



Are We Improving? Update and Critical Appraisal of the Reporting of Decision Process and Quality Measures in Trials Evaluating Patient Decision Aids

Logan Trenaman , Jesse Jansen, Jennifer Blumenthal-Barby, Mirjam Körner, Joanne Lally, Daniel D. Matlock , Lilisbeth Perestelo-Perez, Mary Ropka, Christine Stirling, Kathrene Valentine , Ha Vo, Celia E. Wills, Richard Thomson, and Karen Sepucha 

Background. In 2014, a systematic review found large gaps in the quality of reporting of measures used in 86 published trials evaluating the effectiveness of patient decision aids (PtDAs). The purpose of this study was to update that review. **Methods.** We examined measures of decision making used in 49 randomized controlled trials included in the 2014 and 2017 Cochrane Collaboration systematic review of PtDAs. Data on development of the measures, reliability, validity, responsiveness, precision, interpretability, feasibility, and acceptability were independently abstracted by 2 paired reviewers. **Results.** Information from 273 measures was abstracted, and 109 of these covered the core domains of decision processes ($n = 55$) and decision quality including informed choice/knowledge ($n = 48$) and values-choice concordance ($n = 12$). Very few studies reported data on the performance and clinical sensibility of measures, with reliability (23%) and validity (6%) being the most common. Studies using new measures were less likely to include information about their psychometric performance compared with previously published measures. **Limitations.** The review was limited to reporting of measures in studies included in the Cochrane review and did not consult prior publications. **Conclusion.** There continues to be very little reported about the development or performance of measures used to evaluate the effectiveness of PtDAs in published trials. Minimum reporting standards have been published, and efforts to require investigators to use them are needed.

Keywords

checklist/standards, decision support techniques, patient-centred care, patient decision aids, shared decision making

Date received: January 11, 2021; accepted: March 19, 2021

Introduction

The International Patient Decision Aid Standards (IPDAS) collaboration recommends that patient decision aids (PtDAs) are evaluated by their impact on 2 core domains: decision process and decision quality.¹ Decision process refers to the extent to which a PtDA helps patients to recognize that a decision needs to be made; feel informed about the options; be clear about what

matters most to them in this decision; discuss goals, concerns, and preferences with their health care providers; and be involved in decision making. Decision quality is the extent to which a patient's eventual choice is

Corresponding Author

Karen Sepucha, Massachusetts General Hospital, 100 Cambridge Street, 16th Floor, Boston, MA 02114, USA
(ksepucha@mgh.harvard.edu).

informed and consistent with their values. There are many different measures available for these constructs, with new ones being developed and tailored for specific PtDAs.^{2,3} To understand the impact of PtDAs, it is important that trials report on the psychometric properties of the measures that are used.

Several studies have highlighted issues with reporting of measures for evaluating PtDAs including variability in definitions, methodology, and validity and generally poor reporting of psychometrics and development.⁴⁻⁷ A 2014 systematic review conducted by several of the authors examined measures used in 86 randomized trials included in the 2011 Cochrane systematic review of PtDAs and found that few provided details on psychometric properties of the individual measures.⁸ This work informed the subsequent development and publication of reporting guidelines for evaluations of PtDAs, the SUN-DAE checklist.⁹

This article updates and extends that previous work by conducting a review of the measures used to evaluate decision making in the new trials added to the 2014 and 2017 Cochrane systematic reviews of PtDAs.^{10,11} We focus on the quality of reporting on the development and performance of the outcomes related to decision process and decision quality as recommended by IPDAS.

Methods

This study updates the previous review and follows a similar approach.⁸ Pairs of reviewers independently reviewed the full-text manuscripts of the 49 new randomized controlled trials included in the 2014 and 2017

Cochrane systematic reviews of PtDAs,^{10,11} determined whether they measured 1 or more of the elements of the “quality of the decision-making process” or “decision quality,” and abstracted information using standard forms. The reviewers collected information on study context, description of the measure(s) and their administration, the development process (item generation, cognitive testing, pilot studies), psychometric performance (reliability, validity, responsiveness), and clinical sensibility (interpretability, feasibility, and acceptability). Table 1 includes some of the abstracted data fields and provides examples of evidence from our past review.⁸ The supplemental file includes details on the studies included in this review and the full data extraction tool.

A measure was considered new if there was no cited prior publication and/or it was not a known, named scale. Articles that cited a reference with respect to any of these issues (e.g., “The Decisional Conflict Scale has been shown to be valid and reliable”¹⁶) were given credit for reporting those elements. However, we did not consult cited sources to confirm that information or obtain additional unreported information. The abstraction was limited to the details provided within the published trial papers, based on how a reader might evaluate the measures as described by the trial authors. Frequent calls with the entire coding group were held throughout the data abstraction process to ensure consistency. Discrepancies between reviewers were initially discussed by the paired reviewers, and most were resolved after discussion. The lead authors (K.S. and R.T.) adjudicated unresolved disagreements. The data abstracted from the studies are available from the corresponding author by request.

Analysis

We classified the measures and assessed the presence of reporting for key elements of measure development, psychometric performance, and clinical sensibility. We examined reporting for measures of knowledge, values-choice concordance and decision process. We did not separate out subelements of the decision process (e.g., feel informed), as most measures included multiple elements and did not report separately.

Results

Of the 49 new trials, 44 (90%) measured at least 1 aspect of decision quality or decision process. Most studies included 1 or more measures of the decision process (78%, 38/49 studies) and knowledge (73%, 36/49

University of British Columbia, Vancouver, BC, Canada (LT); Centre for Health Evaluation and Outcome Sciences, Vancouver, Canada (LT); Maastricht University, Maastricht, The Netherlands (JJ); Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, TX, USA (JB-B); Institute of Medical Psychology and Medical Sociology, Medical Faculty, Albert-Ludwigs- University, Freiburg, Baden-Wuerttemberg, Germany (MK); Population Health Sciences Institute, Baddiley Clark Building, Newcastle University, Newcastle upon Tyne, UK (JL, RT); University of Colorado, Aurora, CO, USA (DDM); VA Eastern Colorado Geriatric Research Education and Clinical Center, Denver, CO, USA (DDM); Evaluation Unit, Canary Islands Health Service, REDISSEC, Tenerife Spain (LP-P); University of Virginia, Charlottesville, VA, USA (MR); School of Nursing, University of Tasmania, Hobart, Tasmania, Australia (CS); Harvard Medical School, Boston, MA, USA (KV, KS); Massachusetts General Hospital, Boston, MA, USA (KV, HV, KS); Ohio State University, Columbus, OH, USA (CEW). The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: LT, JJ, JBB, MK, JL, DM, LPP, MR, KS, KDV, HV, CEW, RT, and KRS have no financial conflicts of interest to report. The authors received no financial support for the research, authorship, and/or publication of this article.

Table 1 Elements Abstracted Regarding Measure Development and Psychometric Performance^a

Measure Development		Examples of Evidence from Past Review
Item generation	How were content items developed and by whom?	Showing item generation, pilot study, reliability, validity and interpretability: “First, we determined whether subjects were better informed through a twenty question test of BPH knowledge . . . developed by a panel including a general internist, a urologist, a survey researcher, and a lawyer with a special interest in informed consent. Correct responses were scored + 1, incorrect responses – 1, and “not sure” responses were scored 0 (total range – 20 to + 20). Validation of new outcome measures Cronbach’s alpha statistic for the items testing BPH knowledge was 0.68. The criterion validity of this test was assessed by comparing scores for a convenience sample of 12 urologic nurses with the scores of the 167 BPH patients enrolled in the baseline period. The nurses had a mean score of 14.8 [out of 20], compared to 5.6 for the patients ($p < 0.001$). Nurses answered an average of 85% of the questions correctly, compared to 48% for the patients ($p < 0.001$). Furthermore, a modest correlation between these patients’ knowledge scores and their educational levels was seen, $r = 0.23$ ($p < 0.001$).” ¹²
Cognitive testing	Was the measure tested for understandability before use?	
Pilot studies	Were pilot studies (of any type) conducted to pre-test the measure?	
Measure performance		
Reliability ^b	Were appropriate assessments of the reliability of the measure reported? If so, was there evidence of adequate reliability?	Showing reliability, validity and responsiveness: “The decisional conflict scale measured patients’ uncertainty about which therapy to choose, modifiable factors contributing to uncertainty (believing themselves to be uninformed, unclear about values, and unsupported in decision making), and perceived effective decision making. The scale is reliable, discriminates between those who make or delay decisions, is responsive to change, and discriminates between different decision-supporting interventions. Two items were added to elicit patients’ perceptions that they were informed about the benefits and risks of warfarin and, separately, about benefits and risks of aspirin. This did not affect the scale’s reliability in this study (Cronbach $\alpha = .92$).” ¹³
Validity ^c	Were appropriate assessments of the validity of the measure reported? If so, was there evidence of adequate validity?	
Responsiveness	Is there evidence that the measure is sensitive to changes of importance to patients and clinicians?	
Clinical sensibility		
Interpretability	Are the scores meaningful to clinicians and patients?	Showing interpretability: A score of 25 out of 100 is “associated with implementing decisions,” and a score of 37.5 out of 100 is “associated with decision delay or feeling unsure about implementation.” ¹⁴
Acceptability ^d	Does the measure appear to be acceptable to respondents?	Showing pilot testing and acceptability: Showing interpretability: “In the literature, assessment of values has primarily been measured with probability-based risk-benefit trade-offs. We pretested these items in focus groups ($k = 1$; $n = 15$) and found them unacceptable to a majority of men. Therefore, we developed items to assess the personal importance or relative worth of the advantages and limitations of screening, based on focus groups themes and published literature. Further information about scale development is available elsewhere.” ¹⁵
Feasibility of administration	Are there indicators of the appropriateness of effort, burden, or disruption (of clinical or research team) required to administer and score the measure?	

^aAdapted from Sepucha et al.⁸^bIncludes internal consistency reliability (e.g., Cronbach’s alpha, Kuder-Richardson coefficient), test-retest reliability, and interrater reliability (e.g., percentage agreement, Kappa coefficient; intraclass correlation coefficient).^cIncludes content validity (e.g., Content Validity Index), criterion-related validity (e.g., correlations to demonstrate concurrent, predictive validity), construct validity (e.g., factor analysis to demonstrate predicted convergence/divergence of constructs and/or structural invariance of the measure, discriminant analysis, known groups analysis)^dCould be inferred from patterns of missing data or low response rates.

Table 2 Reporting on Performance of New and Established Measures of Decision Quality and Decision Process in Studies of PtDAs

	Core Outcomes (<i>n</i> = 109)							
	Decision Process (<i>n</i> = 55)		Decision Quality				Other Decision Outcomes (<i>n</i> = 164)	
	<i>n/N</i>	%	Knowledge ^a (<i>n</i> = 48)		Concordance (<i>n</i> = 12)		<i>n/N</i>	%
			<i>n/N</i>	%	<i>n/N</i>	%		
Previously published	50/55	91	28/48	57	5/12	42	75/164	46
Measure development								
Development process	1/55	2	7/48	15	2/12	17	10/164	6
Item generation	1/55	2	6/48	13	1/12	8	6/164	4 ^b
Cognitive testing	0/55	0	4/48	8	0/2	0	4/164	2 ^b
Pilot studies	0/55	0	4/48	8	0/2	0	3/164	2 ^b
Measure performance								
Reliability	18/55	33	6/48	13	1/12	8	19/164	12
Validity	4/55	7	1/48	2	1/12	8	5/164	2
Responsiveness	1/55	2	0/48	0	0/12	0	0/164	0
Clinical sensibility								
Interpretability	3/55	5	2/48	4	0/12	0	0/164	0
Acceptability	5/55	9	1/48	2	1/12	8	5/164	3
Feasibility of administration	0/55	0	1/48	2	1/12	8	1/164	1

^aSix measures covered both knowledge and concordance.

^bMissing data on development process for *n* = 3.

studies), whereas only a minority measured values-choice concordance (24%, 12/49 studies).

We abstracted 273 reported measures related to decision making. Of these, 109 covered 1 or more core constructs of the decision process (*n* = 55) or decision quality, including knowledge (*n* = 48) or values-choice concordance (*n* = 12; Table 2). Of note, 6 measures covered both knowledge and concordance. The most common other type of outcomes included actual choice (*n* = 40), preference or preferred choice (*n* = 25), satisfaction with decision making or chosen option (*n* = 17), depression and/or anxiety (*n* = 14), adherence (*n* = 8), and decision regret (*n* = 7).

Studies included very limited information on psychometric properties of the measures (*n* = 109) such as reliability (23%), validity (6%), and responsiveness (1%). Studies rarely assessed the clinical sensibility of the measures, such as feasibility (2%), acceptability (7%), and interpretability (5%).

Whereas most decision process measures used and cited a previously published measure (50/55), many knowledge (20/48) and concordance measures (7/12) did not cite an existing measure. Few studies using new measures provided information on the development process (4/31, 13%) or psychometric properties (6/31, 19%). Previously published measures were significantly more likely

to have some reporting of psychometrics (41% v. 19%, $\chi^2 = 0.04$).

Discussion

Decision process and quality measures are critical to evaluating the effectiveness of PtDAs.¹ This brief report updates a previous review,⁸ summarizing new evidence on the quality of reporting of measures of decision process and quality captured by 49 new studies included in the 2014 and 2017 Cochrane Collaboration's reviews of PtDAs. This review finds continued shortcomings in the reporting of the development, performance, and clinical sensibility of decision process and quality measures used in published trials.

Reporting of the development process for new measures was poor. Generally speaking, previously published measures were more likely to have some reporting of psychometrics than new measures (41% v. 19%); however, this largely reflects strong reporting of the Decision Conflict Scale (DCS).¹⁷ The DCS was used in more than half of the trials (72/135, 53%), often with detailed descriptions of performance.

Most new trials include decision-making evaluation measures (90%), which is similar to the previous review (88%).⁸ Reliability reporting was also similar (23% v.

21%), whereas validity was worse (6% v. 16%) in these new studies. A common misperception is that validity and reliability are properties of the survey instrument, when in reality they are properties of data and the interpretation of the data (which includes understanding the administration, setting, sample, and analysis procedures).¹⁸ This underscores the importance of reporting relevant information on psychometric performance for each study and each use of an instrument or measure. Detailed reporting of psychometric properties is important to allow appropriate interpretation of results, improve our understanding of the impact of PtDAs on decision process and outcomes, and support replication and synthesis of findings.¹⁹ There are many great resources that describe how to assess the adequacy of psychometric evidence, with the authors recommending a text by Waltz et al.²⁰ The SUNDAE checklist was developed in 2018 to support completeness and transparency of reporting of PtDA evaluation studies, including psychometric properties of the measures used.⁹ While the checklist did not affect this update, which included trials published up to 2017, it may improve reporting in future, particularly if journals adopt the SUNDAE checklist.

Few studies include details on the clinical sensibility of the measures. This information is important to allow appropriate interpretation of the results and to support successful implementation of trials into routine clinical practice. Patient-reported measures provide insight into the outcomes and experience of care from the patients' perspective and are valuable to monitor quality of care and outcomes.^{21–24} However, without information on the acceptability, feasibility, and interpretability of the measures, their implementation into practice may be hindered.

Our study has several limitations. First, we focused on randomized controlled trials included within the Cochrane review, although we would expect these to be the highest-quality evaluations. Second, we did not review the cited sources of previously published measures; hence, our findings reflect only the quality of the reporting of measures not the quality of the measures themselves. Third, it is possible that developers reported more details about the measures elsewhere, and this would not have been captured in our review.

Several questions remain to be answered. What other measures should be used to evaluate PtDAs, if any (e.g., health outcomes, cost-effectiveness, potential harms), and when should they be measured? What components of PtDAs are core to effectiveness? Are different measures needed for disadvantaged patients (e.g., individuals with low literacy or low incomes)? Increasingly, 1 or more options in situations covered in PtDAs involve a

large behavior change component (e.g., surgery versus diet and exercise for obesity/weight management). In what ways does this behavior change component change our strategies (if at all) for the evaluation of PtDAs (e.g., do we need to assess levels of self-efficacy and motivation in addition to knowledge and concordance)? How do we support decisions in which an option is considered of low value (e.g., prostate-specific antigen screening for certain groups)?

There are also theoretical issues. A growing body of research suggests that defining what a good medical decision is, and how to measure it, is more complicated than is often assumed in theoretical decision-making frameworks.²⁵ For example, real-life decision making is influenced by interpersonal factors, structural constraints, and affect/emotions. It provides an argument for consideration of how these factors (and others) contribute to the definition of good medical decision making and a tailored approach to the measurement of decision quality.

There continues to be very little reported about the development or performance of measures used to evaluate the effectiveness of PtDAs within published trials. Minimum reporting standards (SUNDAE) have been published, and wide use should be promoted to support transparent and accurate reporting and clearer interpretation of the outcomes of PtDA trials.

Acknowledgment

The authors would like to acknowledge colleagues who contributed to earlier versions of the IPDAS chapter and the IPDAS steering committee.

ORCID iDs

Logan Trenaman  <https://orcid.org/0000-0002-6399-5741>
 Daniel D. Matlock  <https://orcid.org/0000-0001-9597-9642>
 Kathrene Valentine  <https://orcid.org/0000-0001-6349-5395>
 Karen Sepucha  <https://orcid.org/0000-0002-3762-3880>

Supplemental Material

Supplementary material for this article is available on the *Medical Decision Making* website at <http://journals.sagepub.com/home/mdm>.

References

1. Sepucha KR, Borkhoff CM, Lally J, et al. Establishing the effectiveness of patient decision aids: key constructs and measurement instruments. *BMC Med Inform Decision Making*. 2013;13(suppl 2):S12.
2. Sepucha KR, Stacey D, Clay CF, et al. Decision quality instrument for treatment of hip and knee osteoarthritis: a

- psychometric evaluation. *BMC Musculoskelet Disord*. 2011;12:149.
3. Lee CN, Wetschler MH, Chang Y, et al. Measuring decision quality: psychometric evaluation of a new instrument for breast cancer chemotherapy. *BMC Med Inform Decis Mak*. 2014;14:73.
 4. Kryworuchko J, Stacey D, Bennett C, Graham ID. Appraisal of primary outcome measures used in trials of patient decision support. *Patient Educ Couns*. 2008;73(3):497–503.
 5. Sepucha K, Ozanne EM. How to define and measure concordance between patients' preferences and medical treatments: a systematic review of approaches and recommendations for standardization. *Patient Educ Couns*. 2010;78(1):12–23.
 6. Scholl I, Loon MK, Sepucha K, et al. Measurement of shared decision making—a review of instruments. *Z Evid Fortbild Qual Gesundwes*. 2011;105(4):313–24.
 7. Gärtner FR, Bomhof-Roordink H, Smith IP, Scholl I, Stiggelbout AM, Pieterse AH. The quality of instruments to assess the process of shared decision making: a systematic review. *PLoS One*. 2018;13(2):e0191747.
 8. Sepucha KR, Matlock DD, Wills CE, et al. “It’s valid and reliable” is not enough: critical appraisal of reporting of measures in trials evaluating patient decision aids. *Med Decis Making*. 2014;34(5):560–6.
 9. Sepucha KR, Abhyankar P, Hoffman AS, et al. Standards for UNiversal reporting of patient Decision Aid Evaluation studies: the development of SUNDAE Checklist. 2018;27(5):380–8.
 10. Stacey D, Légaré F, Col NF, et al. Decision aids for people facing health treatment or screening decisions. *Cochrane Database Syst Rev*. 2014;(1):CD001431.
 11. Stacey D, Légaré F, Lewis K, et al. Decision aids for people facing health treatment or screening decisions. *Cochrane Database Syst Rev*. 2017;4(4):CD001431.
 12. Barry M, Cherkin DC. A randomized trial of a multimedia shared decision-making program for men facing a treatment decision for benign prostatic hyperplasia. *Dis Manage Clin Outcomes*. 1997;1(1):5–14.
 13. Man-Son-Hing M, Laupacis A, O'Connor AM, et al. A patient decision aid regarding antithrombotic therapy for stroke prevention in atrial fibrillation: a randomized controlled trial. *JAMA*. 1999;282(8):737–43.
 14. de Achaval S, Fraenkel L, Volk RJ, Cox V, Suarez-Almazor ME. Impact of educational and patient decision aids on decisional conflict associated with total knee arthroplasty. *Arthritis Care Res (Hoboken)*. 2012;64(2):229–37.
 15. Allen JD, Othus MKD, Hart A, et al. A randomized trial of a computer-tailored decision aid to improve prostate cancer screening decisions: results from the Take the Wheel trial. *Cancer Epidemiol Biomarkers Prev*. 2010;19(9):2172–86.
 16. O'Connor AM, Tugwell P, Wells GA, et al. Randomized trial of a portable, self-administered decision aid for postmenopausal women considering long-term preventive hormone therapy. *Medical Decision Making*. 1998;18(3):295–303.
 17. O'Connor AM. Validation of a decisional conflict scale. *Med Decis Making*. 1995;15(1):25–30.
 18. Messick S. Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *Am Psychol*. 1995;50(9):741–9.
 19. Volk RJ, Coulter A. Advancing the science of patient decision aids through reporting guidelines. *BMJ Qual Saf*. 2018;27(5):337–9.
 20. Waltz C, Strickland OL, Lenz E. *Measurement in Nursing and Health Research*. New York: Springer; 2016.
 21. Basch E. Patient-reported outcomes—harnessing patients' voices to improve clinical care. *N Engl J Med*. 2017;376(2):105–8.
 22. Greenhalgh J. The applications of PROs in clinical practice: what are they, do they work, and why? *Qual Life Res*. 2009;18(1):115–23.
 23. Greenhalgh J, Gooding K, Gibbons E, et al. How do patient reported outcome measures (PROMs) support clinician-patient communication and patient care? A realist synthesis. *J Patient Rep Outcomes*. 2018;2(1):42.
 24. Greenhalgh J, Dalkin S, Gibbons E, et al. How do aggregated patient-reported outcome measures data stimulate health care improvement? A realist synthesis. *J Health Serv Res Policy*. 2018;23(1):57–65.
 25. Hamilton JG, Lillie SE, Alden DL, et al. What is a good medical decision? A research agenda guided by perspectives from multiple stakeholders. *J Behav Med*. 2017;40(1):52–68.