

LONDON
SCHOOL of
HYGIENE
& TROPICAL
MEDICINE



LSHTM Research Online

Kaltoft, MK; Nielsen, JB; Dowie, J; (2018) The Evaluation of Decision Support Tools Requires a Measure of Decision Quality That Has Content and Construct Validity in Person-Centred Care. *Studies in health technology and informatics*, 247. pp. 331-335. ISSN 0926-9630 <http://researchonline.lshtm.ac.uk/id/eprint/4649764>

Downloaded from: <http://researchonline.lshtm.ac.uk/4649764/>

DOI:

Usage Guidelines:

Please refer to usage guidelines at <http://researchonline.lshtm.ac.uk/policies.html> or alternatively contact researchonline@lshtm.ac.uk.

Available under license: Creative Commons Attribution Non-commercial
<http://creativecommons.org/licenses/by-nc/3.0/>

<https://researchonline.lshtm.ac.uk>

The Evaluation of Decision Support Tools Requires a Measure of Decision Quality That Has Content and Construct Validity in Person-Centred Care

Mette Kjer KALTOFT^{ab}, Jesper Bo NIELSEN^b, Jack DOWIE^{cb}

^a*Odense University Hospital Svendborg*

^b*University of Southern Denmark*

^c*London School of Hygiene and Tropical Medicine*

Abstract. The evaluation of decision aids and support interventions requires a multi-attribute index which can be calculated for both aid and comparator/s. The Decision Conflict Scale (DCS) is such an index and has been widely used in this context, the recent Cochrane review *Decision aids for people facing health treatment or screening decisions* reporting its use in 63 of the 105 studies included. However, while the DCS may be a valid measure for the eponymous construct – decision conflict - it lacks both content and construct validity for the evaluation of decision aids. It lacks *content validity* for this task because of the 3 items which make up its Uncertainty subscale, which penalize an aid that correctly reports the situation is one of decisional equipoise or near equipoise. A ‘false clarity’ bias in aid presentation is encouraged by being rewarded. In this paper we confirm that the inclusion of the Uncertainty subscale in the DCS has inappropriate empirical consequences for decision aid evaluation. Excluding the Uncertainty items would address this content invalidity, but the DCS would still lack *construct validity*, since the construct being measured is being treated, and inappropriately psychometrically validated, as a reflective rather than formative one. The component scales (items) of an index for a formative construct need to be preference-weighted, in ethical person-centred care by the individual at the point of decision. We argue that the most appropriate formative construct for use in decision aid evaluation is decision quality, and suggest MyDecisionQuality (MDQ) as the first formative index that can claim to both possess content and construct validity for the measurement of decision quality. However, like all multi-criterial formative metrics, the construct that MDQ measures is constructed by the measure and has no existence independent of it.

Keywords. Person-centred care, decision quality, formative, reflective, Decision Conflict Scale, MyDecisionQuality

Introduction

The recent Cochrane review *Decision aids for people facing health treatment or screening decisions* (1) reports that the Decision Conflict Scale (DCS) (2) was used to evaluate aid against comparator in 63 of the 105 included studies. However, the DCS lacks both *content* and *construct* validity for this evaluation task, as opposed to that of measuring the construct in its title, decision conflict.

Decision Conflict Scale: Content validity

A call to end the use of the DCS in evaluating decision aids was made ten years ago (3) on the grounds of its inclusion of items concerning uncertainty reduction. The offending items in the DCS are ‘This decision is easy for me to make’, ‘I feel sure about what to choose’ and ‘I am clear about what choice is best for me’. The call has been repeated occasionally since, most recently by Vickers (4)

Nelson and others argue that the use of decisional conflict as an end point in studies of decision aids involves the assumption that “decisional conflict and uncertainty represent an undesirable state that is detrimental to decision making.” They point out that “appropriate deliberation about alternative outcomes and personal goals, as well as ongoing engagement in the decision-making process,” may increase conflict.

In replying to Vickers, Misra-Hebert and Kattan denied the possibility of fundamental decisional equipoise and implicitly supported the provision of unwarranted certainty for therapeutic reasons.

We agree that decisional conflict is not to be avoided when making rational decisions but it should be a temporary state that eventually leads to certainty about the final choice, whether through a clear physician recommendation or careful patient consideration (5).

Evaluations of decision aids should accept that *decisional* equipoise - irrespective of whether or not there is *clinical* equipoise - is a possible and legitimate outcome from their use, even after full and unbiased processing of evidence and preferences. In ‘toss-up’ or ‘close call’ situations two or more options may be equally or almost equally good (6). Instruments used in the evaluations of decision aids or decision quality should not reward ‘sureness’ or ‘decisional conflict reduction’, since these may encourage a ‘false clarity’ bias (7). A strong ethical case can be made that, in person-centred care, there can be no therapeutic justification for misrepresentation of uncertainties which result in it being unclear which is the single best option for a particular individual. It does not matter whether this misrepresentation comes from an aid, or from a clinician.

These calls have not been sufficient to stop the routine use of the DCS in decision aid evaluations. In the recent Cochrane review *Decision aids for people facing health treatment or screening decisions*, 60% of the included studies used the DCS in aid evaluation. While no empirical results are relevant to the conceptual objection, we examined the impact of including the inappropriate Uncertainty items in the DCS score, in order to establish whether the conceptual flaw - from the perspective of decision aid evaluation - has empirical consequences.

As an adequate sample, we identified the 20 studies in the recently updated Cochrane review that reported all five subscale measures (Uncertainty, Informedness, Value Clarity, Support and Effective Choice) and were part of the pooled analysis. (The pooled analysis included 42 studies, but the remainder did not report all five subscale measures. The mean difference for total DCS scores was -7.22 points out of 100 in the 42 studies, -8.04 in our 20.)

All data are taken from Analysis 4.1 in the Cochrane review (1), which contains the full references. Spreadsheet available on request.

We found that in the Decision Aid arm the Uncertainty subscale score was the highest ('worst') of the five in 18 out of 20 studies (90%). In the Usual care arm it was highest ('worst') in 12 out of the 20 (60%). We can conclude that inclusion of the Uncertainty items almost always increased the score and therefore reduced the effect of the Decision Aid relative to Usual Care. But by how much? In each study, we took the average score of the 4 subscales other than Uncertainty and compared it with the Uncertainty subscale score. Over the 20 studies the Uncertainty score was, on average, 46% higher than the average of the other four in the Decision aid arm (mean 1.46, range 1.0 to 3.3) and 22 % higher in the Usual Care arm (mean 1.22, range 0.8 to 2.2).

In the majority of cases the Uncertainty score was lower in the Decision Aid arm, but in 5 instances (25%) it was higher, a completely tenable result which does not warrant the 'worse' characterisation in the aid evaluation context. These cases may be ones where the aid increased decision quality by increasing uncertainty. In 3 of these 5 the Informed scale was lower (better) in the Decision Aid, as it was in 18 of the 20 studies. However, reporting being better informed is entirely consistent with being more uncertain about the best option, since the information in the aid may well move the situation in the direction of decisional equipoise.

A recent study focusing on making genetic test results available – an alternative form of decision support - had the Uncertainty subscale highest in both before and after phases. The Uncertainty score was 41% higher than the average of the other 4 subscales in both phases (8).

We conclude that the use of the DCS, as a result of its inclusion of the Uncertainty items, is having a detrimental effect on quantitative evaluations of decision support interventions. The conclusion applies even more to the 4-item SURE version of the DCS (9). Its single Uncertainty item ('Do you feel SURE about the best choice for you?') accounts for 25% of the index; the 3 Uncertainty items of the DCS comprise only 19% of the 16.

Decision Conflict Scale: Construct validity

The case against using the DCS in decision aid evaluation is, however, not confined to its inclusion of the Uncertainty subscale and consequent lack of content validity. It lacks construct validity as an outcome metric for a comparative evaluation of decision aids in person-centred care. The required outcome measure is a multi-criteria index measure of a preference-sensitive *formative* construct, in which the item weights are personalised (10).

Decision Quality is a *formative* construct, like Health-Related Quality of Life. HRQOL is constructed – in the act of being measured - by EQ-5D, the Health Utilities Index, and other instruments in this analogous case. The multiple items in these instruments, such as pain or mobility, 'cause'/'form' the index value of HRQOL, are not

‘caused’/‘formed’ by it – as they are in a *reflective* construct. “Health-related quality of life does not determine the ability to move, or pain. On the contrary, the ability to move and pain determine health-related quality of life.” (11). Different measures of Decision Quality, or any formative construct (including Decision Conflict), do not yield the same index value because they do not measure the same ‘it’.

The validation of formative and reflective instruments must be approached very differently. For reflective instruments, the COSMIN set of validation criteria provides guidance. But the COSMIN group clearly accept that most of these standard criteria, including internal consistency and structural validity, are inappropriate when the construct is formative.

A reflective model is a model in which all items are a manifestation of the same underlying construct. These items are called effect indicators and are expected to be highly correlated and interchangeable. Its counterpart is a formative model, in which the items together form a construct. These items do not need to be correlated. Therefore, internal consistency is not relevant for items that form a formative model (12).

What sort of instrument should be used for evaluating decision aids?

We argue that the most appropriate primary outcome in an evaluation of decision aids is a measure of a preference-sensitive formative construct ‘decision quality’. The main instruments being used to measure ‘decision quality’ in the context of decision aid evaluation are those developed by Karen Sepucha and colleagues, and researchers following their method (13–14). From our perspective these DQIs lack both content and construct validity in person-centred care, not being individually preference-sensitive index measures, assessed and available immediately after the point of decision; crucially, being measured before any actions are taken or outcomes known. Meeting psychometric tests appropriate only for reflective constructs cannot compensate for the lack of these two fundamental validities.

My Decision Quality has been advanced as the first instrument to measure Decision Quality as an individually preference-sensitive formative construct (15). It can therefore claim to possess construct validity, as well as content validity, these being the two sufficient conditions for instrument use in decision aid evaluation. It has not been subjected to inappropriate psychometric testing.

Conclusion

Given that all multi-criterial formative constructs are preference-sensitive, searching for an objective definition of decision quality – or indeed decision conflict - is hunting a mythical creature (16). MyDecisionQuality measures the construct of decision quality that MDQ measures. The best instrument to measure decision quality, for whatever purpose, is the one that best reflects the decision maker’s preferences in relation to *their* construct of decision quality.

References

- [1] D. Stacey, F. Légaré, K. Lewis, M.J. Barry, C.L. Bennett, K.B. Eden, et al., Decision aids for people facing health treatment or screening decisions, *Cochrane Database Syst Rev* (2017).
- [2] A. O'Connor, Validation of a Decisional Conflict Scale, *Med Decis Mak* **15** (1995), 25–30.
- [3] W.L. Nelson, P.K.J.Han, A. Fagerlin, M. Stefanek, P.A. Ubel, Rethinking the objectives of decision aids : A call for conceptual clarity, *Med Decis Mak* **27** (2007), 609–18.
- [4] A.J.Vickers, Decisional Conflict, Regret, and the Burden of Rational Decision Making, *Med Decis Mak* **37** (2017), 3–5.
- [5] A. Misra-Hebert, M Kattan, Comment on : Decisional Conflict, Regret, and the Burden of Rational Decision Making, *Med Decis Mak* **37** (2017), e360-1.
- [6] S.G. Pauker, J. Kassirer, The Tossup, *N Engl J Med* **305** (1981), 1467–9.
- [7] M.K. Kaltoft, J. Dowie J, R. Turner, J.B. Nielsen, G. Salkeld, M Cunich, Decisional equipoise is not decisional conflict: Avoiding the false clarity bias in the evaluation of decision aids and shared decision making processes, Poster Eur Soc Med Decis Making, Antwerp (2014).
- [8] K.V. Macdonald, Y. Bombard, K. Deal, M Trudeau, N Leigh, D.A. Marshall, The influence of gene expression profiling on decisional conflict in decision making for early-stage breast cancer chemotherapy, *Eur J Cancer* **61** (2016), 85–93.
- [9] F. Légaré, S. Kearing, K. Clay, S. Gagnon, D. D'Amours, M. Rousseau, et al., Are you SURE? Assessing patient decisional conflict with a 4-item screening test, *Can Fam Physician* **56** (2010), e308–14.
- [10] M.K. Kaltoft, J.B. Nielsen, J. Dowie, Formative preference-sensitive measures are needed in person-centred healthcare at both clinical and policy levels, *Eur J Pers Centered Healthc* **5** (2017), 495-500.
- [11] U. Konerding, Preference-based index measurement of health-related quality of life: when does it reflect only arbitrary settings of the researcher? *Health Econ* **20** (2011), 471–83.
- [12] L.B. Mokkink, C.B. Terwee, D.L. Knol, P.W. Stratford, J. Alonso, D.L. Patrick, et al., The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content, *BMC Med Res Methodol* **10** (2010), 22.
- [13] C. N. Lee, R. Dominik, C.A. Levin, M.J Barry, C. Cosenza, A.M. O'Connor, et al., Development of instruments to measure the quality of breast cancer treatment decisions, *Heal Expect* **13** (2010), 258–272.
- [14] K.R. Sepucha, C.A.Levin, E.E. Uzogara, M.J. Barry, A.M. O'Connor, A.G. Mulley, Developing instruments to measure the quality of decisions: Early results for a set of symptom-driven decisions, *Patient Educ Couns* **73** (2008), 504–10.
- [15] M.K. Kaltoft, M. Cunich, G. Salkeld, J. Dowie, Assessing decision quality in patient-centred care requires a preference-sensitive measure, *J Heal Serv Res Policy* **19** (2014), 110–7.
- [16] R.L. Wears, The hunting of the snark, *Ann Emerg Med* **58** (2011), 465–7.