

Decision Quality Is a Preference-Sensitive Formative Concept: How Do Some Existing Measures Compare?

Jack DOWIE^{ab1}, Mette Kjer KALTOFT^b,
^{ab}*London School of Hygiene and Tropical Medicine, UK*
^b*University of Southern Denmark, Denmark*

Abstract. The primary output of a decision making process is a decision and a key outcome measure is therefore decision quality. However, being a formative construct, ‘decision quality’ is both preference- and context-sensitive and legitimate alternative measures accordingly exist. A decision maker wishing to measure decision quality in the evaluation of a decision or decision making process needs to be aware of the attributes of the measures on offer. This paper establishes some of the key conceptual differences by examining two measures: Decision Quality Instruments and MyDecisionQuality. Four of their main conceptual differences relate to: the timing of the measurement (at the point of decision or at follow-up when the ‘downstream’ outcome is known); (whether or not an objective assessment of the information state of the individual is included (as opposed to self-reported state); whether the instrument itself is preference-sensitive; and whether the measure is to be used in the context of individualised clinical practice at the point of care or only in research to produce group level feedback. Establishing agreed measures of decision quality is necessary and useful, so long as it is accepted that it is a preference- and context-sensitive construct, in the way that is widely acknowledged in relation to, for example, Health-Related Quality of Life, with its many measures.

Keywords: decision quality, Decision Quality Instruments, MyDecisionQuality, formative construct, preference-sensitive.

1. Introduction

If the fundamental aim of informatics is to help decision makers make high-quality decisions, it might be expected that the measurement of ‘decision quality’ would be a major focus of interest. However, there is limited attention to this *outcome* in the literature, compared to the amount devoted to related *input* qualities: the quality of the information inputs being produced, the quality of the transmission and dissemination of information inputs, and the quality of the processing of the information inputs within decision making. In this paper we explore two specific approaches to the measurement of decision quality, an exploration which, while restricted in scope, is sufficient to throw significant light on the reasons for this relative paucity of attention. Among other things, it may help explain the surprising non-appearance of decision quality as a

¹ Corresponding author: Jack Dowie, LSHTM, 15-17 Tavistock Place, London, UK WC1H 9SH; email: jack.dowie@lshtm.ac.uk

Patient-Outcome Measure (PROM). In the comprehensive 2015 survey of the latter [1], the term ‘decision quality’ appears only within the section headed ‘Patient experience’ and not as a PROM.

Among the underlying reasons for this marginalisation is that the latent unobservable construct (‘decision quality’) is a *formative*, not a *reflective* construct, and therefore preference- and context-sensitive [2]. This has both methodological and practical consequences. Among the methodological consequences are the psychometric standards appropriate for validating instruments seeking to measure it. ‘Decision quality’ is an ‘it’ that does not exist in the sense that a reflective construct (such as appendicitis) exists, where psychometric standards such as internal consistency are valid. The main practical consequence of its formative nature is that there are many possible and acceptable definitions of decision quality, none of which can be validated as the ‘gold standard’. The ambition to develop a universally agreed definition of a formative construct is understandable since widely-agreed definitions have their uses and in many senses are essential. The various formative measures of Health-Related Quality of Life exemplify acceptance of their intersubjective, value-based provenance. However, capitalising a construct as Shared Decision Making or Decision Quality cannot produce ontological transformation from formative to reflective, even if it is sometimes institutionally convenient to treat it as if it does.

There are many meaningful distinctions which can and should be made between different proposed measures of decision quality, and the basic point being made here is that the *decision* on which measure of decision quality to select for use, from a set of alternative measures given this label, is both *preference-sensitive* and *context-sensitive*. *Different instruments will measure different constructs and are not to be regarded as different measures of the same construct.*

A decision maker selecting the measure that is optimal for the particular evaluation they are undertaking needs to be informed about their content and construct validity. The limited aim here is to establish *some* of the key *conceptual* differences in measures of decision quality. We set aside most practical differences, the need for alternative versions or adaptations for contexts such as low literacy or ethnicity being taken for granted. Examining two measures will be sufficient for our limited purpose: the Decision Quality Instruments (DQIs) [3-4] and MyDecisionQuality (MDQ) [5-6]. The concept is often used implicitly as an umbrella for a set of individual indicators [7].

2. The Decision Quality Instruments (DQIs)

A generic DQI framework is used in the production of condition-specific instruments, 14 of which are currently available, including ones for breast cancer surgery, prostate cancer screening and treatment, back pain and knee and hip osteoarthritis. (<https://mgHdecisionsciences.org/tools-training/decision-quality-instruments/>)

Since the DQIs are condition-specific, we will focus on that for Breast Cancer Surgery (BCS-DQI) [3-4]. The BCS-DQI comprises a set of 15 ‘objective’ knowledge questions ‘to determine whether patients are informed’ and a set of 8 questions about their ‘goals and concerns’. As a profile measure, it produces two scores. The *DQI-Knowledge Score* is the number of correct responses divided by the number of items, resulting in scores from 0% to 100%. A threshold for considering a patient to be ‘well-informed’ is set, using (if available) the mean knowledge score for a group of patients who have viewed a decision aid. The *DQI-Concordance Score* measures ‘the extent to

which patients received treatments that reflected what is most important to them'. It is derived from a multivariate logistic regression model in which the treatment received (e.g. surgery vs. non-surgical) is the dependent variable. The goals and concerns specified by the respondent are the independent variables, with other factors that should influence treatments (e.g. stage of disease for breast cancer surgery) included, as needed. The dependent variable is a predicted probability of treatment for each patient. Patients with a predicted probability of more than 50% who had the treatment, and those with a predicted probability equal to or less than 50% who did not have the treatment, are classified as having treatments that match their goals and are 'concordant'. (Dividing the number who matched by the total number in the sample then yields a summary percentage concordance score for the group.) A binary *Decision Quality Composite Score* can be created with a score of 1 for patients who were well-informed [i.e. met the pass threshold] and received treatments matching their preferences, and 0 for all others.

The interpretation of both the 2-score profile and the binary composite score is only at the group level and available only after follow-up, months after the decision. So the DQIs are essentially research tools for evaluating a decision process (which is assumed to be 'shared decision making', usually including a decision aid), for the purpose of improving future practice through group level feedback. They are not envisaged as tools to be used as the source of individual feedback in real time within clinical practice.

3. MyDecisionQuality (MDQ)

MDQ is a preference-sensitive instrument in which individuals provide their responses as soon as possible after a decision is made. The responses are of two distinct types - their personal ratings of the preceding decision- and decision-making process on 8 criteria, and their weightings for those 8 criteria. The MDQ score is the weighted sum of the ratings (i.e. the expected value for each option). The abbreviated labels for the 8 items are four which relate to the decision content - Options, Effects, Criterion Importance, and Effect Chances - and four which relate to the decision process - Trust, Support, Control, and Commitment [5-6].

MDQ is a self-reported, preference-sensitive ('dually personalised') multi-criterial index, generic across both conditions and contexts. Its primary purpose is to evaluate the quality of a decision and decision process and do so at the individual level. The evaluation is to occur as soon as possible after the decision is taken. This is to rule out reference to any downstream outcomes, for whatever reason. (Two common ones are to establish whether action or behaviour subsequent to the decision was in accordance with the person's expressed 'goals and values' and to permit their *experienced* regret to affect the assessment of decision quality.) It also excludes 'objective knowledge possession' as a criterion. However, it addresses both the decision and the decision process, being designed to indicate whether the 'reasonable patient' legal standard for *informed and preference-based* consent has been met (items 1-4), as well as the 'subjective patient' ethical standard for the care process (items 5-8). Where clinicians agree to complete an equivalent measure, it can provide decomposed measures of concordance (e.g. between patient and practitioner) and so establish which ratings or weightings are the source of any discordance. A sub-analysis indicates the criteria which hold the greatest potential for improved future decision quality. It is possible to

aggregate MDQ scores across individuals and arms of trials [8], but the interpretation of these results must be approached with caution, given their individual preference-sensitivity.

4. Discussion

A wide variety of constructs of decision quality are possible, with the main distinctions resting on *why* they are done, *who* is asked to provide *which* type of responses/s to *what* items, and *when* and *where* they are asked to do so. The answers determine whether the measure of decision quality is preference-sensitive in relation to the items in the instrument; whether the measure is useful in real time clinical practice or is essentially an output from evaluation research at a group level; and whether it is appropriate in evaluations of a decision, of a decision process, or a decision aid (including policy makers and regulators approving the last of these).

The timing of measurement is perhaps most important, because for many it impinges on the validity of the measure, as well as its use. Elwyn and Miron-Shatz report the widespread view that ‘Good (or bad) decisions may have good or bad outcomes by virtue of chance, at the level of individual consequences.... post hoc assessments, based on the outcomes of decisions, are unsafe measures of good decisions.’ [9] (p141-2). Delayed assessment also permits regret to enter into any evaluation at ‘downstream’ follow-up. *Experienced* regret - indeed all hindsight - is a valuable input into a future decision, but only if it is not allowed to affect the evaluation of a previous one, in which one has had the opportunity to input *anticipated* regret, in foresight. Dissonance reduction is another post hoc influence on evaluations of decision quality widely regarded as undesirable.

Incorporation of an objective knowledge test is a further example of how a different construct of decision quality is created. Decision makers in some contexts may prefer a measure reflecting responses from ‘objectively informed persons’. Others will reject this: ‘... knowledge as a necessary component of a good decision - both as an evaluation of the effectiveness of deliberation and as an evaluation of the determined decision - may be difficult to operationalize and, at a deeper level, may reflect an untested assumption - that a standardized, pre-specified level of knowledge is necessary in order to achieve a high level of deliberation.’ [9] (p144).

5. Conclusion

Since the latent construct of decision quality is formative, each of these instruments measures the construct it measures. An instrument should be chosen for use in an evaluation if its characteristics are preferred and it is context-relevant, as deemed by the decision maker. Ambitions to produce a standardised, preference- and context-free decision quality assessment instrument should be resisted, with the different products accepted as a menu from which a preference- and context-sensitive selection can be made. We note that it is claimed that the definition of decision quality in DQIs has been validated by the International Patient Decision Aid Standards group [4,10]. While this establishes its serious credentials as a construction of decision quality, as with any formative construct these rest on the intersubjective agreement about its contents (items and weightings), reflecting the beliefs and preferences of the producing group. MDQ

does not seek to answer the philosophical question of what constitutes a ‘good decision’ [11], merely to provide a measure of ‘decision quality’ with its particular set of characteristics. Decision quality should be regarded as *the* fundamental outcome of a decision making process, despite the challenges its formative character presents.

Funding

The software used in current implementations of MDQ was installed at <https://ale.rsyd.dk> in a nationally funded project to develop decision support tools. SATS J.nr.1-1010/116/27.

Conflict of Interest

Jack Dowie has a financial interest in the commercial use of the Annalisa software.

References

- [1] D. Cella, E. Hahn, S. Jensen, Z. Butt, C. Nowinski, N. Rothrock, et al., *Patient-Reported Outcomes In Performance Measurement. Research*, Triangle Park (NC); RTI Press 2015.
- [2] M.K. Kalsoft, J.B. Nielsen, J. Dowie, Formative preference-sensitive measures are needed in person-centred healthcare at both clinical and policy levels, *European Journal for Person Centered Healthcare* **5** (2017), 17-20.
- [3] K. Sepucha, E. Ozanne, K. Silvia, A. Partridge, A. Mulley, An approach to measuring the quality of breast cancer decisions, *Patient Education and Counseling* **65** (2007), 261–9.
- [4] K. Sepucha, J. Belkora, Y. Chang, C. Cosenza, C. Levin, B. Moy, et al., Measuring decision quality: psychometric evaluation of a new instrument for breast cancer surgery, *BMC Medical Informatics and Decision Making* **12** (2012), 51.
- [5] M.K. Kalsoft, M. Cunich, G. Salkeld, J. Dowie, Assessing decision quality in patient-centred care requires a preference-sensitive measure, *Journal of Health Services Research and Policy* **19** (2014), 110–7.
- [6] M.K. Kalsoft, J.B. Nielsen, G. Salkeld, J. Dowie, Enhancing healthcare provider feedback and personal health literacy: dual use of a Decision Quality measure, *Studies in Health Technology and Informatics* **218** (2015), 74–9.
- [7] K. McAlpine, K. Lewis, L. Trevena, D. Stacey, What is the effectiveness of Patient Decision Aids for cancer-related decisions? A systematic review subanalysis, *JCO Clinical Cancer Informatics* **2** (2018), 1–13.
- [8] G. Salkeld, M. Cunich, J. Dowie, K. Howard, M. Patel, G. Mann, et al., The role of personalised choice in decision support: A randomized controlled trial of an online decision aid for prostate cancer screening, *PLoS One* **11** (2016), e0152999.
- [9] G. Elwyn, T. Miron-Shatz, Deliberation before determination: the definition and evaluation of good decision making, *Health Expectations* **13** (2010), 139–47.
- [10] G. Elwyn, A. O’Connor, D. Stacey, R. Volk, A. Edwards, A. Coulter, et al., Developing a quality criteria framework for patient decision aids: online international Delphi consensus process, *BMJ* **333** (2006) 417–9.
- [11] J. Hamilton, S. Lillie, D. Alden, L. Scherer, M. Oser, C. Rini, et al., What is a good medical decision? A research agenda guided by perspectives from multiple stakeholders, *Journal of Behavioral Medicine* **40** (2017), 52–68.