concept, etc. For newly identified diseases, for example, our understanding of them can change quickly and quite significantly. Measures for diseases like depression need to be updated as our understanding of the disease changes or as the target of treatment changes. For example, as we have relied on measures developed during the era of tricyclic antidepressants we have struggled to show treatment effect, but modern treatments don't address the same issues in depression.

    Maintenance is critically important, and standards have not been established. This is an area for further development, but certainly this should be done at regular intervals, perhaps every 5 years, as well as when new information is uncovered, or treatment targets change.

13. What organizations are currently using clinical outcome assessments for decision making? What information do they use to inform their decision making, and how is it used?

    Organizations involved in drug development certainly are with the choice of treatments that advance relying on positive outcomes: drug developers, regulatory bodies, HTAs/payers (NICE UK, PBSC MSAC Australia, etc., though to varying extents in different regions).

    Ideally, we would like to see increased use of COA data to inform shared-decision making between patients and their clinical team. Clinical practices/hospitals, insurers increasingly use these data.

14. What steps can reduce patient and other informants' burden?

    Ensuring that the measurement approach is targeted to capturing key concepts and that instrument content is highly relevant, clear, easy to complete and not redundant. Burden is less about the number of items and more about relevancy, ease of completion and lack of redundancy.

    Carefully consider the timing of assessments and collect data at important time points and on a regular schedule. Ensure that COA is considered during study design with input from patients. Follow the latest international guidelines for the inclusion of patient-reported outcomes in clinical trial protocols: The SPIRIT-PRO Extension.

    Allow flexibility in data capture – BYOD ("bring your own device"), using "just in time" reporting processes, leveraging emerging technology such as voice activated assistants and various wearable sensors.

This comment was reviewed and endorsed by the International Society for Quality of Life Research (ISOQOL) Board of Directors and does not reflect an endorsement of the ISOQOL membership.