



OPEN ACCESS

Quality of instruments to assess the process of shared decision-making: a comprehensive systematic review and COSMIN quality appraisal

Christin Hoffmann ,¹ Samuel Lawday,¹ Vignesh Gopalan,^{1,2} Rados Keravica,³ Arwen Pieterse ,⁴ Hilary Bekker,⁵ Denitza Williams,⁶ Natalie Joseph-Williams ,⁶ Jennifer Hall,³ Kerry Avery,¹ Val Snelgrove,⁷ Della Hopkins,² Christie Cabral,⁸ Jane Blazeby,¹ Shireen Hickey,³ Angus G K McNair^{1,2}

10.1136/bmjebm-2025-114092

► Additional supplemental material is published online only. To view, please visit the journal online (<https://doi.org/10.1136/bmjebm-2025-114092>).

For numbered affiliations see end of article.

Correspondence to:
Dr Christin Hoffmann; c.hoffmann@bristol.ac.uk



© Author(s) (or their employer(s)) 2026. Re-use permitted under CC BY. Published by BMJ Group.

To cite: Hoffmann C, Lawday S, Gopalan V, et al. *BMJ Evidence-Based Medicine* Epub ahead of print: [please include Day Month Year]. doi:10.1136/bmjebm-2025-114092

Abstract

Objective Shared decision-making (SDM) is a key component of patient-centred care and major driver of healthcare policy, clinical practice and research globally. Widespread implementation of SDM is hindered by uncertainty about how to best measure decision-making processes. A seminal methodological review of SDM instruments published in 2018 identified widespread deficiencies in instrument quality and provided recommendations for improvement. This systematic review and Consensus-based Standards for the selection of health status Measurement INstruments (COSMIN) quality appraisal aims to provide an update SDM instrument inventory, re-evaluated progress in the field and provided future direction for researchers, funders and policymakers. This systematic review and Consensus-based Standards for the selection of health status Measurement INstruments (COSMIN) quality appraisal aims to provide an update SDM instrument inventory, re-evaluated progress in the field and provided future direction for researchers, funders and policymakers.

Design A systematic search of six databases (September 2017–December 2024) identified studies of instruments measuring the process of SDM. Instrument appraisal followed COSMIN guidelines in a three-step process: appraisal of the methodological quality of instruments and of the quality of the measurement properties of newly identified measures (steps i, ii). This informed the comprehensive best-evidence synthesis which included the data from the original review (step iii).

Results A total of 127 studies were included describing the development and/or evaluation of 105 unique instruments to measure the process of SDM. Some 61 new instruments were identified since the last review, of which 35 (57%) were translations/revisions to existing ones. The best-evidence synthesis revealed positive results for internal consistency (72%), structural validity (75%) and intrarater reliability (50%), but also negative or unknown results for test–retest reliability (58%), content validity (54%) and intrarater reliability (50%), where evaluated. No single instrument consistently showed positive evidence across all

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ A seminal systematic review and quality appraisal of instruments to measure the process of shared decision-making (SDM) was published in 2018 to facilitate evidence-based choices of SDM instruments. This work provided a comprehensive inventory of SDM instruments, but concluded there was an overall lack of evidence on their measurement quality.

WHAT THIS STUDY ADDS

⇒ A total of 61 new instruments to measure the process of SDM were identified. No single instrument was identified to have been sufficiently evaluated and to have consistently obtained ratings of good methodological quality.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ This systematic review has provided an up-to-date inventory of instruments to assess the SDM process in research, clinical care and health policy. Evidence gaps remain among instruments that are widely used in research and clinical practice and are recommended by health policy, suggesting there is a need for prioritisation through international stakeholder consensus to focus future research activities on key measurement instruments to strengthen their psychometric evidence and performance.

relevant measurement properties. Evidence gaps for many properties remain.

Conclusion This updated inventory guides researchers, clinicians and policymakers to the most

appropriate instrument for an intended context/population. There was a proliferation of new, primarily patient-reported measurement instruments, but no single instrument had sufficient evidence of measurement quality. Methods were aligned with the original review and applied the 2011 COSMIN checklist which was updated in 2024, although using a previous checklist version is unlikely to affect overall conclusions. We recommend an international consensus with key interest holders on preferred instrument(s) for further validation and future core measurement set for efficient evidence synthesis.

Registration Prospero: CRD42024485655

Introduction

Shared decision-making (SDM) is a widely endorsed approach in which patients are actively supported to make healthcare decisions in partnership with clinicians, and it is recognised as a key component of patient-centred care.^{1 2} The last two decades have seen a proliferation of SDM research and intervention development to improve SDM internationally.³⁻⁵ There is an ongoing discourse, however, about how to define and accurately measure SDM in different settings. This has led to continuous evolution of conceptualisations of SDM and their methods of assessment, which include measures from the perspectives of patients, providers and observers.⁶⁻¹³

The global integration of SDM into healthcare policies worldwide reflects its widespread recognition as a cornerstone of patient-centred care.^{3 14-16} Data from SDM instruments are increasingly used to inform these policies and evaluate and guide clinical practice. For example, national health bodies in the UK (National Institute for Health and Care Excellence) and Australia (Commission on Safety and Quality in Healthcare) have embedded SDM into national guidelines, supported by evidence from patient-reported outcome and experience measures.^{14 16 17} Routine SDM assessment can provide valuable insights into how effectively SDM is achieved and what impact SDM interventions may have, provided this is underpinned by robust measurement.^{18 19} There is, however, overall weak evidence regarding the effectiveness and impact of SDM interventions on patient, clinical and health service outcomes.²⁰⁻²² This is, to a large part, due to challenges in measurement underpinned by heterogeneity of SDM conceptualisation and measurement instruments.^{20 23 24}

The ability of an SDM instrument to accurately capture experiences and subsequently inform improvements in SDM significantly depends on its psychometric strength.^{25 26} A seminal systematic review and quality appraisal of instruments to measure the process of SDM was published in 2018 to facilitate evidence-based choices of SDM instruments.²⁷ This work provided a comprehensive inventory of SDM instruments, but concluded there was an overall lack of evidence on their measurement quality. It outlined seven key recommendations for future research to address deficiencies and improve SDM assessment. It is now important to consider evidence from the last decade, and to assess developments in measurement quality of SDM instruments.

This study aimed to conduct a systematic review and COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) quality appraisal to provide a revised, up-to-date inventory of instruments assessing the process of SDM and evidence of their psychometric quality.

Methods

We followed COSMIN guidelines^{26 28 29} and a comprehensive description of the methods is published elsewhere²⁷ and any

deviations outlined in the relevant online supplemental appendices. We adhered to the format and structure of the original systematic review for consistency of reporting and used the COSMIN checklist (2011 version) to facilitate best-evidence synthesis.²⁹ This review was registered in PROSPERO (CRD42024485655).

Identification of articles

An updated systematic search of six databases (PubMed, Embase, Cochrane, PsycINFO, Web of Science, Emcare) identified studies investigating instruments (patient-reported/observer coded/combined) to measure the process of SDM between 1 September 2017 and 16 December 2024. This was supplemented by snowball citation searches of included articles and contact with academic SDM networks (Society of Medical Decision Making, International Shared Decision Making Society) to obtain information about ongoing work to develop or validate relevant measurement instruments. The search strategy is presented in online supplemental appendix 1.

Selection of eligible studies

Eligibility criteria are identical to those used in the original review for consistency (box 1). All records underwent screening by title/abstract before full text review of potentially eligible articles. Screening and full text review was completed by at least two independent researchers. Discrepancies in inclusion/exclusion decisions were discussed between the reviewers and referred to the wider study team if no decision could be made. Reasons for inclusion/exclusion were recorded.

Data extraction

Two electronic forms were used to extract information (see online supplemental appendix 2, section 1 and 2), with each capturing details about (1) the article and (2) unique SDM measurement instruments identified in the articles. Any adapted or revised version of an original measurement instrument, including versions that differ in language or perspective, was considered a unique instrument and evaluated separately in line with COSMIN. Data extraction only considered the information presented in the included articles, unless there was specific reference to another information source. Corresponding authors were contacted once via email where further clarification was considered necessary.

Information about the article included details about the study characteristics (author, publication year, study title), research methods (setting, population and recruitment, data collection and if applicable, coding) and results for each of the ten measurement properties originally based on COSMIN^{28 30}: internal consistency, reliability, measurement error, content validity, construct validity (including structural validity, hypotheses testing and cross-cultural validity), criterion validity and responsiveness (see table 1 for definitions). Information about SDM measurement instruments included language, type, information about (sub) scales, the measurement characteristics (construct, aim, viewpoint, measurement model, target setting) and the development process.

There was a three-stage quality assurance process: (1) a reviewer (CH, VG, SL, DW, NJ-W, AGKM) independently extracted data for each included article; (2) a second reviewer then checked/verified all extracted data; and (3) a third independent reviewer (CH, VG, SL) adjudicated on discrepancies. Data extraction, checking and adjudication were performed iteratively alongside regular team meetings to ensure a consistent and robust approach to data extraction.

Box 1 Eligibility criteria³¹**Inclusion criteria**

1. The article had to describe a primary study in which the development or evaluation of one or more instruments occurred.
2. Instruments under investigation:
 - a. Were developed with the aim of measuring the process of SDM between a patient (with or without family) or proxy and a healthcare provider.
 - b. Were evaluated in their ability to measure the process of SDM even though they were not originally developed to measure the process of SDM.
 - c. Were developed or evaluated in their ability to measure patient participation in decision-making. To guarantee a focus on SDM, these instruments should assess at least one of four key steps of SDM¹⁻³:
 - i. Explaining that a decision has to be made.
 - ii. Discussing all relevant treatment options and their associated benefits and harms.
 - iii. Discussing patients' ideas, concerns and expectations and supporting patients in the process of deliberation before reaching a decision.
 - iv. Patient involvement in making the final decision.
3. The article had been peer-reviewed. (Not applicable to unpublished work received via ISDM and SMDM network members.)
4. The article was written in English, Dutch or German.

Exclusion criteria

To guarantee that the instrument under investigation measures a decision-making process that includes both the healthcare provider and the patient, the following two exclusion criteria were applied:

1. Articles investigating instruments that measure interprofessional SDM that do not include the participation of patients.
2. Articles about instruments developed or evaluated for the measurement of SDM about screening or genetic testing. These decisions often rather relate to informed decision-making and thus crucially differ from SDM in two aspects:
 - a. The healthcare provider is not necessarily involved in making the decision.
 - b. A decision usually is not needed by a certain time point.
3. Articles that evaluate instruments measuring antecedents of SDM (eg, preferred role in decision-making) or SDM outcomes (such as decisional regret)

No restrictions were held for

1. The type of measurement instrument (eg, self-report questionnaire or coding scheme).
2. The healthcare setting in which the instrument was evaluated.

Continued

Box 1 Continued

ISDM, International Shared Decision Making; SDM, shared decision-making; SMDM, Society of Medical Decision Making.

Quality appraisal of measurement properties of SDM instruments

Quality appraisal of measurement properties followed a three-step process: (i) appraisal of the methodological quality of included articles, (ii) appraisal of the measurement properties based on reported results of included articles and (iii) best-evidence synthesis for each unique SDM instrument using appraisal data obtained from steps (i) and (ii) and including data from the original review to consider the full body of evidence.

Appraisal of methodological quality and measurement properties was completed independently by one reviewer and subsequently checked by a second reviewer. The review team consisted of coauthors CH, SL, VG, DW, NJ-W, AGKM who met regularly to discuss the quality appraisal process and clarify any criteria as necessary. Discrepancies were discussed between the two relevant reviewers and any remaining disagreements were solved through team discussions. Best-evidence synthesis was completed by two reviewers independently, then compared/consolidated with input from a third reviewer with discrepancies resolved in team discussions. Team members who authored any of the included articles were not involved in data extraction and quality appraisals of their own article.

Appraisal of methodological quality

Ten measurement properties (see online supplemental appendix 2, section 3 for definitions) were evaluated for the articles identified in this updated review, where applicable. Appraisal adhered to COSMIN terminology/definitions throughout, regardless of author-reported measurement properties. Standardised scoring criteria were applied for each measurement property on a four-point scale: 'excellent', 'good', 'fair' or 'poor'. An overall score was determined by the lowest score assigned to any individual criterion within the relevant set.

Reliability of questionnaires and coding schemes were scored according to the relevant methodology, whereby test-retest reliability was applicable to questionnaires only, and inter-rater and intrarater reliability were applicable to coding schemes only. Methodological quality of articles using item response theory models was documented separately to internal consistency and structural validity.

Appraisal of the measurement properties

Appraisal of measurement properties was based on the evaluation of the relevant measurement property using a three-point quality rating: '+' (positive result for a measurement property); '-' (negative result for a measurement property) and '?' (result of measurement property is unknown). The detailed quality criteria are presented in [table 1](#).

Best-evidence synthesis

The best-evidence synthesis (ie, overall quality of the relevant measurement property) was determined for each measurement property per instrument using COSMIN criteria^{31 32} (see [table 2](#)). This synthesis considered the results of the appraisals of methodological quality (2.4.1) and measurement properties per study (2.4.2), the overall number of studies assessing the measurement

Table 1 Quality criteria for results on measurement properties^{28 30}

Measurement property	Criteria for appraisal of the results on measurement properties evaluation
1. Internal consistency	+ Cronbach's alpha(s) are ≥ 0.70 . ? Not able to score because of unclear or missing information, eg, the dimensionality is not known or Cronbach's alpha(s) are not presented. – Criteria for '+' not met.
2. Reliability	+ ICC agreement/weighted Kappa ≥ 0.70 OR ICC consistency/ICC without approach stated/Pearson's $r \geq 0.80$ OR unweighted kappa/or kappa without approach stated ≥ 0.80 . ? Not able to score because of unclear or missing information, eg, neither ICC, Kappa, nor Pearson's r is determined. – Criteria for '+' not met.
3. Measurement error	+ MIC $\geq$ SDC OR MIC outside the LOA OR convincing arguments that agreement is acceptable ? Not able to score because of unclear or missing information, eg, SEM, SDC not calculated or MIC not defined. – Criteria for '+' not met.
4. Content validity	+ Target group and/or experts considered all items to be relevant AND considered the item set to be complete. ? Not able to score because of unclear or missing information, eg, no results on item relevance according to experts reported. – Criteria for '+' not met.
Construct validity	
5. Structural validity	+ For exploratory factor analyses: Factors chosen explain at least 50% of variance OR factors chosen explain less than 50% of variance but the choice is justified by the authors. For confirmatory factor analyses: (The goodness of fit indicators fulfil the following requirements: (CFI or TLI or GFI or comparable measure > 0.90) AND (RMSEA or SRMR < 0.08)) AND (results confirm models with the original factor structure OR results confirm a model with slight changes if these changes are justified by the authors). ? For exploratory factor analyses: Not able to score because of unclear or missing information, eg, explained variance not mentioned. For confirmatory factor analyses: Not able to score because of unclear or missing information, eg, no fit indices are presented. – Criteria for '+' not met.
6. Hypotheses testing	+ (At least 75% of the results are in accordance with the hypotheses AND if calculated, the correlation with an instrument measuring the same construct is ≥ 0.50) AND correlations with related constructs are higher than with unrelated constructs if calculated. ? Not able to score because of unclear or missing information, eg, no correlations with related construct are calculated. – Criteria for '+' not met.
7. Cross-cultural validity	+ The original factor structure is confirmed AND no important DIF found. If only one of these properties is investigated: either the factor structure is confirmed OR no important DIF found. ? Not able to score because of unclear or missing information, eg, no confirmative factor analyses are performed nor the DIF is investigated. – Criteria for '+' not met
8. Criterion validity	+ Correlations with chosen gold standard ≥ 0.70 OR AUC ≥ 0.80 OR (specificity AND sensitivity ≥ 80). ? Not able to score because of unclear or missing information. – Criteria for '+' not met.
9. Responsiveness	+ Correlations of change scores of the target instrument with an instrument measuring the same construct are ≥ 0.40 OR at least 75% of the results are in accordance with the hypotheses OR AUC ≥ 0.70) AND correlations of change scores of the target instrument with an instrument measuring related constructs are higher than with unrelated construct if calculated. ? Not able to score because of unclear or missing information, eg, no correlations of change score with related constructs are calculated or no AUC investigated. – Change score correlation with an instrument measuring the same construct < 0.40 OR $< 75\%$ of the results are in accordance with the hypotheses OR AUC < 0.70 OR change score correlations with related constructs are lower than with unrelated constructs.
10. Interpretability	No quality scoring performed.
IRT	+ At least limited evidence for unidimensionality or positive structural validity AND no evidence for violation of local independence: Rasch: standardised item-person fit residuals between -2.5 and 2.5 ; OR IRT: residual correlations among the items after controlling for the dominant factor < 0.20 OR Q3's < 0.37 AND no evidence for violation of monotonicity: adequate looking graphs OR item scalability > 0.30 AND adequate model fit: Rasch: infit and outfit mean squares ≥ 0.5 and ≤ 1.5 OR Z-standardised values > -2 and < 2 OR IRT: G2 > 0.01 . Optional additional evidence: Adequate targeting; Rasch: adequate person-item threshold distribution; IRT: adequate threshold range. No important DIF for relevant subject characteristics (such as age, gender, education), McFadden's $R^2 < 0.02$. ? Model fit not reported. – Criteria for '+' not met.

Continued

Table 1 Continued

Measurement property	Criteria for appraisal of the results on measurement properties evaluation
+ = positive result for a measurement property.	
? = result of measurement property is unknown.	
– = negative result for a measurement property.	
AUC, area under the curve; CFI, Comparative Fit Index; DIF, Differential Item Function; GFI, Goodness of Fit Index; ICC, intra-class correlation coefficient; IRT, item response theory; LOA, limits of agreement; MIC, minimal important change; RMSEA, Root Mean Square Error of Approximation; SDC, smallest detectable change; SRMR, Standardised Root Mean Squared Residual; TLI, Tucker-Lewis Index.	

property and, if applicable, the consistency of results from multiple validation studies. Findings from the original review and current study were pooled to provide an up-to-date best-evidence synthesis.

An overall rating was determined by applying five levels of evidence, either positive or negative: strong (labelled '+++' or '---'), moderate (labelled '++' or '--'), limited (labelled '+' or '-'), conflicting (labelled '+/-') and unknown evidence (labelled '?').

Results

Search results

A total of 127 articles were included in this comprehensive review, combining included articles from the original review²⁷ and the current systematic review. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram (figure 1) shows the combined search results. One article in Spanish language³³ was excluded during the full-text screening process of this systematic review and its eligibility could consequently not be assessed.

A total of 105 SDM instruments (77 questionnaires and 28 observer-based coding schemes) were available for best-evidence synthesis, combining included SDM instruments from the original and updated review. An overview of SDM instrument characteristics is presented in table 3. The up-to-date inventory of SDM measurement instruments, their key characteristics and results of the best-evidence synthesis is shown in online supplemental appendix 3. Full details of all measurement instrument characteristics are presented in online supplemental appendix 4, table S1.

This update identified a total of 61 new instruments compared with the original review. Of those, 35 (57%) reported the translation (n=28) or revision (n=7) of existing measures. New instruments included 42 additional patient/proxy questionnaires (23 translations/revisions), 6 additional provider questionnaires (4 translations/revisions), 9 additional coding schemes (7 translations/

revisions) and 3 additional mixed instruments (1 translation/revision) published since the original systematic review in 2018.²⁷ No new psychometric evidence was identified for 21 of the 40 SDM instruments identified in the original systematic review.

During this systematic review update, corresponding authors of two articles were contacted to clarify study details. One author responded but did not provide the requested information, while the other did not respond.

Quality appraisal

Separate ratings for both the appraisal of the methodological quality (step i) and the quality of measurement properties (step ii) for each included article are presented in online supplemental appendix 4, table S2.

The results of the best-evidence synthesis (step iii) for each measurement property per instrument are included in online supplemental appendix 3. The full results are presented in online supplemental appendix 4, table S3. Information about how often measurement properties were applicable and how often they were assessed is included in online supplemental appendix 4, table S4.

The most frequently evaluated questionnaires were the SDM Process scale (n=6), the SDM-Q-9 versions in Spanish (n=6) and English (n=5) and the CollaboRATE (n=5). The OPTION-12 (12-item Observing Patient Involvement in Decision Making instrument) (also known as OPTION (revised)), the OPTION-5 and the observer-based tools included in the MAPPIN'SDM Inventory were among the most frequently evaluated coding schemes with seven, five and six validation studies respectively.

10 instruments demonstrated positive levels of evidence across two measurement properties, potentially offering greater confidence in their quality: the SDM-Q-9 versions in German, Dutch and Finnish, the SDM-Q-Doc versions in German, Dutch and Japanese, the patient-centred primary care instrument, the Dietician PCC—Patient Version (Patient version of the Dietician Patient Centered Care Measure) and the SDM-C Japanese (Care SDM-Questionnaire for care receivers) (patient and provider versions). No single instrument showed only positive evidence across all relevant measurement properties assessed.

Overall results on the best-evidence synthesis (online supplemental table S3) revealed positive levels of evidence for three measurement properties (internal consistency, structural validity, intrarater reliability) for 50% or more of the instruments. In particular, strong positive evidence was most frequently found for internal consistency (23 questionnaires, 44% out of 52 questionnaires evaluating this measurement property) and/or structural validity (16 questionnaires, 36% out of 45 questionnaires). Only three coding schemes showed strong positive evidence for inter-rater reliability (Observer OPTION-MCC) (Observing Patient Involvement scale for Multiple Chronic Conditions), hypothesis testing (SDM Process Scale (Observer)) and content validity (Observer-based measurement of SDM in web-based

Table 2 Levels of evidence for the best-evidence synthesis^{28 30}

Level of evidence	Rating	Criteria
Strong	+++ or ---	Consistent findings in multiple studies of good methodological quality OR one study of excellent methodological quality
Moderate	++ or --	Consistent findings in multiple studies of fair methodological quality OR one study of good methodological quality
Limited	+ or –	One study of fair methodological quality
Conflicting	±	Conflicting findings
Unknown	?	Only studies of poor methodological quality
Note: a plus sign (+) indicates positive results for a measurement property evaluation and a minus sign (–) indicates negative results for a measurement property evaluation, eg, + stands for limited evidence for positive results and --- stands for strong evidence for negative results for a measurement property.		

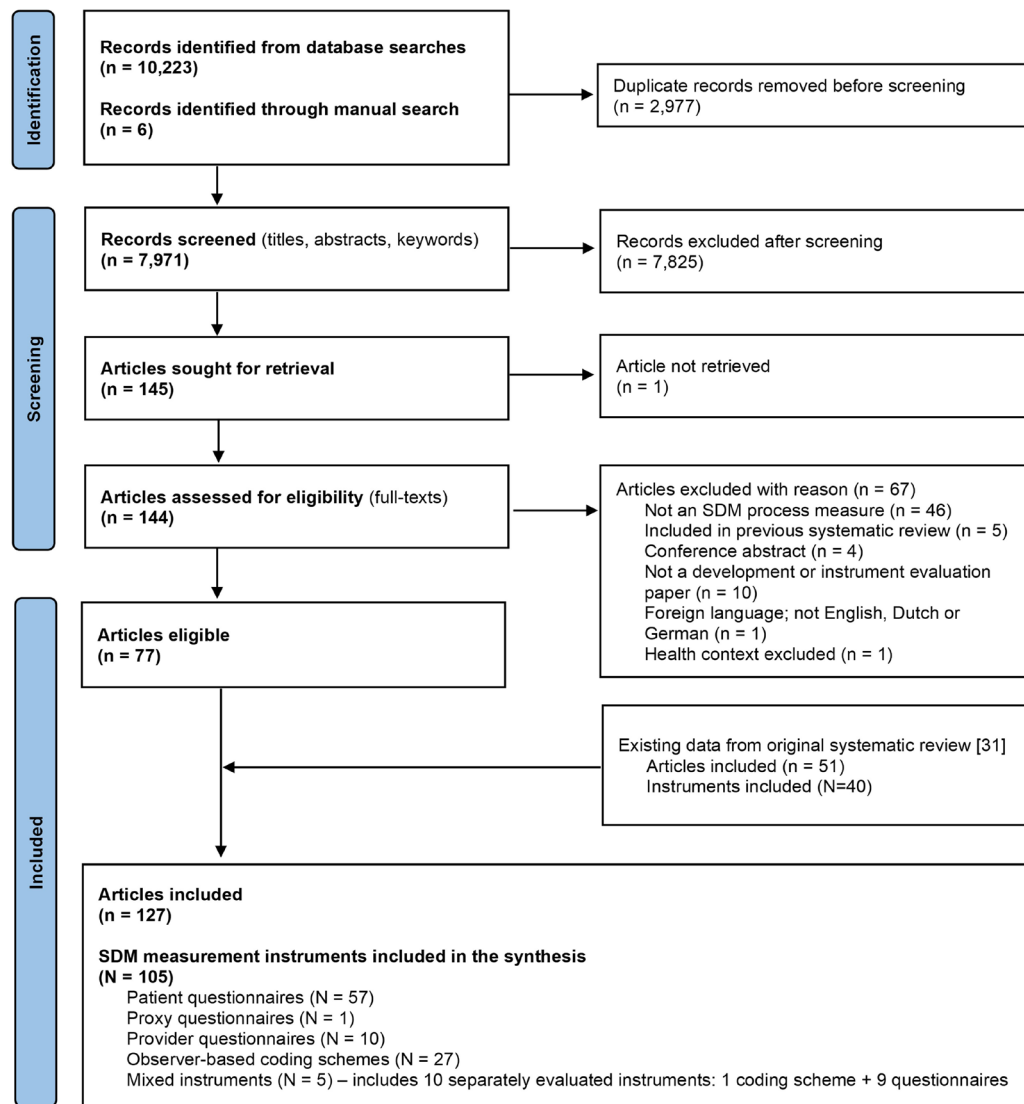


Figure 1 PRISMA flow diagram. PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; SDM, shared decision-making.

pharmaceutical care consultation). Four measurement properties evaluated showed negative (test-retest reliability) or unknown (internal consistency and intra-rater reliability for coding schemes, content validity for questionnaires) levels of evidence for 50% or more instruments. There is inconclusive evidence about hypothesis testing, and a lack of evidence for the appraisal of cross-cultural validity, measurement error, responsiveness and content validity for coding schemes.

Detailed data extracted for the methodology and evaluated measurement property for each article included in this systematic review update can be found in online supplemental appendix 5, table S5. The overall results of their methodological quality are summarised in online supplemental appendix 5, table S6.

Discussion

This systematic review has provided an up-to-date inventory of instruments to assess the process of SDM in research and clinical care. Overall, a large number of studies and instruments were identified, but no single instrument was found to consistently show good evidence for measurement quality. This means evidence gaps remain among instruments that are widely used in research and clinical practice^{34 35} and are recommended by health policy.³⁶ This

review summarised best available psychometric evidence for 105 instruments (77 questionnaires, 28 coding schemes) using COSMIN quality appraisal data from a previous systematic review²⁷ and this current updated systematic review. Best-evidence synthesis revealed positive results for at least half of the instruments that have investigated internal consistency, structural validity and intra-rater reliability, but there was negative, unknown or a general lack of evidence for the remaining measurement properties.

Results also show a marked increase in instrument development activity and potential multiplicity of SDM instruments. More effort is needed by the research community, funders and journals to agree and coordinate strategies for producing targeted and well-designed research in this area. This can include, for example, prioritising robust supplemental validation of existing SDM instruments while carefully considering the context of use (research vs clinical practice) and addressing key recommendations outlined in the previous systematic review. We therefore recommend an international consensus is established with the remit to prioritise and coordinate future validation of preferred existing instrument(s) and standardised measurement, carefully considering the complexities and requirements of SDM processes within a multi-stakeholder context.

Table 3 Overview of SDM instrument characteristics (N=105)

	n	%*
Measurement perspective		
Patient questionnaires	57	57
Proxy questionnaires	1	1
Provider questionnaires	10	10
Observer-based coding schemes	27	27
Mixed instruments (including nine questionnaires and one coding scheme)	5	5
Instrument version		
Original	45	45
Translation	45	45
Revision	11	11
Target setting		
Generic	69	69
Oncology	5	5
Maternity care	4	4
Psychiatry and mental healthcare	4	4
Paediatric care	4	4
General practice	3	3
Other		
Number of evaluation studies; mode (range)	1 (0–7)	
*Percentages use the denominator of 100 as the five mixed instruments (including nine questionnaires and one coding scheme) are counted as single instrument. This is because their characteristics were extracted for the whole instrument and not split into individual instruments.		
SDM, shared decision-making.		

Results in the context of the previous systematic review and future research

The present study identified more new SDM instruments (n=61) developed over the last 7 years compared with the original review (n=40). Most were representing the patient's perspective (n=48). While only few new patient-reported SDM instruments showed good content validity (EPAT - Experienced Patient-Centeredness Questionnaire, PPQ - Patient Participation Questionnaire, A-MADM - Australian version of the Mother's Autonomy in Decision Making scale/MORi - The Mothers on Respect Index, PCCoc/rheum - Person-Centered Care instrument for outpatient care in rheumatology), most lacked consistent methodological and psychometric quality. Encouragingly, 35 newly translated or revised versions of existing SDM instruments were identified (n=19 in the original review) to enable evaluation of SDM across different healthcare contexts. While enhancing potential for cross-cultural comparisons and broader application of SDM instruments in different cultural settings, cross-cultural validity of translated instruments rarely followed COSMIN guidelines and evidence of their content validity remains largely unknown or untested.

Seven recommendations to focus future research activities were proposed in the 2018 review, including suggestions to improve study design and transparency and explicit recommendations for measurement properties to evaluate. Overall results continue to highlight methodological limitations and incomplete psychometric assessments. This suggests that previous recommendations that were based on articles published until 2017, still apply to-date. These include, in particular: (1) provide transparent measurement conceptualisation by providing a clear and comprehensive explanation of the conceptual framework that underpins the construct to be measured (eg, using guidance by OMERACT - Outcome Measures in Rheumatology³⁷) and specify

any key components/steps of the SDM process (eg, using evidence syntheses by Bomhof-Roordink *et al*¹⁰) that are being assessed, (2) perform content validity analyses prior to further validation and (3) determine minimal important change values to inform the interpretation of change scores in intervention studies. This study reiterates that these recommendations should be implemented immediately in ongoing and future work.

There are areas of improvement since the original review. Findings for structural validity are consistent with those reported in the original systematic review which showed 'excellent' or 'good' methodological quality and most consistently demonstrated strong positive evidence for this measurement property. There have been improvements in the available evidence for inter-rater reliability testing of coding schemes, as well as increased sample sizes in validation studies. Some improvements in methodological quality were demonstrated for internal consistency and test-retest reliability, which previously received 'poor' ratings. A higher number of validation studies have evaluated test-retest reliability of measures, but overall quality of evidence remains 'fair'. The appropriateness of internal consistency testing for SDM instruments remains an important discussion point. Internal consistency (and structural validity) assessments are only relevant in cases of a reflective measurement model, but not for formative measurement models. The key distinction between those two models lies in causality, whereby constructs are either shaped by their indicators or items (formative) or constructs generate or determine them (reflective). Previous work has called for greater clarity and consensus on this issue and argued that assuming a formative measurement model is most appropriate to assess the process of SDM,^{27 38–40} few articles included in this review clearly defined the construct underlying their measurement instrument leading to uncertainty about the measurement model and what psychometric properties are relevant to assess. Authors ought to clearly define the construct and explicitly state the underlying measurement model to explain the choice of assessing a measurement property.^{25 38} In addition, results for test-retest reliability, particularly in the context of SDM instruments measuring the patients perspective, warrant careful interpretation. This form of reliability assumes stability of perceptions during the test interval which may not apply due to continued cognitive deliberation.^{41 42} Included validation studies often used long test-retest intervals and thus may reflect actual changes in perception rather than weak reliability.

There are also areas where evidence remained unchanged or weakened compared with the original review. Results showed no improvement for hypothesis testing because studies often lacked robust, a priori formulated hypotheses. There was weaker evidence for content validity of questionnaires which received 'good' and 'poor' methodological quality ratings for >50% of studies in the original review, but no 'good' or 'excellent' rating in >50% of studies in the current review.

Strengths and limitations of the review

Rigorous methodology was applied using published methods following COSMIN quality standards. Particular strengths included an extensive search strategy and broad inclusion criteria to identify wide-ranging instruments assessing SDM process to provide a comprehensive inventory of SDM instruments. A robust quality assurance process ensured that all quality assessment steps were completed by second/third independent reviewers and a comprehensive and up-to-date overview of methodological and psychometric evidence of SDM instruments by meta-synthesising all published data from past and present reviews.

There were some limitations. We aligned methods described in the original review to ensure methodological consistency across the reviews. In particular, we adopted the measurement property definitions reported in the original review and applied scoring criteria using the 2011 version of the COMSIN checklist to facilitate meta-synthesis. These have now been updated by consensus definitions²⁶ and more recent iterations of the checklist (V.2.0, published in 2024),⁴³ the use of which may have affected the ratings. We do, however, believe that the impact of the new checklist is minor and overall conclusions of the review would remain unchanged. A further limitation is that only studies published in English, German or Dutch were included, consistent with the original review eligibility criteria. This may have introduced language bias, however, as only one Spanish article was excluded during eligibility screening, the impact is likely minimal.

Implications

Accurate, meaningful and reliable measurement of SDM is necessary to inform evidence-based health policy and practice and effectively assess and implement high-quality SDM in systems. Variation in conceptualising/measuring SDM and a lack of high-quality instruments leads to ongoing challenges for research and service improvement.

Uncertainty in SDM measurement continues to limit the ability to synthesise evidence and draw conclusions about the effectiveness of interventions intended to support SDM. Although SDM is best understood as a *process* and not an *outcome*, consistent and appropriate measurement is essential to evaluating implementation of SDM interventions. Successful interventions to improve the use of SDM are often complex, comprising multiple components targeting different actors with varying needs/priorities across multiple levels within a healthcare system.^{16 18 19 44–47} Meta-analytical evidence revealed a lack of consistent evidence about what intervention components effectively support the SDM process in routine clinical practice^{20 48} and how to best measure their effectiveness across different contexts¹⁷ and taking into account different perspectives.^{46 47} The absence of standardised measurement ultimately contributes to persistent gaps in understanding the impact of SDM on patient and health service outcomes. To address this, future research should prioritise the use of consensus-based, validated measures and work towards the development of a core outcome/measurement set for SDM. This would strengthen the evidence base for SDM and support organisations and policymakers prioritise and select implementation of the most effective SDM interventions.

Routine collection of SDM data has become an important driver for quality improvement initiatives. Health services use these data to identify gaps in practice, inform clinician training and benchmark performance. Current health initiatives promoting and incentivising the implementation and evaluation of SDM, however, may not be underpinned by robust measurement and may not consider complexities of varying SDM priorities.⁴⁶ While some instruments are more commonly used or even recommended by health policy,³⁶ this reflects pragmatic choices rather than definitive evidence of superiority in measurement quality. This raises questions about whether accurate care gaps are identified and hampers the assessment of whether personalised care is realised.^{17 20} An updated inventory of instruments measuring the process of SDM will provide an important resource to inform policy and guide care improvements. The current state of the literature, however, has not sufficiently evolved to recommend the most high-quality instruments that can reliably evaluate the impact of SDM implemented in routine clinical practice.

Conclusions

There has been substantial research activity to develop and evaluate measurement instruments to assess the process of SDM, resulting in many entirely new or translated versions. While best evidence synthesis revealed some SDM instruments showed strengths in certain measurement properties, none demonstrated consistently positive evidence across all relevant properties evaluated. Evidence gaps for many measurement properties remain. There is a need for prioritisation through international interest holder consensus to focus future research activities on key measurement instruments to strengthen their psychometric evidence base and performance.

Author affiliations

¹National Institute for Health and Care Research Bristol Biomedical Research Centre, Bristol Centre for Surgical Research, Bristol Medical School: Population Health Sciences, University of Bristol, Bristol, UK

²NHS North Bristol NHS Trust, Bristol, UK

³Bradford Institute for Health Research, Bradford Teaching Hospitals NHS Foundation Trust, Bradford, UK

⁴Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands

⁵Leeds Unit of Complex Intervention Development (LUCID), Leeds Institute of Health Sciences, University of Leeds School of Medicine, Leeds, UK

⁶Division of Population Medicine, Cardiff University, Cardiff, UK

⁷Patient representative, Bristol, UK

⁸Centre for Academic Primary Care, Bristol Medical School: Population Health Sciences, University of Bristol, Bristol, UK

Acknowledgements We would like to acknowledge the following individuals: We thank the coauthors of the original review, Fania R Gärtner, Hanna Bomhof-Roordink, Ian P Smith, Isabelle Scholl and Anne M Stiggelbout, whose earlier work provided the basis for this update. We are also grateful to Rhiannon Macefield for her guidance about psychometric properties, and to Paul Kilgariff for his help with generating automated queries in Excel to assist data analysis.

Contributors Conceptualisation: AGKM and JB. Data Curation: CH, SL, VG and AP. Formal analysis: CH, SL, VG and AP. Funding acquisition: AGKM, JB, CH, JH, KA, VS, DH, CC and SH. Investigation: CH, SL, VG, RK, NJ-W, DW and AGKM. Project administration: CH, SL and VG. Resources: AP. Supervision: AGKM. Validation: AGKM, CH, SL, VG, AP and HB. Visualisation: CH, SL, VG and AP. Writing—original draft: CH, AGKM, AP and HB. Writing—review and editing: all authors. AGKM is the guarantor.

Funding This project is funded by the National Institute for Health and Care Research (NIHR) Programme Development Grant (Award ID: NIHR205174) and is supported by the NIHR Biomedical Research Centre (BRC) at the University Hospitals Bristol and Weston NHS Foundation Trust and the University of Bristol (BRC-1215-20011). SL is funded by a NIHR Doctoral Research Fellowship (NIHR303276). VG is funded by an NIHR Academic Clinical Fellowship (NIHR ACF-2023-25-006).

Competing interests None declared.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

ORCID iDs

Christin Hoffmann <https://orcid.org/0000-0002-6293-3813>

Arwen Pieterse <https://orcid.org/0000-0001-6395-0052>

Natalie Joseph-Williams <https://orcid.org/0000-0002-8944-2969>

References

- Carmona C, Crutwell J, Burnham M, *et al*. Shared decision-making: summary of NICE guidance. *BMJ* 2021;n1430.
- Barry MJ, Edgman-Levitan S. Shared decision making--pinnacle of patient-centered care. *N Engl J Med* 2012;366:780–1.
- Bravo P, Härter M, McCaffery K, *et al*. Editorial: 20 years after the start of international Shared Decision-Making activities: Is it time to celebrate? Probably.... *Z Evid Fortbild Qual Gesundheitswes* 2022;171:1–4.
- Tracy MC, Thompson R, Muscat DM, *et al*. Implementing shared decision-making in Australia. *Z Evid Fortbild Qual Gesundheitswes* 2022;171:15–21.
- Coulter A, Edwards A, Elwyn G, *et al*. Implementing shared decision making in the UK. *Z Evid Fortbild Qual Gesundheitswes* 2011;105:300–4.
- Callon W, Beach MC, Links AR, *et al*. An expanded framework to define and measure shared decision-making in dialogue: A “top-down” and “bottom-up” approach. *Patient Educ Couns* 2018;101:1368–77.
- Sepucha KR, Scholl I. Measuring shared decision making: a review of constructs, measures, and opportunities for cardiovascular care. *Circ Cardiovasc Qual Outcomes* 2014;7:620–6.
- Barr PJ, Elwyn G. Measurement challenges in shared decision making: putting the ‘patient’ in patient-reported measures. *Health Expect* 2016;19:993–1001.
- Stiggelbout AM, Pieterse AH, De Haes J. Shared decision making: Concepts, evidence, and practice. *Patient Educ Couns* 2015;98:1172–9.
- Bomhof-Roordink H, Gärtner FR, Stiggelbout AM, *et al*. Key components of shared decision making models: a systematic review. *BMJ Open* 2019;9:e031763.
- Bekker HL, Winterbottom AE, Gavaruzzi T, *et al*. Decision aids to assist patients and professionals in choosing the right treatment for kidney failure. *Clin Kidney J* 2023;16:i20–38.
- Matlock DD, Scherer L. Shared Decision Making “Ought” to Be Done, but Definitions Need Simplicity: Response to “Reframing SDM Using Implementation Science: SDM Is the Intervention”. *Med Decis Making* 2024;44:865–6.
- Fattori F, O'Donnell D, Rodríguez-Martín B, *et al*. Which instruments are used to measure shared, supported and assisted healthcare decision-making between patients who have limited, impaired or fluctuating capacity, their family carers and healthcare professionals? A systematic review protocol. *HRB Open Res* 2019;2:19.
- Australian Commission on Safety and Quality in Health Care. National safety and quality health service standards (second edition). 2021. Available: <https://www.safetyandquality.gov.au/publications-and-resources/resource-library/national-safety-and-quality-health-service-standards-second-edition> [Accessed 10 Sep 2024].
- National Health and Medical Research Council. General guidelines for medical practitioners on providing information to patients. 2017. Available: <https://webarchive.nla.gov.au/awa/20170816072344/https://www.nhmrc.gov.au/guidelines-publications/e57> [Accessed 14 Mar 2025].
- National Institute for Health and Care Excellence. Shared decision making. NICE guideline [NG197]. 2021. Available: <https://www.nice.org.uk/guidance/ng197/> [Accessed 18 Mar 2025].
- National Institute for Health and Care Excellence. Shared decision making [b] evidence review for interventions to support effective shared decision making. 2021. Available: <https://www.nice.org.uk/guidance/ng197/evidence/b-interventions-to-support-effective-shared-decision-making-pdf-9142344255> [Accessed 26 Jan 2023].
- NHS Rightcare. Your health – your decision. London, 2013. Available: <https://aqua.nhs.uk/wp-content/uploads/2021/10/Your-Health-Your-Decision-Evaluation-Report.pdf>
- Joseph-Williams N, Lloyd A, Edwards A, *et al*. Implementing shared decision making in the NHS: lessons from the MAGIC programme. *BMJ* 2017;357:j1744.
- Légaré F, Adekpedjou R, Stacey D, *et al*. Interventions for increasing the use of shared decision making by healthcare professionals. *Cochrane Database Syst Rev* 2018;7:CD006732.
- Shay LA, Lafata JE. Where Is the Evidence? A Systematic Review of Shared Decision Making and Patient Outcomes. *Med Decis Making* 2015;35:114–31.
- Hauser K, Koerfer A, Kuhr K, *et al*. Outcome-Relevant Effects of Shared Decision Making. *Dtsch Arztebl Int* 2015;112:665–71.
- Scholl I, Koelewijn-van Loon M, Sepucha K, *et al*. Measurement of shared decision making – a review of instruments. *Z Evid Fortbild Qual Gesundheitswes* 2011;105:313–24.
- Simon D, Loh A, Härter M. Measuring (shared) decision-making – a review of psychometric instruments. *Z Arztl Fortbild Qualitatssich* 2007;101:259–67.
- Terwee CB, Prinsen CAC, Chiarotto A, *et al*. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res* 2018;27:1159–70.
- Mokkink LB, Terwee CB, Patrick DL, *et al*. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737–45.
- Gärtner FR, Bomhof-Roordink H, Smith IP, *et al*. The quality of instruments to assess the process of shared decision making: A systematic review. *PLoS ONE* 2018;13:e0191747.
- Mokkink LB, Terwee CB, Patrick DL, *et al*. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010;19:539–49.
- Terwee CB, Mokkink LB, Knol DL, *et al*. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012;21:651–7.
- Terwee CB, Bot SDM, de Boer MR, *et al*. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.
- Schellingerhout JM, Heymans MW, Verhagen AP, *et al*. Measurement properties of translated versions of neck-specific questionnaires: a systematic review. *BMC Med Res Methodol* 2011;11:1.
- Schellingerhout JM, Verhagen AP, Heymans MW, *et al*. Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic review. *Qual Life Res* 2012;21:659–70.

- 33 Martínez Ques AA, Braña Marcos B, Martín Arribas C, *et al.* Diseño y validación de un instrumento sobre calidad de la planificación anticipada de decisiones para profesionales. *Gac Sanit* 2022;36:401–8.
- 34 Doherr H, Christalle E, Kriston L, *et al.* Use of the 9-item Shared Decision Making Questionnaire (SDM-Q-9 and SDM-Q-Doc) in intervention studies—A systematic review. *PLoS ONE* 2017;12:e0173904.
- 35 Echtherhoff J, Petter J, Scholl I, *et al.* Measurement properties of the 9-item shared decision making questionnaire (sdm-q-9): a systematic psychometric review and meta-analysis. *PsyArXiv* [Preprint].
- 36 NHS England. Commissioning for quality and innovation (CQUIN): 2022/23. 2022. Available: <https://www.england.nhs.uk/nhs-standard-contract/cquin/2022-23-cquin/> [Accessed 26 Jan 2023].
- 37 Boers M, Beaton DE, Shea BJ, *et al.* OMERACT Filter 2.1: Elaboration of the Conceptual Framework for Outcome Measurement in Health Intervention Studies. *J Rheumatol* 2019;46:1021–7.
- 38 Wollschläger D. Short communication: Where is SDM at home? putting theoretical constraints on the way shared decision making is measured. *Z Evid Fortbild Qual Gesundheitswes* 2012;106:272–4.
- 39 Jarvis CB, MacKenzie SB, Podsakoff PM. A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research. *J Consum Res* 2003;30:199–218.
- 40 Bomhof-Roordink H, Gärtner FR, van Duijn-Bakker N, *et al.* Measuring shared decision making in oncology: Development and first testing of the iSHAREpatient and iSHAREphysician questionnaires. *Health Expect* 2020;23:496–508.
- 41 Bekker HL, Hewison J, Thornton JG. Understanding why decision aids work: linking process with outcome. *Patient Educ Couns* 2003;50:323–9.
- 42 Gopalan V, Hoffmann C, Bekker HL, *et al.* Reliability testing of patient-reported measurement instruments of shared decision-making within surgical treatment pathways: a mixed-methods study. *medRxiv* [Preprint] 2025.
- 43 Mokkink LB, Elsmann EBM, Terwee CB. COSMIN guideline for systematic reviews of patient-reported outcome measures version 2.0. *Qual Life Res* 2024;33:2929–39.
- 44 Härter M, van der Weijden T, Elwyn G. Policy and practice developments in the implementation of shared decision making: an international perspective. *Z Evid Fortbild Qual Gesundheitswes* 2011;105:229–33.
- 45 Stolz-Klingenberg C, Bünzen C, Coors M, *et al.* Comprehensive Implementation of Shared Decision Making in a Neuromedical Center Using the SHARE TO CARE Program. *Patient Prefer Adherence* 2023;17:131–9.
- 46 Toft BS, Ellegaard T, Nielsen BK, *et al.* Health service provider views on measuring patient involvement in healthcare: an interview study with researchers, clinicians, service managers, and policymakers. *BMC Health Serv Res* 2024;24.
- 47 Toft BS, Rodkjaer L, Andersen AB, *et al.* Measures used to assess interventions for increasing patient involvement in Danish healthcare setting: a rapid review. *BMJ Open* 2022;12:e064067.
- 48 Stacey D, Bennett CL, Barry MJ, *et al.* Decision aids for people facing health treatment or screening decisions. *Cochrane Database Syst Rev* 2011;CD001431.