

# Measures of Decision Aid Quality Are Preference-Sensitive and Interest-Conflicted - 1: Normative Measures

Jack DOWIE<sup>ab1</sup>, Mette Kjer KALTOFT<sup>b</sup> and Vije Kumar RAJPUT<sup>c</sup>

<sup>a</sup>*London School of Hygiene and Tropical Medicine*

<sup>b</sup>*University of Southern Denmark*

<sup>c</sup>*Stonydelph Health Centre, Tamworth, UK*

**Abstract.** The belief that following rigorous inclusive methods will eliminate bias from ‘quality’ measures ignores the preferences necessarily embedded in any formative instrument. These preferences almost always reflect the interests of its developers when one uses the wide definition of ‘interest’ appropriate in healthcare research and provision. We focus on the International Patient Decision Aid Standards instrument, a popular normative measure of decision aid quality. Drawing on its application to a set of 23 breast cancer screening decision aids, we show the effects of modifications that reflect our own different interest-conflicted preferences. It is emphasised that the only objection is to the implication that any formative instrument should be promoted or treated as the ‘the gold standard’, without a conflict of interests disclaimer, and to the implication that other instruments cannot provide equally valid, high-quality measures.

**Keywords.** Decision aid, preferences, normative, IPDASI, conflict of interest

## 1. Introduction

There is growing agreement that the future will be dominated by the social and self-production of health by citizens, optionally supplemented by its co-creation with healthcare professionals. This means that the provision of apomediative decision support direct to the person in the community is of increasing relevance and importance, in addition to intermediative decision support direct from the clinician to the *citizen-as-patient*. For clarity we will refer to the former as apomediative Person Decision Support Tools (PnDSTs) and the latter as intermediative Patient Decision Aids (PtDAs). The terms ‘intermediative’ and ‘apomediative’ are those of Eysenbach [1] and characterise, respectively, the presence or absence of a dependent relationship between the supplier of the support and the professional involved in the decision.

Despite this distinction the quality of both types must be assessed appropriately, and this includes addressing possible bias. Unfortunately, the belief that by following rigorous ‘scientific’ methods, all biases can be eliminated, involves confusing whether they are seen as acceptable, justifiable, or desirable, with their inevitability. This

---

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

inevitability is at its clearest in indexes developed to measure the *quality* of some state (e.g. of health), of some action (e.g. of surgery), of some process (e.g. shared decision making), or of some intervention (e.g. a decision aid or support tool). All quality indexes are *formative* measures and therefore dependent on the preferences - the value judgments - involved in (a) the selection of the component scales (instrument items) and (b) the weighting of those selected scales to form a quality index. Evaluations using such quality measures are therefore 'preference-sensitive' [2] and, given preferences are rarely in conflict with *interests*, they are usually biased in the sense of 'interest-conflicted'.

The definition of 'interest' used here is the wide one we regard as relevant in healthcare, including research and provision, not a narrow legal/financial one that ignores the serious conflicts arising for a variety of institutional and personal reasons [3]. Among these interests are the ones listed in the first section of Rodwin's typology:

Intellectual commitments (e.g. working within a theoretical framework, school of thought, or having proposed a hypothesis). Interest in a positive outcome to a study that will support your previous findings. Interest in maintaining professional reputation. Interest in career advancement. Interest in finding potential practical applications of research. Interest in maintaining good relations with entities that can provide future research funding [4].

With this wider definition of interest - particularly the inclusion of commitments to theoretical frameworks or schools of thought - the first issue is not whether a quality measure is 'interest-conflicted', but the nature and origins of the interest-conflicted preferences that are necessarily reflected in its development. Whether or not these interests are regarded by some (many) as acceptable or desirable, the second, but equally important issue, is the attitude and behaviour of those responsible towards alternative measures, which, by definition, also reflect interest-conflicted preferences. Quality being a formative construct, any alternative is a measure of a different construct (e.g. of decision aid quality), not an alternative measure of the same one.

The wide range of interests underlying the preferences reflected in a formative measure should therefore be declared alongside any legal requirement, accompanied by a denial of any intention to seek to establish the measure as '*the gold standard*', with an effective monopoly on professional endorsement and regulatory approval. An example of good behaviour is provided by the generic health-related quality of life measures, where alternatives co-exist in friendly, if robust, rivalry.

In this first of two papers we focus on the *normative* measurement of the quality of decision aids, in other words on measures which consider only the content of the aid and/or its development process. In the companion paper [5], we see how the argument applies in the *empirical* measurement of decision aid quality, as implemented. In both we focus on the products of the International Patient Decision Aid Standards (IPDAS) consortium, but it is important to see that the argument is completely generic.

We take the International Patient Decision Aid Standards instrument (IPDASi) as our specific focus [6]. It should already have been inferred that we will not be 'criticising' or even 'critiquing' IPDASi from some purportedly neutral, unbiased (interest-unconflicted) position. We will be pointing out the way the preferences and interests it reflects do not coincide with our own, and arguing why it should not be promoted and/or treated as the 'gold standard', as opposed to a standard based on a widely-agreed, but particular, set of preferences and interests. The setting of an IPDASi standard for decision aid 'certification' [7] is not in itself objectionable, only any implication or inference that aids that do not meet this certification standard should not

be regarded as usable *for this reason alone*. To repeat, this is a widespread phenomenon and the same danger can be detected with other quality measures [8,9].

To give the argument empirical flesh we draw on the recent paper by Hild and colleagues [10]. They assessed the quality of 23 decision aids for women at average risk of breast cancer (and eligible for mammographic screening) using the original IPDASi 47 item instrument. We take advantage of the complete data set they provided to see the effects of modifying the instrument to reflect our different preferences.

## 2. Modifying IPDASi

Our preferences are fully in agreement with IPDASi in distinguishing conceptually between binary *checklist* items and scalar *index* items. The former must be met. The latter may be only partially fulfilled and, most importantly, are compensable - lower ratings on some may be countered by higher ratings on others. IPDASi recognise this distinction in partitioning their later 44 item version into three parts [7] with a certification sub-set.

Six binary *qualifying* (checklist) items are ones to be met for something to qualify as a decision aid: “1) the intervention should relate to a specific decision that has to be made; 2) patients should be helped to choose deliberately among options; 3) positive and negative features of the options should be presented; 4) outcomes given should be relevant to health status; 5) the intervention should not promote compliance with a recommended option; and 6) the intervention should help patients to clarify values.”

Ten *certification* criteria, scored on a 1–4 scale (‘strongly disagree’ to ‘strongly agree’), are “deemed essential in order to avoid risk of harmful bias... Decision aids must meet all of these criteria to be certified. The 6 certification criteria selected relate to the quality of the evidence synthesis process, open disclosure of funding source, and a balanced presentation of options, with 4 additional items for screening/test aids.”

Twenty-eight *quality* items are deemed “desirable because they would enhance a decision aid but are not essential for reducing risk of harmful bias... These items would improve the experience of using the decision aid, but absence of the item would not be expected to influence the individual’s decision in a negative way.”

Tools should meet all *qualifying* criteria and score 3 or 4 on each of the 10 *certification* criteria in order to reach the certification standard. (Hild, et al. do not cite the 2013 paper in which this item classification and certification standard is introduced and hence make no reference to certification in their analysis.)

Our preferences would reduce the IPDASi list from 44 to just 5 items. We retain the IPDASi wording here, as sufficient for the present purpose.

1. The aid makes it possible to compare the positive and negative features of the available options.
2. The aid provides information about outcome probabilities associated with the options (i.e. the likely consequences of decisions)
3. The aid asks patients to think about which positive and negative features of the options matter most to them (implicitly or explicitly).
4. The aid (or associated documentation) describes how research evidence was selected or synthesized
5. The aid was field tested with patients who were facing the decision

All 5 of these fall only in the residual ‘quality’ category for IPDASi, in other words none are required for certification. It follows we must regard their 10 certification items

as either merely ones necessary for acceptance as an aid (we indeed regard 5 of them as ‘qualifiers’), or as redundant (the other 5 certifying items, including all the screening test ones, are indeed covered in item 2 above *if this is fulfilled properly*. Spelling them out as separate items either implies that item 2 is not required to be fulfilled properly, or involves double counting. (There is a major omission in the IPDASi test set in our view. The prior odds of the target condition are essential in decision support, prevalence being the usual proxy. Providing only the True and False Positive and Negative rates, as IPDASi requires, is likely to be misleading if the user cannot also see the False Alarm and False Reassurance Rates. These appear in item 2 if done properly.)

The correlation between the scores on our 5 item selection and the Hild 47 items scores is 0.89, confirming the feasibility of serious reduction. But the correlation with the certification items is only 0.66, confirming that preferences make a major difference. Applying the certification standard (rated 3 or 4), to our 5 items, only 3 of the 23 aids are certifiable. In contrast to IPDASi our preferences would certify the Schonberg aid, which fails IPDASi by not providing an update policy and not indicating the next steps if the target condition is not detected. (The latter is a redundant item for us.) Contrariwise, we would not certify the Keevil aid, because it is not describing well enough ‘how research evidence was selected or synthesized’. (The analysis is at <http://bit.ly/hildanalysis>, but the argument does not depend on its details.)

### 3. Interest-Conflicted Preferences

Where and how do *interests* come in? First, in the inclusion of *process* criteria regarding the development process, criteria made redundant for us by rigorous testing using *outcome* criteria. If this testing is conducted properly, we see no justification for using any aspect of the development process (including the credentials of the people involved in it) in establishing the quality of an aid.

Second, through the omission of items that preserve the interests of healthcare professionals in not having to deal with a preliminary opinion based on numerical analytic calculation, as opposed to one designed to fit a verbal deliberative reasoning process. Items 2 and 3, when fulfilled according to our preferences, provide all the ingredients necessary to calculate the expected value of each option, using the importance weights of the person. Calculating and displaying those scores is absent from IPDASi, whereas it is the central feature of aids based on other techniques.

Third, through preference specification. Our preference is for decision support tools that are preference-sensitive at the point of care. These cannot be based on group average tariffs, let alone be those of a panel whose expertise and eminence relate to belief judgments about option performance rates, not value judgements about those criteria. In the online spreadsheet we enable the preference-sensitive weighting of the 47 items IPDASi set. To illustrate, we assign weight to only our 5 preferred items. The reader is free to explore alternative weights.

### 4. Conclusion

That IPDASi was built by a large international consortium in a prolonged and rigorous Delphi process gives it unquestioned credibility and the right to have aids promote themselves as being ‘certified by the IPDASi standard’. But implying that it is a ‘gold

standard' and that an aid that fails to meet it cannot, *by that fact*, be a valid, possibly excellent, decision aid, conflicts with the scientific standards its developers - and the informatics community - would undoubtedly wish to uphold. The same applies to all alternative measures of decision support quality and indeed to all quality metrics.

## References

- [1] Eysenbach G. Medicine 2.0: Social Networking, Collaboration, Participation, Apomediation, and Openness. *J Med Internet Res*. 2008 10(3):e22.
- [2] Kalsoft MK, Nielsen JB, Dowie J. Formative preference-sensitive measures are needed in person-centred healthcare at both clinical and policy levels. *Euro J Pers Cent Healthcare*. 2017 5(4):495-500.
- [3] Saver RS. Is it really all about the money? Reconsidering non-financial interests in medical research, *J Law Med Ethics*. 2012 40(3):467–81.
- [4] Wiersma M, Kerridge I, Lipworth W, Rodwin M. Should we try to manage non-financial interests? *BMJ*. 2018;361:10
- [5] Kalsoft MK, Dowie J, Rajput VK. Measures of decision aid quality are preference-sensitive and interest-conflicted 1: normative measures. *Stud Health Technol Inform*. 2020: submitted
- [6] Elwyn G, O'Connor AM, Bennett C, Newcombe RG, Politi M, Durand M-A, et al. Assessing the quality of decision support technologies using the International Patient Decision Aid Standards Instrument (IPDASi). *PLoS ONE*. 2009 4(3):e4705.
- [7] Joseph-Williams N, Newcombe R, Politi M, Durand M-A, Sivell S, Stacey D, et al. Toward minimum standards for certifying Patient Decision Aids: A modified Delphi consensus process. *Med Decis Making*. 2013 34(6):699–710.
- [8] Dowie J, Kalsoft MK. Why a Global PROMIS can't be kept, *Stud Health Technol Inform*. 2019 262: 114–17.
- [9] Kalsoft MK, Dowie J. The evaluation of decision support tools needs to be preference context-sensitive, *Stud Health Technol Inform*. 2019 265:163-68
- [10] Hild S, Johanet M, Valenza A, Thabaud M, Laforest F, Ferrat E, Rat C. Quality of decision aids developed for women at average risk of breast cancer eligible for mammographic screening: Systematic Review and assessment according to IPDASi. *Cancer*. 2020 126(12):2765-74.