PCR39

Reporting and Interpretation of Cognitive Debriefing Results in Patient Reported Outcome Measure Validation: Are There Standards?

Authors: Nathan Johnson[1]; Martha Gauthier[1]; Julie Whyte[1]

Affiliations: 1. Patient-Centered Outcomes, Lumanity Inc., Boston, MA, US

INTRODUCTION

- Cognitive debriefing (CD) studies are conducted to assess if patients interpret patient-reported outcome (PRO) measures as intended
- While regulatory agencies emphasize the importance of providing evidence that respondents understand contents of a PRO measure (e.g. instructions, items, response options) as intended[1], there are no clear recommendations for interpretation cutoffs when evaluating the CD results
- Determining whether revisions are warranted can be especially challenging when results are ambiguous; when interpretation challenges are reported by a minority of participants, the implications are less straightforward and require more nuanced judgment
- Pragmatic thresholds could help researchers determine whether changes should be made to the item alongside qualitative judgment.

OBJECTIVES

- This study aimed to examine how CD results have been categorized in published research and existing regulatory and organizational guidances and to determine whether standard quantitative thresholds or cutoffs are used in their interpretation

METHODS

Conference abstracts

- A targeted review of ISPOR and ISOQOL conference abstract databases was conducted using the terms “cognitive debriefing” and “content validation”
  - The ISPOR database included conference abstracts from 2010-2025, while the ISOQOL database included abstracts from 2025
- Each abstract was screened to identify those discussing providing CD results; abstracts describing CD studies of PRO measures were included for full review
- For each eligible abstract, the following was extracted and assessed:
  - Whether CD results were reported quantitatively
  - Whether a predefined threshold/cutoff was specified for interpreting participant understanding
- Any reported thresholds/cutoffs for participant understanding were recorded

Regulatory and organizational guidances

- Key regulatory guidances (EMA, FDA, NMPA, PMDA) and qualitative research methodological societies (COSMIN, ISOQOL, ISPOR) and consortia were reviewed for specific recommendations on interpretation cutoffs.

RESULTS

Conference abstracts

- The search yielded 126 ISPOR and 11 ISOQOL abstracts for review
  - 40 reported CD results, while 15 abstracts (10 ISPOR, 5 ISOQOL) met the eligibility criteria (**Figure 1**)
  - Few abstracts reported CD results quantitatively (e.g. the proportion of participants interpreting the items as intended). Among those that did, explicit thresholds were rare (**Table 1**)
  - When thresholds were mentioned, they varied considerably (e.g. ≥ 75% item comprehension) and were typically applied descriptively rather than as a formal decision rule
  - No consistent or widely endorsed standard for acceptable comprehension emerged across studies

Figure 1: Search results

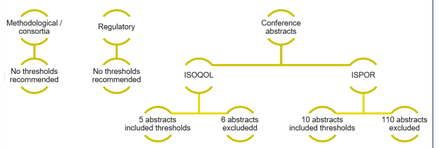


Table 1: Cognitive debriefing thresholds specified

Attribute	Result
Threshold min – max	60 – 92%
Threshold mean	74%
Threshold median	75%
Comments on thresholds	• 80-100% considered “high level of understanding” for multiple studies • Most studies did not specify what occurred if item interpretability did not meet threshold. Only 2 studies mentioned that items were removed based on not meeting thresholds • Thresholds varied for specific aspects of interpretability (item interpretation, relevance, response options, instructions)

Regulatory and organizational guidances

- Review of regulatory guidances and methodological frameworks found no published quantitative thresholds for interpreting cognitive debriefing results (**Table 2**)
- While the FDA PFDD Guidance Series 3[1] emphasizes documenting participant understanding and the ISPOR Task Force[2] recommends iterative revision until issues are resolved, neither provides specific cutoffs for acceptable comprehension rates
- The ISPOR PRO Good Research Practices Task Force[2] and ISOQOL PRO reporting standards[3] emphasize documentation, iteration, and qualitative judgment in cognitive debriefing but do not specify quantitative thresholds for acceptable comprehension. This creates ambiguity in how results should be reported and interpreted, particularly for regulatory submissions
- COSMIN provides methodological quality criteria for cognitive interview studies, including recommendations for sample size[4]
  - Like the ISPOR task force, COSMIN does not specify quantitative thresholds for acceptable item comprehension, leaving interpretation to researcher judgment.

Table 2: Regulatory and organizational guidances

Organization	Guidance
Organizational guidances	
COSMIN	Sample size benchmark of >7 is recommended for a “good' rating
ISOQOL	Interpretation is context-driven, supported by holistic evidence
ISPOR	• Emphasizes iterative revision until comprehension issues resolved • Recognition of nuances in interpreting results and challenges when issues are presented for a minority of respondents
Regulatory agencies	
EMA	No guidelines specific to methodology of development or validation of instruments
FDA	Emphasis on documenting respondent understanding of each item, response options, an instructions with reference to the intended interpretation
NMPA	Discusses PRO development process, but no specific quantitative threshold published
PMDA	• References PFDD • Notes that PRO measures should be developed and validated according to guidelines, but does not specify

CONCLUSIONS

- This review found limited transparency and consistency in the reporting and interpretation of CD results in PRO measure content validation. While some studies apply quantitative thresholds, these are neither common nor standardized, and often lack explicit justification.
- The limited quantitative reporting may reflect differing expectations between publication and regulatory audiences. Regardless, researchers should ensure data are collected and analyzed with sufficient granularity to provide the transparency required for regulatory submissions, even when such detail is not included in publications.
- Variability in reporting practices may reflect the inherently interpretive nature of CD data, where conclusions must be informed by study context, the target population, and item complexity.
- Developing standardized reporting guidance could strengthen comparability and methodological rigor across CD studies; however, any thresholds should be applied alongside contextual judgment, since concerns raised by even a single participant may represent meaningful feedback that warrants revision.
- In the absence of established guidance, we propose pragmatic thresholds informed by observed practices in the literature and methodological precedent from adjacent content validation standards (e.g., CVI) (**Table 3**).[5,6] An 80% threshold aligns with conventions seen in some content validity studies and reflects a practical "supermajority" standard. The 60% lower bound flags items where at least a third of participants had comprehension issues — a meaningful signal.
- With small samples (N<10), absolute counts provide a more stable signal than percentages. A single misunderstanding may reflect idiosyncratic interpretation; two or more suggests a pattern warranting review the acceptability threshold.
- Unlike item interpretation, item relevancy is expected to vary across participants based on disease severity, symptom heterogeneity, and individual experience.

Table 3: Recommended cognitive debriefing thresholds for interpretation (N≥10)

Category	Result	Interpretation
Acceptable	≥80% of participants interpret the item as intended	No modification required
Marginal	60–79%	Review qualitative feedback; minor wording adjustments may be warranted
Problematic	< 60%	Item revision strongly recommended; re-testing in subsequent CD round

KEY TAKEAWAYS

- Regulatory agencies and methodological organizations do not specify quantitative thresholds for acceptable comprehension in cognitive debriefing studies.
- Few published CD studies report results quantitatively. When thresholds are applied, they vary considerably and lack explicit justification.
- In the absence of consensus guidance, we propose: ≥80% comprehension (acceptable), 60–79% (marginal), <60% (problematic). These thresholds should be applied alongside qualitative judgment.
- With <10 participants, percentage thresholds are unreliable. Two or more participants with comprehension issues should trigger review, regardless of percentage.
- Thresholds should complement but not replace qualitative interpretation. Concerns raised by even a single participant may warrant revision depending on the nature and severity of the issue.

REFERENCES

1. US FDA. Patient-focused drug development: selecting, developing, or modifying fit-for-purpose clinical outcome assessments. Guidance for industry, Food and Drug Administration staff, and other stakeholders. FDA; 2025 Oct **2.** Patrick D, et al. Content validity—establishing and reporting the evidence in newly developed PRO instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force Report: Part 2—assessing respondent understanding. Value Health. 2011;14(8):978–88. **3.** Brundage M, et al. Patient-reported outcomes in randomized clinical trials: development of ISOQOL reporting standards. Qual Life Res. 2013;22(6):1161–75. **4.** Terwee CB, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res. 2018;27(5):1159–70. **5.** Lynn MR. Determination and quantification of content validity. Nurs Res. 1986;35(6):382–5. **6.** Polit DF, Beck CT. The content validity index: are you sure you know what's being reported? Critique and recommendations. Res Nurs Health. 2006;29(5):489–97.

Key: CD, cognitive debriefing; COSMIN, COnsensus-based Standards for the selection of health Measurement INstruments; CVI, content validity index; EMA, European Medicines Agency; FDA, Food and Drug Administration; ISPOR, International Society for Pharmacoeconomics and Outcomes Research; ISOQOL, International Society for Quality of Life Research; NMPA, National Medical Products Administration; PMDA, Pharmaceuticals and Medical Devices Agency; PRO, patient-reported outcome.

DISCLOSURES: **NJ**, **JW**, and **MG** are employees of Lumanity, Inc.

An electronic version of the poster can be viewed by scanning the QR code.

Poster presented at the 2026 ISPOR Conference; May 17-20, 2026; Philadelphia, Pennsylvania, USA.