



Evaluating Comprehensibility of 157 Patient-Reported Outcome Measures (PROMs) in the Nationwide Dutch Outcome-Based Healthcare Program: More Attention for Comprehensibility of PROMs is Needed

Attie Tuinenburg¹ · Domino Determann¹ · Elise H. Quik¹ · Esmee M. van der Willik¹ · Geeske Hofstra¹ · Joannes M. Hallegraef^{1,2} · Ingrid Vriend¹ · Lisanne Warmerdam¹ · Hester E. van Bommel³ · Gudule Boland³ · Martijn A. H. Oude Voshaar^{1,4}

Accepted: 28 July 2024 / Published online: 13 August 2024
© The Author(s) 2024

Abstract

Introduction Patient-reported outcomes measures (PROMs) are increasingly prevalent in healthcare and used for shared decision-making and healthcare quality evaluation. However, the extent to which patients with varying health literacy levels can complete PROMs is often overlooked. This may lead to biased aggregated data and patients being excluded from studies or other PROM collection initiatives. This cross-sectional study evaluates the comprehensibility of 157 well-known and widely used PROM scales using a comprehensibility checklist.

Methods Pairs of two independent raters scored 157 PROM scales designed for adults included in the 35 sets of outcome information developed as part of the Dutch Outcome-Based Healthcare Program. The PROM scales were scored on the eight comprehensibility domains of the Pharos Checklist for Questionnaires in Healthcare (PCQH). Interrater agreement of domain ratings was assessed using Intraclass Correlation Coefficients or Cohen's kappa. Subsequently, final ratings were established through discussion and used to evaluate the domain-specific comprehensibility rating for each PROM scale.

Results Comprehensibility of a large number of PROM scales ($n = 157$), which cover a wide range of diseases and conditions across Dutch medical specialist care, was assessed. While most PROM scales were written at an accessible language level, with minimal use of medical terms, instruction clarity, number of questions, and response options emerged as significant issues, affecting a substantial proportion of PROM scales. Interrater agreement was high for most domains of the PCQH.

Conclusion This study highlights the need for greater attention to the comprehensibility of PROMs to ensure their accessibility to all patients, including those with low health literacy. The PCQH can be a valuable tool in PROM development in addition to qualitative methods and in selection processes enabling comparison of comprehensibility between PROMs. However, the PCQH needs further development and validation for these purposes. Enhancing the comprehensibility of PROMs is essential for their effective incorporation in healthcare evaluation and decision-making processes.

1 Introduction

Patient-reported outcomes (PROs) are becoming increasingly important in clinical trials as well as in daily routine healthcare and are considered essential in value-based healthcare [1–3]. Various national health authorities now recommend routine collection of PROs from patients receiving medical specialist care [2, 4–6]. Studies show that patient-level PRO data can enhance patient engagement,

shared decision-making and personalized treatment [7–9], while aggregated data across care providers can be used in quality improvement through benchmarking and shared learning [10–12]. Additionally, aggregated PRO data serves to provide real-world evidence on treatment effectiveness and safety [13, 14] and ensures accountability to payers and the public [15]. Therefore, patient-reported outcome measures (PROMs) are increasingly included in core outcome data collection initiatives [16–18]. To achieve the potential benefits of using PROMs in all patients and to obtain high-quality aggregated data, it is important that every patient can participate, particularly when PROMs are embedded in the clinical workflow. However, PROM implementation studies

Attie Tuinenburg and Domino Determann Shared first authorship.

Extended author information available on the last page of the article

Key Points for Decision-Makers

As the use of patient-reported outcomes measures (PROMs) becomes more widespread in healthcare, including their integration into clinical workflows, it is essential that all patients, including those with low health literacy, can understand and complete PROMs.

This study shows that comprehensibility of PROMs can be improved. Especially by including clear instruction for patients and paying attention to the number of questions and answers.

More attention is needed for comprehensibility when developing or selecting PROMs to be implemented in initiatives supporting the decision-making process or healthcare evaluation. The Pharos Checklist for Questionnaires in Healthcare (PCQH) is a valuable tool but needs further development and validation for these purposes.

show challenges regarding the comprehensibility of PROMs (i.e., the degree to which the PROM is correctly understood by patients) [19–22].

Problems with comprehensibility are of particular concern in people with low (health) literacy, which is a worldwide problem [23–25]. For instance, in the Netherlands, one in four people have insufficient or limited health literacy skills [26, 27]. They face difficulties in obtaining, understanding, appraising, and using health information when making health-related decisions [28], for example, when filling in PROMs. A part of this population also has low basic literacy skills [29]. Health literacy is recognized as a major determinant of health and socioeconomic health disparities by the World Health Organization [30] and lower health literacy is associated with poorer health outcomes and increased mortality [25, 31].

To ensure that every patient, regardless of their health literacy skills, can participate and benefit from discussing their own PRO data during consultations and to obtain PRO data that represents the entire population for quality assessment purposes, it is crucial that everyone can understand PROMs¹ [32, 33]. Consequently, various well-defined frameworks to support PROM development [e.g., International Society for

Pharmacoeconomics and Outcomes Research (ISPOR) [34] and selection [e.g. International Consortium for Health Outcomes (ICHOM) [35], Consensus-based Standards for the selection of health Measurements Instruments (COSMIN) [36], PROM-cycle [37], and International Society for Quality of Life Research (ISOQOL) [38] emphasize the importance of comprehensibility [16, 17, 22–40]. For instance, the ISPOR PRO Good Research Practices Task Force advises to use cognitive interviews to ensure that respondents understand how to complete the PROM, the meaning of the questions, and how to use the response scales [41, 42]. In addition, according to the COSMIN guideline and the PROM-cycle, the comprehensibility of PROMs impacts content validity, which is often considered to be the most important psychometric property of a PROM [43]. Thus, it is recommended to use qualitative methods that involve patients, such as cognitive interviews, to assess the comprehensibility of PROMs [43]. However, a scoping review by Wiering et al. shows that only in 51% of developed PROMs, patients were involved in testing for comprehensibility [44]. Other studies show that, for example, the readability level of PROMs is higher than recommended [45–47]. This may result in lower completion rates or hinder accurate completion, compromising PROM validity and excluding people with low (health) literacy [32, 33, 45].

Little is known, however, about how the degree to which PROMs are easy to use, varies between PROMs as qualitative methods assessing the comprehensibility of PROMs make it difficult to compare the comprehensibility of PROMs between populations or PROMs. Moreover, in situations where a large number of PROMs need to be evaluated for comprehensibility, for example, during PROM selection, it is often impractical to conduct qualitative studies on multiple PROMs as this is time-consuming and requires specific expertise.

The aim of this study was to evaluate the comprehensibility of the PROMs included in the 35 core outcome sets that were developed as part of the Dutch Outcome-Based Healthcare Program (2018–2023), which was a national initiative to stimulate the collection of routine patient-reported and clinical outcomes in daily medical specialist care. A core outcome set, further referred to as “outcome set,” is an agreed standardized collection of important outcomes and how they should be measured in a specific area of health or health care [48]. The program was initiated as part of the national policy agenda for medical specialist care agreed upon by all relevant umbrella organizations and conducted under the auspices of the Dutch Ministry of Health, Welfare and Sport [2, 49].

¹ Besides comprehensibility of PROMs other factors, e.g. accessibility, time constraints, interest in the subject matter, perceived relevance of questions and situational contexts, can influence the patient's ability and willingness to complete a questionnaire accurately.

2 Methods

2.1 Dutch Outcome-Based Healthcare Program

In the Outcome-Based Healthcare Program, disease/condition-specific working groups consisting of mandated representatives of the umbrella organizations representing all stakeholders involved in Dutch medical specialist care (including patient organizations), have developed 33 outcome sets to support shared decision-making and healthcare quality evaluation [50]. The process was facilitated by the Dutch National Health Care Institute by providing methodological and organizational support. The program was assigned with the task of developing 36 outcome sets for diseases and conditions representing a considerable part of the Dutch national disease burden [50, 51]. Due to practical considerations, 33 sets were eventually developed (Online Resource 1). In addition, two sets of Generic PROMs (for adults and children) were developed, consisting of PROs that reflect common areas of disease impact for all patients in medical specialist care [49, 52, 53]. In this study on the comprehensibility of PROMs, outcome set exclusion criteria were: (1) sets without PROMs, (2) sets developed for children, and (3) sets that were not approved (yet) by all participating umbrella organizations.

2.2 Pharos Checklist for Questionnaires in Healthcare

Pharos, the Dutch Centre of Expertise on Health Disparities, developed a checklist to evaluate and improve the comprehensibility and accessibility of questionnaires for adults with low health literacy, the so-called “Pharos Checklist for Questionnaires in Healthcare” (PCQH) [54]. Types of healthcare questionnaires that can be assessed with the PCQH include, for example, PROMs, patient reported experience measures (PREMs), and questionnaires for scientific studies or nationwide surveys. Pharos is committed to reduce avoidable health disparities due to socio-economic conditions [55] and is frequently involved in the development and adaptation of health questionnaires. To do this, Pharos employs a standardized qualitative methodology [56] based on think aloud principles and the use of probing questions for rewording and testing with persons with low literacy levels [57, 58]. Based on the experiences gained through this work and substantiated by scientific publications [32, 59–64], the PCQH was developed.

The PCQH consists of four components: (1) comprehensibility, (2) accessibility, (3) layout, and (4) validation. In our study we focus on comprehensibility, which we define as: the ease with which the intended patient population of the PROMs can understand, interpret and accurately respond to

the questions or items presented. In the PCQH, comprehensibility comprises the following eight domains: (1) language level, (2) presence of a brief and clear instruction, (3) number of questions, (4) number of answer options, and the use of: (5) questions in active voice, (6) medical terms/abbreviations, (7) concrete questions, and (8) statements. Each domain is rated using a color-coded ordinal or nominal scale where “green” denotes “optimal” adherence to the domain criteria, “orange” denotes “acceptable” adherence and “red” denotes “significant inadequacies” that require substantial revisions for enhanced clarity and accessibility.

Since the PCQH was originally intended to support the development of new or adapting existing complete questionnaires in healthcare (i.e., evaluation of complete PROMs instead of rating multiple PROM scales), several changes were made to the criteria used for rating some of the eight domains to better align with the goals of this study. A detailed overview of the domain-specific definitions and ratings, as well as the changes that were made to the PCQH, are presented in Table 1. We evaluated comprehensibility at the scale level of PROMs, meaning that for multidimensional PROMs the PCQH was applied to each individual scale. For example, the Michigan Hand Outcomes Questionnaire consists of six scales that measure multiple PROs [65]; only the scale activities of daily living was included in the outcome set for “hand and thumb base osteoarthritis” and assessed for comprehensibility in this study. Therefore, the number of questions allowed for a green or orange rating was reduced relative to the original version of the PCQH, which is intended to be applied to a complete questionnaire. Threshold values for acceptable length of individual scales were derived from the balance between minimizing patient burden and maintaining scale reliability. In this regard, we established that a maximum of six questions would receive a green rating, as this represents the minimum number of questions that still achieves acceptable internal consistency reliability (Cronbach’s $\alpha = 0.72$) at the group level assuming relatively low inter-item correlations of 0.3 [66]. Furthermore, for a green rating in the domain “questions in active voice,” an absolute number of questions with the unfavorable property is tolerated in the original version of the PCQH. Applying this threshold disproportionately penalizes longer questionnaires. To ensure a fair and balanced evaluation, we adjusted the criteria so that ratings in this domain consider the proportion of questions in active voice rather than their absolute number. To be able to rate more objectively, the criteria used to rate the domains of “concrete questions” and “statements” have been adapted from nominal in the original version of the PCQH to rational in this study. Again, to not penalize longer PROM scales, percentages are used instead of absolute numbers.

Table 1 Eight comprehensibility domains and their ratings

Domain	Definition	Rating
Language level	The language level, according to the Common European Framework of Reference for Languages (CEFR) [Council of Europe, 2023], of a PROM, was determined using “Klinkende Taal” [Klinkende Taal, 2023]. “Klinkende Taal” is an online tool that provides its judgment of the (Dutch) language level of a text based on a list of difficult words and a linguistic and syntactic analysis. Assessment of language level was structured, so texts were consistently entered into “Klinkende Taal.” Language level was determined successively for (1) the instruction, (2) the questions and answer categories, and (3) instruction, questions, and answer categories.	Green: A1, A2, and B1 Orange: B2 Red: C1
Presence of a brief and clear instruction	Good instruction is defined as being brief and clear.	Green: maximum of ten lines and clear instructions, which include the purpose or subject of the PROM, contain a fill-in instruction, and is addressed to the patient Orange: all other kinds of instructions Red: > 20 lines and/or absence of clear instructions, which lack either the purpose or subject of the PROM, a fill-in instruction, and/or is not addressed to the patient
Number of questions ^a	Number of questions in the PROM.	Green: ≤ 6 questions Orange: 7–10 questions Red: ≥ 11 questions
Number of answer options	Maximum number of answer options in the PROM.	Green: maximum number of answer options = 4 or NRS (0–10 or 1–10) Orange: maximum number of answer options = 5 Red: maximum number of answer options > 5 or PROM contains an open question
Questions in active voice ^b	Percentage of questions in active voice. In an active sentence, the subject is also the doer or acting person. In a passive sentence, the subject undergoes something instead of performing the action.	Green: 100% of questions in active voice Orange: 26–99% of questions in active voice Red: ≤ 25% of questions in active voice
Medical terms/abbreviations	Commonly used medical terms (e.g., diabetes) and abbreviations (e.g., CT scan, MRI) are rated green.	Green: no medicals term or abbreviations Orange: medical terms and abbreviations are explained Red: medical terms or abbreviations without explanation
Concrete questions ^c	Percentage of concrete questions. With concrete we mean that questions are specific rather than general and are about everyday situations with clear examples and avoid hypothetical descriptions.	Green: 100% concrete questions Orange: 26–99% concrete questions Red: ≤ 25% concrete questions
Statements ^c	Percentage of questions not formulated as a statement.	Green: 100% of questions is not a statement Orange: 26–99% of questions is not a statement Red: ≤ 25% of questions is <i>not</i> a statement

CT scan computed tomography, *MRI* magnetic resonance imaging, *NRS* numerical rating scale, *PROM* patient-reported outcome measure

^aThe PCQH is intended to be applied to a complete questionnaire. In this study, PROM scales were assessed and therefore the rating of the domain number of questions has been adjusted (original rating was: green ≤ 20 questions, orange 21–25 questions, red ≥ 26 questions)

^bThe criteria used to rate this domain have been adjusted compared with the original version of the PCQH (where absolute numbers were used) to not penalize longer PROMs (original rating was: green, absence of auxiliary verbs; orange, ≤ 3 questions containing auxiliary verbs; red, ≥ 4 questions containing auxiliary verbs)

^cTo be able to rate more objectively and to provide a more detailed insight into the characteristics of different PROMs, the criteria used to rate these domains have been adapted from nominal in the original version of the PCQH, to rational in this study. In order to not penalize longer PROMs, percentages are used instead of absolute numbers (original rating of concrete questions domain was: green, questions on everyday situations with a clear example; orange, hypothetical questions; red, general questions. Original rating of statements domain was: green, no statements; orange, statements with the answer options ja/no or for/con; red, statements with response scales)

2.3 Analyses

All eight raters were methodologists with expertise in PROMs, employed by the Outcome-Based Healthcare

Program and were trained by Pharos (H.v.B. and G.B.) to apply the PCQH. For calibration, the comprehensibility of the eight scales of the SF-36 [67] was assessed by the

Table 2 Agreement between assessors

Domain	Absolute agreement	95% CI	Statistic	Value ^a	95% CI
Language level	99%	98–100%	Kappa	0.92	0.82–1.00
Presence of a brief and clear instruction	77%	70–84%	Kappa	0.61	0.50–0.80
Number of questions	96%	94–99%	ICC	1.00	1.00–1.00
Number of answer options	97%	94–100%	Kappa	0.94	0.90–1.00
Questions in active voice	79%	72–85%	ICC	0.95	0.93–0.96
Medical terms/abbreviations	94%	91–98%	Kappa	−0.02	−0.03–0.00
Concrete questions	72%	66–80%	ICC	0.98	0.98–0.99
Statements	83%	77–89%	ICC	0.91	0.88–0.93

CI confidence interval, ICC intraclass correlation coefficient

^aInterpretation of the ICC: an ICC > 0.90 indicates good agreement; a kappa value of 0.81–1 is considered “almost perfect agreement,” a value of 0.61–0.80 is considered “substantial agreement;” lower values indicate insufficient agreement between assessments [69, 70]

methodologists. Differences and similarities were discussed to reach a consistent interpretation of each domain.

The methodologists were divided into fixed pairs and PROM scales were assigned to these pairs. When a PROM consisted of multiple scales, all included scales of the PROM were assigned to the same pair. Subsequently, each PROM scale was independently assessed by both members of a pair. An online tool was used to assess the (Dutch) language level of a text (Klinkende Taal) based on a computer algorithm consisting of a preexisting list of difficult words and a linguistic and syntactic analysis [68]. Differences in assessment were discussed in a consensus meeting within each pair and if no conclusion was reached a third assessor was consulted. As a result, each pair provided one final assessment per PROM scale on the eight comprehensibility domains of the PCQH. All descriptive results regarding the comprehensibility of the PROM scales in this study are based on these final ratings.

The interrater agreement between the assessment of comprehensibility by the two methodologists, prior to the consensus meeting, was evaluated for each domain by examining the absolute percentage of agreement and its corresponding 95% confidence interval. For the four domains measured on a ratio scale (number of questions, active voice in questions, concrete questions, statements), the intraclass correlation coefficient (ICC) was determined. This was based on a two-way analysis of variance (ANOVA) model for single measurements and absolute agreement (ICC type 2,1 according to Shrout & Fleiss) [69]. For the interpretation of the ICC, it is assumed that an ICC > 0.90 indicates good agreement. For the four domains measured at an ordinal level (all other domains), the kappa statistic was calculated. The interpretation of the kappa values follows the classification of Landis and Koch, where a kappa value of 0.81–1 is considered “almost perfect agreement,” and a value of 0.61–0.80 is considered “substantial agreement” [70]. Lower values indicate insufficient agreement between assessments.

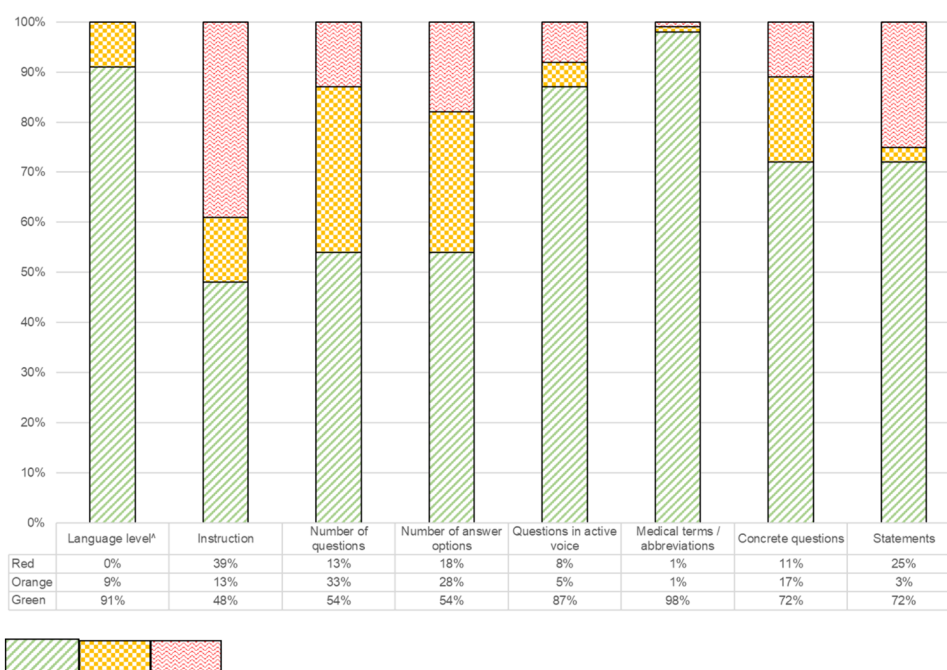
3 Results

3.1 Outcome Sets and PROMs Selected for Analyses

A total of 6 of the 35 outcome sets (33 disease/condition-specific sets and 2 Generic PROM sets) developed in the Outcome-Based Healthcare Program were excluded from the study. The cataract set and macular degeneration set did not contain PROMs, and the pancreatic cancer set and renal cell carcinoma set were not approved (yet) by the boards of all participating umbrella organizations. In addition, the asthma in children set and the generic PROM set for children were excluded as the focus of this study is on PROMs for adults. The remaining 29 outcome sets included a total of 157 PROM scales. All 157 PROM scales were independently assessed as described in the “Methods” section. A third assessor was consulted for six PROM scales. Online Resource 2 contains the comprehensibility profiles of each of the 157 PROM scales.

3.2 Agreement Between Assessors

Agreement between assessors was predominantly high, with ICC/kappa in most domains compatible with good agreement according to the specified cutoff values (Table 2). Absolute agreement was lower for most ratio variables (active voice in questions, concrete questions, statements). This can be explained by a wider range of possible values compared with ordinal variables, making the chance of exact agreement between assessors smaller. The negative kappa value for the domain medical terms/abbreviations indicates lower agreement than could be expected by chance, despite high absolute agreement. Inspection of the data revealed that this was caused by uneven distribution of assessments across categories, with nearly all assessments falling into the green category. The domain of instructions is the only domain with a low degree of agreement that cannot be explained by

Fig.1 Results per domain

reasons other than assessment discrepancies. These discrepancies were due to varying interpretations of which section of a PROM should be identified as the instruction.

3.3 Comprehensibility Results per PCQH Domain

Next, we address the results per domain based on the final ratings among the pairs of assessors (Fig. 1).

Almost all PROM scales (91%) were judged to be written on A1, A2, or B1 language level according to the Common European Framework of Reference for Languages (CEFR), which constitutes optimal adherence with the criteria for language level (Fig. 1) [71]. No PROM scale was written on language level C1 (red). We observed that in general the language level of the PROMs' instruction was rated more difficult (68% of the instruction's language level ratings were green) than that of the questions (91% of the questions' language level ratings were green) (data not shown in Fig. 1).

Only 48% of PROM scales included a general instruction (that was directed at the patient) and contained the subject/purpose of the PROM scale and a fill-in instruction. For a slightly smaller proportion, the rating was red (39%). These include PROM scales that lacked any form of instruction (15% of all PROMs).

With respect to the domain “number of questions,” we found that PROM scales with one question were most common, followed by PROM scales with ten questions (Fig. 2). More than half (54%, green rating) of the PROM scales contained six or fewer questions (Figs. 1 and 2). Only 13% of PROM scales contained 11 or more questions

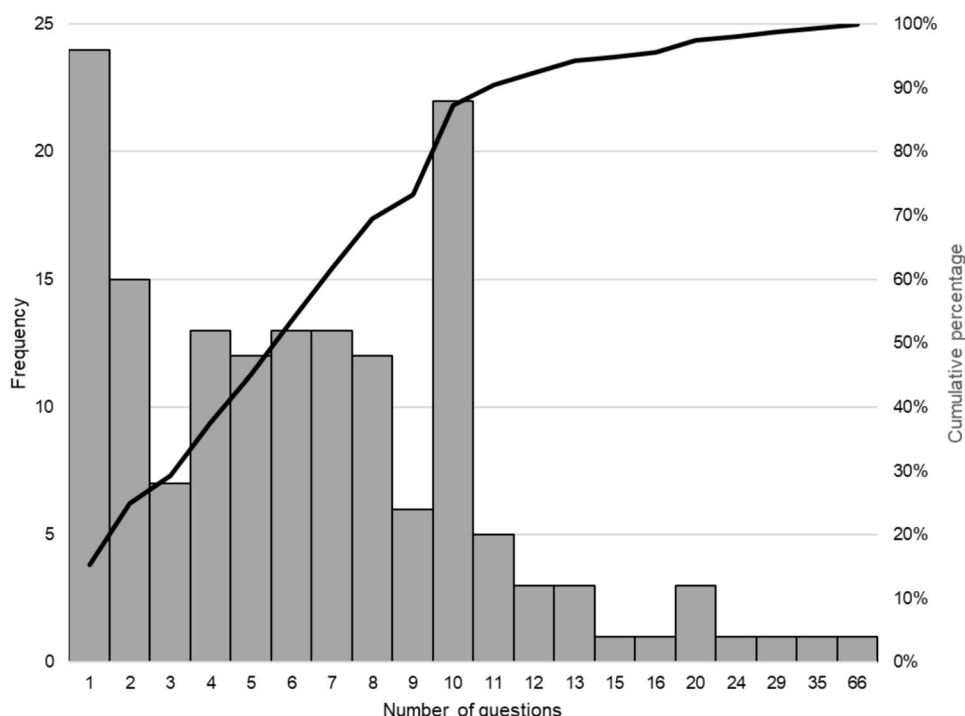
(red rating), with a maximum of 66 questions. Approximately half of all PROM scales (54%) contained a maximum of four answer options and/or used a numeric rating scale (NRS) ranging from 0 or 1 to 10. For almost one in five PROM scales (18%), the number of answer options was more than five or the PROM scale contained an open question.

Questions in a PROM scale were either all formulated actively (87%) or predominantly formulated passively (8%). This pattern was also seen in the concrete questions domain (72% exclusively concrete questions versus 11% predominantly non-concrete questions).

Abbreviations were not used at all in the assessed PROM scales and the use of medical terms was almost absent (98%). In four PROM scales, medical terms were used with (orange rating) or without (red rating) explaining them in lay terms.

Most PROM questions consisted of only interrogative sentences (green, 72%). Although the red category theoretically also includes PROMs that consist of a combination of interrogative sentences and statements ($\leq 25\%$ interrogative sentences), all PROM scales that are rated red on this domain (25%) consisted of statements only.

Table 3 shows the PROM scales that had a green rating on all eight domains of the comprehensibility component of the PCQH.

Fig.2 Domain “number of questions” in PROM scales**Table 3** Most comprehensible PROM scales

PROM	Scales
EORTC-QLQ-C30 ^a	Physical functioning, fatigue, cognitive functioning, mental health, sleep problems, shortness of breath, anorexia, nausea and vomiting, diarrhea
EORTC-QLQ-Br23 ^b	Body image, sexual functioning, arm/hand-symptoms, breast-symptoms
NRS fatigue	Not applicable
NRS pain intensity	Not applicable
NRS itch	Not applicable
VAS tch	Not applicable
TOPICS-SF ^c	NRS Quality of Life

EORTC-QLQ European Organization for Research and Treatment for Cancer Quality of Life Questionnaire, *NRS* numerical rating scale, *PROM* patient-reported outcome measure, *TOPICS-SF* The Older Persons and Informal Caregiver Survey—Short Form, *VAS* visual analogue scale

^aThe EORTC-QLQ-C30 scales obstipation, role functioning, and pain received an orange overall evaluation. The scales social functioning, financial problems and quality of life received a red overall evaluation

^bThe EORTC-QLQ-Br23 scale systemic side-effects received an orange overall evaluation

^cThe TOPICS-SF scales pain and physical functioning received an orange overall evaluation. The scales anxiety, perceived global health, and social functioning received a red overall evaluation

4 Discussion

The current study provides an extensive evaluation of the comprehensibility of 157 widely used and previously validated PROM scales included in 28 disease/condition-specific outcome sets and one generic PROM set that were developed as part of the National Outcome-Based Healthcare Program in the Netherlands.

A total of 18 out of 157 PROM scales (11%) had a green rating on all eight domains of the comprehensibility component of the PCQH and should be easy to use for all patients. Most PROM scales are on an appropriate language level and largely free from medical terms and passively formulated questions. However, approximately half of all PROMs lacked clear instructions. In fact, in 15% of all cases instructions were completely absent. Moreover, we found that individual PROM scales regularly comprised more than ten questions and five response options

per question. These factors reduce comprehensibility and may result in obtaining biased outcomes and preclude the participation of patients with lower (health) literacy skills in PROM initiatives.

A strength of this study is that a large number of PROM scales was assessed, which covers a wide range of diseases and conditions across Dutch medical specialist care. This suggests that the evaluated PROM scales constitute a representative sample of those that are now in use. Furthermore, the analysis that is presented results from final ratings by trained PhD level raters with expertise in outcome assessment and PROMs. The interrater reliability of the individual ratings was high for most domains. An exception is the instruction domain. Further examination of this finding showed that this was explained by differences in the interpretation of what part of a PROM to mark as an instruction. For example, some assessors only included instructions at the beginning of a PROM, while other assessors also included instructions that are provided with individual questions. During the consensus meetings, this was aligned and therefore resulted in one way of interpretation (i.e. instructions at the beginning of a PROM) for the final rating of this domain.

A potential limitation of this study is that the use of a computer algorithm embedded in the “Klinkende Taal” online tool for assessing language level [68] resulted in relatively little differentiation between PROMs. This tool evaluates the (Dutch) language level of a text based on a list of difficult words and a linguistic and syntactic analysis [68] and was used to assess the language level of the PROM scales in this study in a standardized way. To obtain an impression of the agreement with an expert judgment of the language level assessments, five carefully selected PROM scales were assessed both with “Klinkende Taal” and by an expert from Pharos (G.B.). In this small sample, the assessed language levels showed little difference between the expert from Pharos and “Klinkende Taal;” one PROM scale was judged to have a more difficult language level by the expert from Pharos, one PROM scale was judged to have a less difficult language level, and three PROM scales were judged equal. Another limitation is that the PCQH did not take into account whether or not qualitative methods were applied when developing the PROM scale. A final limitation of the study is that the level of complexity of questions also influences the comprehensibility and this is not captured by the domains of the PCQH. However, previous studies show that PROMs with complex questions may result in higher average time to complete. For instance, Van der Willik et al. (2019) show that the average time to complete the 66-item Dialysis Symptom Index (DSI) was 5.4 min [standard deviation (SD) 1.6] versus 7.5 min (SD 1.8) to complete the more

complex 21 symptom-items from the Palliative Care Outcome Scale—Renal Version (IPOS—Renal) [72].

The results of our study suggest that more attention is needed for the comprehensibility in the development and appraisal of PROMs. For example, qualitative methods and pilot tests are currently the recommended approach to assess comprehensibility of PROMs [43]. However, many of the more recent and well-known PROM scales evaluated in the present study, such as those derived from the PROMIS item banks or the SF-36 version 2, were developed through extensive qualitative and quantitative approaches to guarantee comprehensibility [73, 74]. Although improvements in preliminary versions of the respective PROMs through the application of these methods are well documented, our results show that this has not resulted in measures that meet all eight comprehensibility domains that were examined in our study. This finding suggests that use of the PCQH in PROM development might complement established methods for appraising comprehensibility and ultimately lead to the development of new PROMs that are easier to use for the intended patient populations. In addition, the fact that in our study domain scores are reliable across raters suggests that the PCQH could also be used to compare comprehensibility between different PROMs, which is difficult to do with qualitative approaches. Comparing comprehensibility of PROMs may be useful in systematic reviews or to inform PROM selection in addition to appraisal of psychometric properties. Online Resource 2 contains comprehensibility profiles per PROM scale and can be used as a practical tool during a PROM selection process to gain insight into and compare the comprehensibility of PROMs.

In this study, the PCQH was used for the first time as a tool for large scale evaluation of PROM comprehensibility, as it was originally developed to assess and subsequently improve comprehensibility of individual healthcare questionnaires. Although we are able to present some conclusions on the comprehensibility of the assessed PROM scales, more research is needed to evaluate the comprehensiveness and validity of the PCQH. This research should focus on the domains and the cutoff values for the domain ratings which might be considered arbitrary or subjective without further evidence to support them. However, for some domain ratings, such evidence does exist. For example, literature has shown that patients have trouble distinguishing between more than five Likert/rating scale response options and prefer four to five. This aligns with the PCQH’s classification which rates four response options as optimal, five as acceptable, and more than five as problematic. [73, 75]. Furthermore, there is experimental evidence that adding clear instructions substantially reduced the missing item rate and increased the overall response rate [78]. Finally, the number of questions is frequently used as a proxy measure of patient burden in validation studies [76, 77].

When using the PCQH to support PROM selection, it might be beneficial to have an overall evaluation measure of comprehensibility. For example, a graded scoring system in which numerical values (e.g. 0, 1, 2) are first assigned to the domains, which are then summed to provide an overall comprehensibility rating of a PROM scale. Future research on this should also focus on whether or not all domains need to be equally represented in such a measure.

Finally, the PCQH consists of three additional components not covered in this study: accessibility, layout and validation [54]. To improve the suitability of PROMs for all patients, we recommend consideration of these components as well. In addition, given the target group of the questionnaire, it might be needed to offer PROMs in multiple languages.

In conclusion, we have provided an extensive evaluation of 157 well-known and widely used PROM scales in eight relevant domains of comprehensibility. Our results provide actionable insights to improve the comprehensibility of PROMs, such as including clear instructions for patients and paying attention to the number of questions and answers. The use of PROMs is increasingly ubiquitous, including integration of PROMs in clinical workflows. For such applications it is essential that all patients, including those with low health literacy, can understand and complete their PROMs, ultimately facilitating the delivery of person-centered and more effective healthcare. Therefore, more attention is needed for comprehensibility when developing or selecting PROMs to be implemented in such initiatives.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40271-024-00710-w>.

Acknowledgments The authors acknowledge Maarten de Haan for reviewing this manuscript and offering many helpful suggestions for improvement.

Declarations

Funding The research leading to these results received funding from the Dutch Ministry of Health, Welfare and Sport.

Competing interests The authors have no relevant financial or non-financial interests to disclose.

Availability of data All descriptive results regarding the comprehensibility of the PROM scales in this study are based on final ratings available in Online Resource 2. Data on the individual assessment of comprehensibility of each PROM scale by the two methodologists, supporting the final ratings and the results on the interrater agreement reported in this study, are available on request from the authors.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Code availability Not applicable.

Author contributions All authors contributed to the study conception and design. Material preparation and data collection were performed by Attie Tuinenburg, Domino Determann, Elise H. Quik, Esmee M. van der Willik, Geeske Hofstra, Joannes M. Hallegraef, Ingrid Vriend, Lisanne Warmerdam, and Martijn A.H. Oude Voshaar. Analysis was performed by Martijn Oude Voshaar. The first draft of the manuscript was written by Attie Tuinenburg and Domino Determann and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

References

1. Crossnohere NL, Brundage M, Calvert MJ, et al. International guidance on the selection of patient-reported outcome measures in clinical trials: a review. *Qual Life Res.* 2021. <https://doi.org/10.1007/s11136-020-02625-z>.
2. Ministerie van Volksgezondheid, Welzijn en Sport. Ontwikkeling Uitkomstgerichte Zorg 2018-2022. Available at: <https://open.overheid.nl/documenten/ronl-b0848781-3c90-4b03-9515-f6b6a4cc168e/pdf>. Accessed Nov 2023
3. Turner M, Louie K, Chow C, et al. Advancing PROMs for health system use in Canada and beyond. *J Patient Rep Outcomes.* 2021. <https://doi.org/10.1186/s41687-021-00370-6>.
4. Emilsson L, Lindahl B, Köster M, et al. Review of 103 swedish healthcare quality registries. *J Intern Med.* 2015. <https://doi.org/10.1111/joim.12303>.
5. National Health Services England. National Patient Reported Outcome Measures (PROMs) Programme Guidance. 2017. Available at: <https://www.england.nhs.uk/wp-content/uploads/2017/09/proms-programme-guidance.pdf> (england.nhs.uk). Accessed Nov 2023.
6. PRO Secretariat. The Danish National Work on Patient Reported Outcomes. Available at: PRO – patient reported outcome - PRO (pro-danmark.dk). Accessed Nov 2023
7. Bennett AV, Jensen RE, Basch E. Electronic patient-reported outcome systems in oncology clinical practice. *CA Cancer J Clin.* 2012. <https://doi.org/10.3322/caac.21150>.
8. Katzan IL, Thompson NR, Lapin B, et al. Added value of patient-reported outcome measures in stroke clinical practice. *J Am Heart Assoc.* 2017. <https://doi.org/10.1161/JAHA.116.005356>.
9. Holmes MM, Lewith G, Newell D, et al. The impact of patient-reported outcome measures in clinical practice for pain: a systematic review. *Qual Life Res.* 2017. <https://doi.org/10.1007/s11136-016-1449-5>.
10. Kool M, Van der Sijp JRM, Kroep JR, et al. Importance of patient-reported outcome measures versus clinical outcomes for breast

- cancer patients evaluation on quality of care. *Breast*. 2016. <https://doi.org/10.1016/j.breast.2016.02.015>.
11. Proddinger B, Taylor P. Improving quality of care through patient-reported outcome measures (PROMs): expert interviews using the NHS PROMs Programme and the Swedish quality registers for knee and hip arthroplasty as examples. *BMC Health Serv Res*. 2018. <https://doi.org/10.1186/s12913-018-2898-z>.
 12. Arends D, Van Kooij Y, Loos N, et al. Uitkomstinformatie in de dagelijkse zorg: van verzamelen naar gebruiken. 2022. Available at: https://www.zonmw.nl/sites/zonmw/files/typo3-migrated-files/05160472110006_Rapport_Uitkomstinformatie_in_de_dagelijkse_zorg_van_verzamelen_naar_gebruiken_voor_beleidsmakers.pdf naargebruiken (zonmw.nl). Accessed Nov 2023.
 13. Calvert MJ, O'Connor DJ, Basch EM. Harnessing the patient voice in real-world evidence: the essential role of patient-reported outcomes. *Nat Rev Drug Discov*. 2019. <https://doi.org/10.1038/d41573-019-00088-7>.
 14. Roberts MH, Ferguson GT. Real-world evidence: bridging gaps in evidence to guide payer decisions. *Pharmacoecon Open*. 2021. <https://doi.org/10.1007/s41669-020-00221-y>.
 15. Porter M, Teisberg E. Redefining health care: creating value-based competition on results. 2006. Available at: <https://books.google.com/books?hl=en&lr=&id=Kp5fCkAzzS8C&oi=fnd&pg=PR10&ots=V-v3Oihodw&sig=uPUzb9JbIB5p984W0vvPAZ2M8jc>. Accessed Nov 2023
 16. Kim AH, Roberts C, Feagan BG, et al. Developing a standard set of patient-centred outcomes for inflammatory bowel disease—an international, cross-disciplinary consensus. *J Crohns Colitis*. 2018. <https://doi.org/10.1093/ecco-jcc/jjx161>.
 17. Wouters RM, Jobi-Odeneye AO, De la Torre A, et al. A standard set for outcome measurement in patients with hand and wrist conditions: consensus by the International Consortium for Health Outcomes Measurement Hand and Wrist Working Group. *J Hand Surg Am*. 2021. <https://doi.org/10.1016/j.jhsa.2021.06.004>.
 18. OMERACT. The OMERACT Handbook for establishing and implementing core outcomes in clinical trials across the spectrum of rheumatologic conditions. 2021. Available at: https://omeract.org/wp-content/uploads/2021/06/OMERACT-Handbook-Chapter-5_Final_June-2-2021_a.pdf. Accessed Nov 2023
 19. Nguyen H, Butow P, Dhillon H, et al. A review of the barriers to using patient-reported outcomes (PROs) and patient-reported outcome measures (PROMs) in routine cancer care. *J Med Radiat Sci*. 2021. <https://doi.org/10.1002/jmrs.421>.
 20. Dawson J, Doll H, Fitzpatrick R, et al. The routine use of patient reported outcome measures in healthcare settings. *BMJ*. 2010. <https://doi.org/10.1136/bmj.c186>.
 21. Van der Willik EM. Doctoral thesis: Implementation and use of patient-reported outcome measures in routine nephrology care. General discussion page 234-235. 2023. Available at: https://scholarlypublications.universiteit leiden.nl/handle/1887/3619965?solr_nav%5Bbid%5D=02a8286825af920e84c0&solr_nav%5Bpage%5D=0&solr_nav%5Boffset%5D=0. Accessed Dec 2023.
 22. Prinsen CAC, Vohra S, Rose RR, et al. Guideline for selecting outcome measurement instruments for outcomes included in a core outcome set. COSMIN. Available at: https://static-content.springer.com/esm/art%3A10.1186%2F13063-016-1555-2/MediaObjects/13063_2016_1555_MOESM2_ESM.pdf. Accessed Nov 2023
 23. The HLS19 Consortium of the WHO Action Network M-POHL. International Report on the methodology, results, and recommendations of the European Health Literacy Population Survey 2019-2021 (HLS19) of M-POHL. 2021. Available at: <https://m-pohl.net/sites/m-pohl.net/files/inline-files/HLS19%20International%20Report.pdf>. Accessed Dec 2023.
 24. Wikkeling-Scott LF, Ajja RJM, Vann RR. Health literacy research in the Eastern Mediterranean Region: an integrative review. *Int J Public Health*. 2019. <https://doi.org/10.1007/s00038-018-01200-1>.
 25. Berkman ND, Sheridan SL, Donahue KE, et al. Low health literacy and health outcomes: an updated systematic review. *Ann Intern Med*. 2011. <https://doi.org/10.7326/0003-4819-155-2-201107190-00005>.
 26. Van der Heide I, Rademakers J, Schipper M, et al. Health literacy of Dutch adults: a cross sectional survey. *BMC Public Health*. 2013. <https://doi.org/10.1186/1471-2458-13-179>.
 27. Willems A, Heijmans M, Barbers A, et al. Gezondheidsvaardigheden in Nederland: factsheet cijfers 2021. Nivel. 2022. Available at: <https://www.nivel.nl/sites/default/files/bestanden/1004162.pdf>. Accessed Dec 2023.
 28. Sørensen K, Van den Broucke S, Fullam J, et al. Health literacy and public health: a systematic review and integration of definitions and models. *BMC Public Health*. 2012. <https://doi.org/10.1186/1471-2458-12-80>.
 29. Heijmans M, Brabers A, Rademakers J. Hoe gezondheidsvaardig is Nederland? Factsheet gezondheidsvaardigheden - Cijfers 2019. Nivel. 2019. Available at: <https://www.nivel.nl/nl/publicatie/hoe-gezondheidsvaardig-nederland-factsheet-gezondheidsvaardigheden-cijfers-2019>. Accessed Nov 2023.
 30. World Health Organization. Health literacy, the solid facts. 2013. Available at: <https://www.who.int/europe/publications/i/item/9789289000154>. Accessed Nov 2023
 31. Stormacq C, Wosinski J, Biollat E, et al. Effects of health literacy interventions on health-related outcomes in socioeconomically disadvantaged adults living in the community: a systematic review. *JBI Evid Synth*. 2020. <https://doi.org/10.1112/JBISR-IR-D-18-00023>.
 32. Calvert MJ, Cruz Rivera S, Retzer A, et al. Patient reported outcome assessment must be inclusive and equitable. *Nat Med*. 2022. <https://doi.org/10.1038/s41591-022-01781-8>.
 33. Long C, Beres LK, Wu AW, et al. Patient-level barriers and facilitators to completion of patient-reported outcomes measures. *Qual Life Res*. 2022. <https://doi.org/10.1007/s11136-021-02999-8>.
 34. ISPOR. Available at: <https://www.ispor.org/home>. Accessed Jun 2024
 35. ICHOM. Available at: <https://www.ichom.org/>. Accessed Jun 2024
 36. COSMIN. Available at: <https://www.cosmin.nl/>. Accessed Jun 2024
 37. PROM-cycle. Available at: <https://www.zorginzicht.nl/ondersteuning/prom-cyclus/over-de-prom-cyclus>. Accessed Jun 2024
 38. ISOQOL. Available at: <https://www.isoqol.org/>. Accessed Jun 2024
 39. Van der Wees PJ, Verkerk EW, Verbiest MEA, et al. Development of a framework with tools to support the selection and implementation of patient-reported outcome measures. *J Patient Rep Outcomes*. 2019. <https://doi.org/10.1186/s41687-019-0171-9>.
 40. Reeve BB, Wyrwich KW, Wu AW. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res*. 2013. <https://doi.org/10.1007/s11136-012-0344-y>.
 41. Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity—Establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1—eliciting concepts for a new PRO instrument. *Value Health*. 2011
 42. Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity—Establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 2—assessing respondent understanding. *Value Health*. 2011

43. Terwee CB, Prinsen CAC, Chiarotti A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res.* 2018. <https://doi.org/10.1007/s11136-018-1829-0>.
44. Wiering B, De Boer D, Delnoij D. Patient involvement in the development of patient-reported outcome measures: a scoping review. *Health Expect.* 2017.
45. Issa TZ, Lee Y, Mazmudar AS, et al. Readability of patient reported outcomes in spine surgery and implications for health literacy. *Spine.* 2023. <https://doi.org/10.1097/BRS.00000000000004761>.
46. Lee SE, Farzal Z, CS Ebert Jr. et al. Readability of patient-reported outcome measures for head and neck oncology. *Laryngoscope.* 2022. <https://doi.org/10.1002/lary.28555>
47. Rao SJ, Nickel JC, Kiell EP, et al. Readability of commonly used patient-reported outcome measures of laryngoscopy. *Laryngoscope.* 2022. <https://doi.org/10.1002/lary.29849>.
48. Clarke M, Williamson P. Core outcome sets and trial registries. *Trials.* 2025. <https://doi.org/10.1186/s13063-015-0738-6>.
49. Oude Voshaar M, Terwee CB, Haverman L, et al. Development of a standard set of PROs and generic PROMs for Dutch medical specialist care: Recommendations from the Outcome-based Healthcare Program Working Group Generic PROMs. *Qual Life Res.* 2023. <https://doi.org/10.1007/s11136-022-03328-3>.
50. Platform Uitkomstgerichte Zorg. Inzicht in uitkomsten. Available at: <https://platformuitkomstgerichtezorg.nl/themas/inzicht-in-uitkomsten/default.aspx>. Accessed Nov 2023
51. Zorginstituut Nederland. Overzicht 50% van de Nederlandse ziektelast: aandoeningen met een voorsprong op het gebied van uitkomstinformatie en geschikt voor samen beslissen. 2018. Available at: <https://www.zorginstituutnederland.nl/publicaties/rapport/2018/06/28/rapport-overzicht-50-van-de-nederlandse-ziektelast>. Accessed Nov 2023
52. Programma Uitkomstgerichte Zorg Lijn 1 'Meer inzicht in uitkomsten'. Adviesrapport set Generieke PRO(M)s. 2022. Available at: <https://www.platformuitkomstgerichtezorg.nl/aan-de-slag/documenten/handlerdownloadfiles.ashx?idnv=2148205>. Accessed Nov 2023
53. Programma Uitkomstgerichte Zorg Lijn 1 'Meer inzicht in uitkomsten'. Adviesrapport set Generieke PRO(M)s voor kinderen. 2023. Available at: <https://www.platformuitkomstgerichtezorg.nl/aan-de-slag/documenten/handlerdownloadfiles.ashx?idnv=2471004>. Accessed Nov 2023
54. Pharos. Pharos Checklist for Questionnaires in Healthcare. Available at: <https://www.pharos.nl/kennisbank/test-je-vragenlijst-op-begrijpelijkheid/>. Accessed Jun 2023
55. Pharos. Gezondheidsverschillen duurzaam aanpakken - de 9 principes voor een succesvolle strategie. Available at: <https://www.pharos.nl/gezondheidsverschillen-duurzaam-aanpakken/een-succesvolle-strategie> - Pharos. Accessed Nov 2023
56. Pharos. Methodiek: begrijpelijke medische informatie in woord en beeld ter ondersteuning bij het uitleggen en samen beslissen. 2021. Available at: <https://www.pharos.nl/nieuws/methodiek-begrijpelijke-medische-informatie-in-woord-en-beeld/>. Accessed Nov 2023
57. Charters E. The use of Think-aloud methods in qualitative research an introduction to Think-aloud methods. *Brock Educ J.* 2003. <https://doi.org/10.26522/BROCKED.V12I2.38>.
58. Eccles DW, Arsal G. The think aloud method: what is it and how do I use it? *Qual Res Sport Exerc Health.* 2017. <https://doi.org/10.1080/2159676X.2017.1331501>.
59. Fang J, Fleck MP, Green A, et al. The response scale for the intellectual disability module of the WHOQOL: 5-point or 3-point. *J Intellect Disabil Res.* 2011. <https://doi.org/10.1111/j.1365-2788.2011.01401.x>.
60. Schoemaker SJ, Wolf MS, Brach C. Development of the patient education materials assessment tool (PEMAT): a new measure of understandability and actionability for print and audiovisual patient information. *Patient Educ Couns.* 2014. <https://doi.org/10.1016/j.pec.2014.05.027>.
61. Berger U, Fehlinger M, Mühleck J, et al. Inclusive research: validation of the general self-efficacy scale in simple language in a sample of students with special educational needs. *Psychother Psychosom Med Psychol.* 2019. <https://doi.org/10.1055/a-0831-2270>.
62. Kooijmans R, Mercera G, Langdon PE, et al. The adaptation of self-report measures to the needs of people with intellectual disabilities: a systematic review. *Clin Psychol Sci Pract.* 2022. <https://doi.org/10.1037/cps0000058>.
63. Taylor S, Guirguis M, Raney EM. Can patients and families read the questionnaires for patient-related outcome measures? *J Pediatr Orthop.* 2019. <https://doi.org/10.1097/BPO.0000000000001327>.
64. Clerehan R, Guillemin F, Epstein J, et al. Using the evaluative linguistic framework for questionnaires to assess comprehensibility of self-report health questionnaires. *Value Health.* 2016. <https://doi.org/10.1016/j.jval.2016.01.008>.
65. Chung KC, Pillsbury MS, Walters MR, et al. Reliability and validity testing of the Michigan hand outcomes questionnaire. *J Hand Surg Am.* 1998. [https://doi.org/10.1016/S0363-5023\(98\)80042-7](https://doi.org/10.1016/S0363-5023(98)80042-7).
66. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika.* 1951. <https://doi.org/10.1007/BF02310555>.
67. Aaronson NK, Muller M, Cohen PD, et al. Translation, validation, and norming of the Dutch language version of the SF-36 health survey in community and chronic disease populations. *J Clin Epidemiol.* 1998. [https://doi.org/10.1016/s0895-4356\(98\)00097-3](https://doi.org/10.1016/s0895-4356(98)00097-3).
68. Klinkende Taal. Available at: <https://beoordeel-tekst.klinkendetaal.nl/>. Accessed Aug 2023.
69. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979. <https://doi.org/10.1037/0033-2909.86.2.420>.
70. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159–74.
71. Council of Europe. Common European Framework of References and Languages (CEFR). Available at: <https://www.coe.int/en/web/common-european-framework-reference-languages>. Accessed Oct 2023
72. Van der Willik EM, Meuleman Y, Prantl K, et al. Patient-reported outcome measures: selection of a valid questionnaire for routine symptom assessment in patients with advanced chronic kidney disease - a four-phase mixed methods study. *BMC Nephrol.* 2019. <https://doi.org/10.1186/s12882-019-1521-9>.
73. Bruce B, Fries JF, Ambrosini D, et al. Better assessment of physical function: item improvement is neglected but essential. *Arthritis Res Ther.* 2009. <https://doi.org/10.1186/ar2890>.
74. Ware JE, Kosinski M. The SF-36 Health Survey (Version 2.0). Technical Note. Boston, MA, Health Assessment Lab, Sep. 1996
75. Kroenke K, Monohan PO, Kean J. Pragmatic characteristics of patient-reported outcome measures are important for use in clinical practice. *J Clin Epidemiol.* 2015. <https://doi.org/10.1016/j.jclinepi.2015.03.02>.
76. Oude Voshaar MAH, Das Gupta Z, Bijlsma JWI, et al. International Consortium for Health Outcome Measurement Set of Outcomes that matter to people living with inflammatory arthritis: consensus from an international working group. *Arthritis Care Res.* 2019. <https://doi.org/10.1002/acr.23799>.

77. Verberne WR, Das Gupta Z, Allegretti AS, et al. Development of an international standard set of value based outcome measures for patients with chronic kidney disease: a report of the International Consortium for Health Outcome Measurement (ICHOM) CDK Working Group. *Am J Kidney Dis*. 2019. <https://doi.org/10.1053/j.ajkd.2018.10.007>.
78. Van Beers LWAH, Scholtes VAB, Van Wermeskerken M. Clear instructions reduce missing responses in pen-and-paper collected patient reported outcome measures: a randomized study. *Netherlands Tijdschrift voor Orthopedie*. 2015;22:51–3.

Authors and Affiliations

Attie Tuinenburg¹ · Domino Determann¹ · Elise H. Quik¹ · Esmee M. van der Willik¹ · Geeske Hofstra¹ · Joannes M. Hallegraeff^{1,2} · Ingrid Vriend¹ · Lianne Warmerdam¹ · Hester E. van Bommel³ · Gudule Boland³ · Martijn A. H. Oude Voshaar^{1,4}

✉ Martijn A. H. Oude Voshaar
a.oudevoshaar@erasmusmc.nl

- ¹ National Health Care Institute (Zorginstituut Nederland), Diemen, The Netherlands
- ² Experimental Anatomy Research Group, Department of Physiotherapy, Human Physiology and Anatomy, Faculty

of Physical Education and Physiotherapy, Vrije Universiteit Brussel, Brussel, Belgium

- ³ Dutch Centre of Expertise on Health Disparities, Utrecht, The Netherlands
- ⁴ Department of Public Health, Erasmus Medical Center, Rotterdam, The Netherlands