



Published in final edited form as:

*Med Decis Making*. 2014 July ; 34(5): 560–566. doi:10.1177/0272989X14528381.

## “It's valid and reliable” is not enough: critical appraisal of reporting of measures in trials evaluating patient decision aids

Karen Sepucha, PhD<sup>1</sup>, Daniel D. Matlock, MD, MPH<sup>2</sup>, Celia E. Wills, PhD, RN<sup>3</sup>, Mary Ropka, PhD, RN<sup>4</sup>, Natalie Joseph-Williams, GDipPsych<sup>5</sup>, Dawn Stacey, RN, PhD<sup>6</sup>, ChirkJenn Ng, MBBS, PhD<sup>7</sup>, Carrie Levin, PhD<sup>9</sup>, Joanne Lally<sup>9</sup>, Cornelia M. Borkhoff<sup>10</sup>, and Richard Thomson, MD<sup>9</sup>

<sup>1</sup>General Medicine Division, Massachusetts General Hospital and Harvard Medical School

<sup>2</sup>University of Colorado School of Medicine

<sup>3</sup>The Ohio State University

<sup>4</sup>Department of Public Health Sciences, University of Virginia School of Medicine

<sup>5</sup>Cochrane Institute of Primary Care and Public Health, Cardiff University

<sup>6</sup>University of Ottawa and Ottawa Hospital Research Institute, Ottawa, Canada

<sup>7</sup>Department of Primary Care Medicine, University of Malaya

<sup>8</sup>Informed Medical Decisions Foundation

<sup>9</sup>Institute of Health and Society, Newcastle University

<sup>10</sup>University of Toronto and The Hospital for Sick Children

### Abstract

**Background**—This review systematically appraises the quality of reporting of measures used in trials to evaluate the effectiveness of patient decision aids (PtDAs) and presents recommendations for minimum reporting standards.

**Methods**—We reviewed measures of decision quality and decision process in 86 randomized controlled trials (RCTs) from the 2011 Cochrane Collaboration systematic review of PtDAs. Data on development of the measures, reliability, validity, responsiveness, precision, interpretability, feasibility and acceptability were independently abstracted by two reviewers.

**Results**—Information from 178 instances of use of measures was abstracted. Very few studies reported data on the performance of measures, with reliability (21%) and validity (16%) being the most common. Studies using new measures were less likely to include information about their psychometric performance.

**Limitations**—The review was limited to reporting of measures in studies included in the Cochrane review and did not consult prior publications.

---

Corresponding Author And Reprint Requests: Karen Sepucha, PhD, Massachusetts General Hospital, 50 Staniford Street, 9<sup>th</sup> floor, Boston, MA 02114, ksepucha@partners.org.

Meetings Where Paper Was Presented: None

**Conclusion**—There is very little reported about the development or performance of measures used to evaluate the effectiveness of PtDAs in published trials. Minimum reporting standards are proposed to enable authors to prepare study reports, editors and reviewers to evaluate submitted papers, and readers to appraise published studies.

---

## Introduction

The Institute of Medicine (IOM) has identified patient-centered care as central to healthcare quality, and shared decision making is often discussed as a means to promote more patient-centered care (1,2). Patient decision aids (PtDAs) are evidence-based interventions designed to support shared decision making by preparing patients to participate in treatment decisions. Evaluation of patient-centered interventions, such as PtDAs, requires patient-reported measures and it is important that these measures have demonstrated strong psychometric performance.

Useful measures should address appropriate or “meaningful” processes and outcomes that are essential to high quality decision making (3). The International Patient Decision Aids Standards (IPDAS) has recommended two core domains for evaluating the effectiveness of PtDAs: decision quality and decision process (4). Decision quality includes sub domains of decision-specific knowledge, realistic expectations and value concordance (or extent to which treatments match patient's goals). Decision process measures include sub domains of recognition of a decision; feeling informed about options and outcomes; feeling clear about what matters most; discussing the goals of treatment with providers; and being involved in decision making. The Cochrane Collaboration systematic review has found that PtDAs increase decision quality and improve decision process (5).

Previous reviews have evaluated the quality of selected measures used to assess the effectiveness of PtDAs and shared decision making. Kryworuchko et al. (2008) appraised the quality of eight primary outcome measures used in randomized controlled trials (RCTs) evaluating PtDAs and found that only two had evidence of strong psychometric performance (6). Recently, Scholl et al. (2011) appraised the quality of 19 different measures of shared decision making and also found limited reporting on the performance of measures (7). Without clear data on the performance of measurement instruments, it is difficult for investigators to decide on which measures to use and it is difficult to interpret the results of the studies. Further, the lack of information about performance of measures hampers the ability to generate consensus on a core set of shared measures to facilitate knowledge synthesis and theory building in the field.

This paper extends previous work by conducting a comprehensive review of the measures used to evaluate decision quality or decision process in the RCTs in the 2011 Cochrane review of PtDAs, focusing on the quality of *reporting* of their development and performance, in order to propose standards to enable authors to prepare their study reports, editors and reviewers to evaluate submitted papers, and readers to appraise published studies.

## Methods

Two reviewers independently reviewed the full text manuscripts of the 86 RCTs included in the 2011 systematic review of PtDAs (5) and abstracted information using standard forms. The reviewers abstracted information on each time a measure was reported that compared one or more aspects of decision quality or decision process for the intervention and control groups. We collected information on study context, description of the measure(s) and their administration, development process, and psychometric performance. Table 1 includes some of the abstracted data fields on the performance of the measures. The appendix includes the full set of fields. The data abstracted from the studies included in the systematic review are available from the corresponding author by request.

Definitions and examples of items were established prior to data abstraction and regular discussion among co-authors ensured consistency. 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 DCS has been shown to be valid and reliable (O'Connor, 1998),” were given credit for reporting those elements. However, we did not consult cited sources to confirm that information, nor to obtain additional unreported information. Frequent calls with the entire coding group were held throughout the data abstraction process. Discrepancies between reviewers were initially discussed by the two reviewers and the majority were resolved after consulting the full text. Common and unresolved discrepancies were brought for discussion by the entire group, with the lead authors (KS and RT) adjudicating to ensure consistency across studies and resolve any remaining disagreements. For example, we clarified that the reporting of response rates for the overall study did not provide evidence of acceptability of an individual measure used in the study.

## Analysis

We classified the measures and assessed the presence of reporting for key elements of scale development and psychometric performance. We examined reporting for measures of knowledge, values-choice concordance and decision process. We did not separate out specific elements of decision process (e.g. feel informed), as most measures included multiple elements and did not report separately. We hypothesized that new measures would have more substantial reporting of psychometric performance than previously-validated measures; hence, we compared reporting of new to previously published measures. Given that the Decisional Conflict Scale (DCS) was the most extensively used measure, we analyze and report it separately from other measures in the results section (4).

## Results

Of the 86 trials in the 2011 Cochrane review, 76/86 (88%) measured at least one aspect of decision quality or decision process. Most of the remaining ten studies (7/10, 70%) evaluated the impact of decision aid on choices or uptake of tests or treatments.

Across all studies, we abstracted 178 instances where a measure of one or more aspect of decision quality and decision making process was reported. Of the 178, 73/178 (41%) were

related to knowledge and/or realistic expectations, 13/178 (7%) covered value-choice concordance, and 92/178 (52%) covered one or more aspects of decision process. The following results summarize the reporting on the development process for the knowledge, value-choice concordance and decision process measures. Table 2 presents details about the frequency and type of information reported about the 178 abstracted measures, comparing new and established measures.

Sixty studies included 73 instances of use of measures of knowledge and/or realistic expectations. Only two of the knowledge measures were named, hence it was difficult to ascertain whether the same measures were used across studies. About half (41/73, 56%) appeared to be new measures used for the first time in the study. Thirty-five knowledge measure descriptions (48%) included either information about item generation (13/73) or referenced another publication reporting on development of the measure (22/73), and about half (52%) included no information about development of the measure.

Of the thirteen studies that included a measure of values-choice concordance, about half (6/13) were new measures developed for that study. Five studies (5/13, 38%) included information either on item generation in the paper (2/13) or referenced a prior publication for item development (3/13).

Sixty-one studies included 92 instances of measures of decision process. Most (78/92; 85%) were named, established measures while 14/92 (15%) were developed specifically for that study. About half (47/92, 51%) were some version of the DCS. None included data on item generation in the paper itself, although about half (50/92, 54%) referenced a prior publication. Not all studies cited prior work when using a named scale; for example, 15 studies using the DCS did not include a citation for it.

## Discussion

Decision quality and decision process have been identified as core measures to evaluate the effectiveness of PtDAs (4). This review of the 86 RCTs of PtDAs found that reporting on the development and performance of measures was extremely limited. About 1/3 of the studies used new instruments and, contrary to our expectation, the amount reported on new measures was actually less than that reported on existing measures.

For the measures included in the review, reliability and validity were most commonly reported aspects of performance, but only for 21% and 16% of instances respectively. In several instances, this was limited to statements such as “the scale is valid and reliable” with a reference to a prior citation. Other important features such as pilot testing, responsiveness, precision, and acceptability were reported in less than 10% of instances, and virtually all of those reports were from the Decisional Conflict Scale. A particular gap was found with cognitive interviews and feasibility which none of the studies mentioned.

Studies should include relevant details on the psychometric properties of measures used, so that readers can appropriately interpret results and conclusions. Psychometric performance can vary across settings, samples, and measurement contexts (8-10)(8). A common misperception is that if one simply picks a validated and reliable survey instrument, then

there is no more work to be done. In reality, validity and reliability are not properties of a survey instrument; rather it is the data and the interpretation of the data (which includes understanding the administration, setting, sample, and analysis procedures) that support construct validity (10). Thus, relevant information on psychometric performance ideally needs to be reported for each study and each use of an instrument or measure.

To support researchers in reporting the development and performance of measures, we outline a set of reporting standards for both new and existing measures used to evaluate PtDAs in Table 3. The recommended standards are applicable for any PtDA evaluation measure. The first five items are basic information required for a reader to understand what data were collected and how they were analyzed. For decision aid studies, it is important to understand the context and timing of the assessment, (such as how long after the decision aid the measure was administered, whether a consultation with a provider had occurred or not, whether treatment had been completed or not) and the mode of administration. Even with respect to these basic issues, information was often lacking in the reviewed studies.

Items 6-8 in Table 3 focus on the performance of the measure, and for each of these items there is considerable discretion as to the amount of detail to include. A full assessment of these items could provide enough material for an entire manuscript, or even a book, in the case of well-tested measures. The next three paragraphs provide some guidance on what would satisfy a minimum reporting requirement and then what would be considered a strong or enhanced reporting for these three items.

Some information on current experience with a measure should be included. All studies using new or previously published instruments will have data to report on some psychometric assessments such as, internal consistency reliability, acceptability, and even validity. Not all of these elements are required for each measure, and the type of information most relevant will vary depending on the measure. For studies reporting knowledge using a new measure, it is more important to provide some evidence that the measure demonstrates discriminant validity than internal consistency. The example in Table 3 taken from Barry et al 1997 (11) illustrates how they demonstrated discriminant validity by comparing responses of experts and patients and found significant differences in knowledge scores. On the other hand, studies using the total score from the Decisional Conflict Scale should report on the internal consistency in that sample and may refer to prior citations for evidence of validity, as illustrated in Table 3 from Vodemeier et al 2009 (12).

Often, there are prior publications that can be cited to provide evidence of psychometric properties. When including a citation, at a minimum, there should be specific acknowledgement of the kind of information is included in the citation (e.g. development process, reliability, acceptability). Even better would be to provide some details on the strength of performance of the measure, such as including an internal consistency reliability coefficient to describe reliability, naming the other measures used to establish divergent and/or convergent validity, and presenting the magnitude of the association(s). There are several textbooks that have detailed information on how to assess the adequacy of the psychometric evidence, one recommended by the authors is Waltz et al 2010 (13). Finally, an overall assessment should occur, perhaps in the discussion, that critically reflects on the

performance of the measure in the current study and how aspects of the sample, or administration, or scoring extends the known properties of the measure is important to advance the field.

In some situations, a measure may not exist that is appropriate for a given study. For decision aid studies, this situation often arises with knowledge measures and measures of patients' goals and preferences, which are specific to the decision. Researchers should be aware of several resources to search for potential measures such as the National Cancer Institute's GEM ([www.gem-beta.org](http://www.gem-beta.org)), the Ottawa Health Research Institute's evaluation section (<http://decisionaid.ohri.ca/eval.html>), and the National Institute of Health's PROMIS ([www.nihpromis.org](http://www.nihpromis.org)). Sometimes it may be preferable or necessary to develop a new measure as opposed to adapting an existing one. Often, these measures are designed and used in one study, and the authors may not have any plans to use the measure again or to make it widely available. The limited use, however, does not mean that it is not important to develop a strong measure, particularly if it is a main outcome measure for a study of an intervention. As Table 3 lists, reporting on newly-developed measures requires the basic information (items 1-5) as well as details on the development. In particular, it is necessary at a minimum to include some information on content validity, or how the items were generated (e.g. was the approach driven by theory? How much input did patients have? Was there any empirical testing?), and on how the developer assured understandability (e.g. focus group testing, cognitive interviewing, pilot testing). Additional details about the development process are desirable, but it is often not feasible to include in the manuscript itself. A link to an online supplement or reference to another publication where readers may obtain additional details is helpful. Finally, for new measures, it is important to provide the items or access to the items in the manuscript, an online appendix or website.

Other organizations have promoted increased attention to measurement properties and have made recommendations for patient-reported outcome measures in treatment trials (14). The minimum standards proposed here are aligned with broader guidelines and adapted for measures of decision aids. The purpose of this effort is to support researchers in selecting measures and enabling more complete reporting within the scientific literature. However, tensions exist in both the reporting and the development of measures which must be acknowledged. With regards to reporting, there is a practical tension around how much can reasonably be reported within a manuscript. Given the word limits imposed by many journals, it may be necessary to limit the amount of detail reported on psychometrics in the manuscript. However, many journals will allow authors to include online supplements that may alleviate the restrictions on word limits and make this information available to readers, reviewers and editors. Finally, measure developers can post user guides on their websites or on a centralized website such as the National Cancer Institute's Grid Enabled Measures project (see for example, [www.gem-beta.org](http://www.gem-beta.org), <http://decisionaid.ohri.ca/eval.html>, or <http://www.massgeneral.org/decisionsciences/>.)

A second tension exists with regards to balancing the requirements for strong psychometrics with the need to develop new measures. If every measure has to be extensively examined prior to its use in a study, this could have the unintended consequence of inhibiting innovation and/or slowing the pace of research. Researchers may avoid creating new

measures, even in situations where no measure exists. Over-reliance on existing measures may prevent testing of new theories or hypotheses. The standards outlined in Table 3 can be achieved with a reasonable amount of effort; particularly if researchers are thinking about evaluation early in the process. For instance, it would not require much additional effort to develop and test decision-specific measures of knowledge or goals in conjunction with the development of a decision aid.

There is a benefit to having a core set of measures that is used consistently across studies. If every investigator creates their own measures, then cross study comparisons become impossible and the field as a whole will have a more difficult time moving forward. To help bring standardization to the field, a new resource is available for formal and informal reporting on measure constructs and instruments through the National Cancer Institute's Grid Enabled Measures shared decision making database. Ideally, this effort will support researchers in selecting strong measures and enable more complete reporting (15).

Our study has several limitations. First, we only included RCTs from the Cochrane review of PtDAs. Second, we did not review the cited sources for previously published measures, as we were focused on the quality of reporting, not the quality of the measures themselves. As a result, these findings should not be interpreted to mean that the measures used were necessarily poor; rather, the reporting of measures was inadequate.

While not every psychometric property should be reported for every measure in every publication, current practice is clearly inadequate. Improving the process and quality of decision making for patients is of paramount importance; accurate measurement and reporting of the performance of the measures used is essential to moving the field forward.

## Acknowledgments

**Funding Source:** None<sup>1</sup>

## Appendix

### Data abstraction fields

- Reviewer 1 Initials
- Reviewer 2 Initials
- Paper (Author, Date)
- Instrument name (or Not Named)

### Decision Quality

- Knowledge Measure Yes=1/ No=0
- Realistic expectations Yes=1/ No=0
- Preferences-Choice [None=0; Preferences=1; Choice=2; Match calculated=3]

---

<sup>1</sup>No direct financial support was provided to authors for this manuscript. Dr. Matlock was supported by the National Institutes on Aging (K23AG040696). Dr. Sepucha received research and salary support from Informed Medical Decisions Foundation (0100).



## Decision Process

- Recognize decision Yes=1/ No=0
- Feel that understand options and outcomes Yes=1/ No=0
- Feel clear about what matters most Yes=1/ No=0
- Discuss goals with providers Yes=1/ No=0
- Involved in decision making Yes=1/ No=0

## General Information

- Addl studies that used instrument and/or validation reference that is/are cited in this paper
- Medical Topic / Decision
- Description of measure (e.g. items, response format, mode of administration)
- New? [1= instrument developed for this study or 2=instrument previously published (and/or adapted from published)]
- Quality of Instrument reported in this paper? Yes=1 (at least some data reported) / No=0
- Notes and comments

## Development process

- Item generation description
- Cognitive interviews Yes=1/ No=0
- Pilot testing Yes=1/ No=0

## Psychometric Properties

- Reliability reported? (Instrument is reproducible and when appropriate internally consistent) Yes=1/ No=0
- If yes, describe type and result (Cronbach alpha, retest reliability intraclass correlation coefficient, inter-rater reliability, etc)
- Validity reported? (the instrument measures what it purports to measure) Yes=1/ No=0
- If yes, describe type and results (e.g. content, criterion, predictive, discriminant, factor analysis, etc)
- Responsiveness: instrument is sensitive to changes of importance to patients and clinicians Yes=1/ No=0
- If yes, describe type of analysis and result
- Precision: accuracy and number of distinctions made by the instrument Yes=1/ No=0



- If yes, describe type of analysis and result
- Interpretability: evidence the scores are meaningful to clinicians and patients  
Yes=1/ No=0
- If yes, describe type of analysis and result
- Acceptability: instrument is acceptable to respondent (in most cases patients) if evidence cited from response rates and/or missing data Yes=1/ No=0
- If yes, describe type of analysis and result
- Feasibility: effort, burden, disruption required to administer on part of staff and clinical team reported Yes=1/ No=0
- If yes, describe type of analysis and result

## References

1. Institute of Medicine. Crossing the quality chasm: A new health system for the 21st century. Mar 1-8. 2001
2. Oshima EL, Emanuel EJ. Shared Decision Making to Improve Care and Reduce Costs. *N Engl J Med.* 2013; 368(1):6–8. [PubMed: 23281971]
3. Sechrest L, McKnight P, McKnight K. Calibration of measures for psychotherapy outcome studies. *Am Psychol.* 1996; 51(10):1065–1071. [PubMed: 8870543]
4. Sepucha KR, Borkhoff CM, Lally J, Levin CA, Matlock DD, Ng CJ, et al. Establishing the effectiveness of patient decision aids: key constructs and measurement instruments. *BMC Medical Informatics and Decision Making.* 2013; 13(Suppl 2):S12. [PubMed: 24625035]
5. Stacey D, Bennett CL, Barry MJ, Col NF, Eden KB, Holmes-Rovner M, et al. Decision aids for people facing health treatment or screening decisions. *Cochrane Database of Systematic Reviews.* 2011; 10, Art. No.: CD001431. doi: 10.1002/14651858.CD001431.pub3
6. Kryworuchko J, Stacey D, Bennett C, Graham ID. Appraisal of primary outcome measures used in trials of patient decision support. *Patient Educ Couns.* 2008 Dec; 73(3):497–503. [PubMed: 18701235]
7. Scholl I, Koelewijn-van Loon M, Sepucha K, Elwyn G, Härter M, et al. Measurement of shared decision-making - a review of instruments. *Z Evid Fortbild Qual Gesundheitswes.* 2011; 105(4):314–324.
8. Messick S. Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist.* 1995; 50:741–749.
9. Osterlind, SJ. *Modern measurement: theory, principles, and applications of mental appraisal.* 2nd. Upper Saddle, NJ: Pearson; 2009.
10. Sechrest L. Validity of measures is no simple matter. *Health Services Research.* 2005 Oct; 40(5 Pt 2):1584–1604. [PubMed: 16178997]
11. Barry MJ, Cherkin DC. A randomized trial of a multimedia shared decision-making program for men facing a treatment decision for benign prostatic hyperplasia. *Disease Management and Clinical Outcomes.* 1997; 1(1):5–14.
12. Vodermaier A, Caspari C, Koehm J, Kahlert S, Ditsch N, Untch M. Contextual factors in shared decision making: a randomised controlled trial in women with a strong suspicion of breast cancer. *Br J Cancer.* 2009; 100(4):590–597. [PubMed: 19209172]
13. Waltz, CF.; Strickland, OL.; Lenz, ER. *Measurement in nursing and health research.* 4th. NY: Springer; 2010.
14. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage M, et al. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *JAMA.* 2013; 309(8):814–822. [PubMed: 23443445]

15. Moser RP, Shaikh AR, Courtney P, Morgan G, Auguston E, Kobrin S, et al. Grid-enabled measures: using Science 2.0 to standardize measures and share data. *Am J Prev Med.* 2011 May; 50(5 Suppl 2):S134–S143. [PubMed: 21521586]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 1**  
**Elements of measure development and psychometric performance: abstraction criteria and descriptions**

<b>Measure Development</b>	
Item generation Cognitive testing Pilot studies	How were content items developed and by whom? Was the measure tested for understandability before use? Were pilot studies (of any type) conducted to pretest the measure?
<b>Measure Performance</b>	
Reliability	Were appropriate assessments of the reliability of the measure reported? If so, was there evidence of adequate reliability? Examples of assessments: internal consistency reliability (e.g., Cronbach's alpha, Kuder-Richardson coefficient); test-retest reliability; inter-rater reliability (e.g., percentage agreement, Kappa coefficient; intra-class correlation coefficient)
Validity (extent to which the measure assesses what is intended)	Were appropriate assessments of the validity of the measure reported? If so, was there there evidence of adequate validity? Types of validity assessment for self-report measures: 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)
<b>Measure Performance (other)</b>	
Responsiveness (sensitivity)	Is there evidence that the measure is sensitive to changes of importance to patients and clinicians?
Accuracy and precision	What is known about the measure performance in comparison to "gold standard" measures (accuracy) and/or the number of distinctions or extent of random error in use of the measure (precision)?
Interpretability	Are the scores meaningful to clinicians and patients?
Acceptability	Does the measure appear to be acceptable to respondents (usually patients; could include others); e.g., are there patterns to missing data or low response rates that could signal a problem with acceptability of the measure?
Feasibility of administration	Are there indicators of the appropriateness of effort, burden, or disruption (of clinical or research team) required to administer the measure?

**Table 2**  
**Reporting on performance of new and established measures of decision quality and decision process in studies of PtDAs**

	New Measures (n=61)	Prior published measures (n=117)		All Measures (n=178)
		DCS* (n=47)	Non DCS Prior published (n=70)	
Item Generation <sup>‡</sup>	20%	68%	66%	51%
Cognitive testing	0	0	0	0
Pilot studies <sup>‡</sup>	11%	0	1%	5%
Reliability	15%	30%	21%	21%
Validity	15%	17%	17%	16%
Responsiveness <sup>‡</sup>	0%	19%	0%	5%
Accuracy/Precision	3%	15%	0%	5%
Interpretability <sup>‡</sup>	0%	13%	1%	4%
Acceptability	2%	4%	0%	2%
Feasibility of administration	0	0	0	0

PtDAs=patient decision aids; DCS=Decisional Conflict Scale

\* Includes all versions of the DCS, as well as studies that only included some of the DCS subscales.

<sup>‡</sup> p<0.05 comparing frequency of reporting on new and prior published measures;

<sup>‡</sup> p<0.01 comparing frequency of reporting on new and prior published measures.

**Table 3**  
**Proposed minimum reporting standards for measures used to evaluate the effectiveness of PtDAs**

Type of measure	Sample reports extracted from published studies
<p><u>Previously published measures</u></p> <ol style="list-style-type: none"> <li>1 Name and version of instrument</li> <li>2 Domains measured</li> <li>3 Mode and timing of administration</li> <li>4 Number of items and response format</li> <li>5 How the measure is scored (including missing items if applicable).</li> <li>6 Experience with the measure in the current study (e.g., reliability, validity, acceptability)</li> <li>7 Type of psychometric performance previously published</li> <li>8 Adequacy and relevance of prior performance to current samples and settings</li> </ol>	<p>Extract from Vodermaier A et al., 2009.</p> <p><i>“Primary outcome variable Decisional conflict. The Decisional Conflict Scale (DCS; O’Connor, 1995) measures patients’ uncertainty about which treatment to choose, factors contributing to uncertainty (believing to be uninformed, unclear values, and unsupported in decision making), and perceived effectiveness of decision making. Questions have to be answered on a 5-point Likert scale [from strongly agree to strongly disagree]. Higher scores on the scale or subscales reflect higher decisional conflict, uncertainty, and a less effective choice. The German version of the scale demonstrated subscale and total score internal consistencies in the present sample between 0.73 and 0.94. The scale discriminates between patients who make and those who delay decisions (O’Connor, 1995; Bunn and O’Connor, 1996) and is sensitive to change (O’Connor et al., 1998).”</i></p>
<p><u>New measure</u>            Newly-developed measures should describe items 1-6 above, and should include:</p> <ol style="list-style-type: none"> <li>1 Measure development with a focus on content validity and understandability</li> <li>2 If permitted by publisher include the entire measure in the paper or as an appendix, otherwise state how the measure can be accessed.</li> </ol>	<p>Extracts from Barry et al., 1997.</p> <p><i>“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). The test of knowledge was administered two weeks after exposure to either the [patient decision aid]) or the control brochure.”</i></p> <p>and</p> <p><i>“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 &lt; 0.001). Nurses answered an average of 85% of the questions correctly, compared to 48% for the patients (p &lt; 0.001). Furthermore, a modest correlation between these patients’ knowledge scores and their educational levels was seen, r = 0.23 (p &lt; 0.001).”</i></p>