Translating the results of Discrete Choice Experiments into p-/e-/m-health decision support tools

Jack DOWIE[ab1], Mette Kjer KALTOFT[b],

[ab]*London School of Hygiene and Tropical Medicine*
[b]*University of Southern Denmark*

Abstract. The rapidly growing number of health-related Discrete Choice Experiments (DCEs) has not been matched by studies of their impact on decision or policymaking. However, it is widely assumed that this impact has been very limited, despite the potential relevance of the resulting average preferences to group policy development. The main, but at the moment essentially speculative, explanation offered, focuses on the methodological quality of the DCEs and their reporting. An alternative explanation, equally speculative, lies in the research-practice gap created by the conceptualisation of the DCE as a purely research exercise, not supplemented by any attempt to translate the findings into analytic decision support form. This also applies in the clinical decision context, where there are frequent claims that DCE results can assist in an individual's decision making. In the absence of suggestions as to how group results can analytically facilitate preference-sensitive care (and legally informed consent), we propose a generic add-on for DCEs with 'real' options, attributes, and attribute levels. This takes the form of a multi-criteria analysis-based decision support tool. Exemplars, showing how preference-sensitive individualised opinions can be derived from published DCEs for Heavy Menstrual Bleeding and Prostate Cancer Screening, may be consulted online.

Keywords: Discrete Choice Experiment**,** Multi-Criteria Decision Analysis, decision support, heavy menstrual bleeding, prostate cancer screening

Introduction

In their recent systematic review, Soekhai and colleagues identified no less than 301 health-related empirical Discrete Choice Experiments (DCEs) published between 2013 and 2017 [1]. They noted that the number of DCEs had continued to increase, relative to the number in similar prior surveys [2], with broader areas of application and increased geographic scope. However, they also found there was no description of how health-related discrete choice experiments were being employed, and the number of studies testing external validity remained small. 'Key barriers to their wider use in policy include concerns about the robustness and validity of the method and the quality of applied studies… future research should focus on identifying and resolving the methodological and practical challenges involved in validity testing.., ~~and on guiding the incorporation of DCEs into actual decision making in healthcare..~~.' Vass and Payne also endorse a

[1] Corresponding author: jack.dowie@lshtm.ac.uk. London School of Hygiene and Tropical Medicine 15-17 Tavistock Place, London, WC1H 9SH UK

Formatted: Default Paragraph Font, English (United States)

primarily methodological explanation, pointing to several key areas that need to be addressed in order to demonstrate that discrete choice experiments are an appropriate and valid stated preference elicitation method [3].

Unfortunately, the dearth of empirical studies on the extent to which, and ways in which, DCEs are being employed - or not - makes proposed explanations for their non-use, including the favoured methodological and reporting ones, largely speculative.

This future vision paper reflects an alternative speculative explanation of the limited impact of DCEs on decision making, which is indeed remarkable in the light of the number produced and published, and the resources consumed in their production. It is suggested here that lack of detail provided, or methodological weakness, is probably of marginal relevance. The main explanation is likely to be found in the absence of a way to bridge the research-practice gap created by the conceptualisation of the DCE task as a purely research one - where the researchers' obligations are limited to 'presenting' the results clearly, so that the decision makers can fully 'understand' them, but purely as 'information' support, lacking any decision-relevant translation. In both policy and clinical settings this alternative hypothesis leads to bridging the research-practice gap as the route to greater impact of, and higher Return of Investment on, DCEs.

Method

Bridging the gap by an add-on that enables preference variation to be undertaken, can address the gap in the policy context. But it can go much further in the clinical setting, transforming the DCE results into ones of direct relevance to the individual decision maker. Our focus in this paper is on this latter, clinical, situation, where we demonstrate how a DCE can be translated into an online personalised decision support tool for use in person-centred care.

Not all DCEs can be made clinically relevant. While they vary in numerous other ways, particularly in study design and methods of analysing the results, for the present purpose we are primarily interested in the reality of their key components. In the policy context and in relation to future development of interventions or services, especially ones that will be influenced by Willingness to Pay, hypothetical components are often essential. However, to be relevant in today's clinical practice the alternatives need to be real (even if they are unlabeled in the DCE study), the attributes need to be real (this is almost always the case because of the precursor study of what matters to individuals) and the attribute levels need to be real and assigned/assignable to options.

Given its cognitive and logistic demands, person-centred decision making requires an interactive, computer-based decision support tool (DST), capable of integrating the individual's preferences and producing an *opinion* in the form of a complete set of personalised option scores. We use Multi-Criteria Decision Analysis (MCDA) as the basis for developing such a personalised DST from the DCE.

The type of MCDA most compatible with ethical person-centred decision making and most able to ensure informed and preference-based consent is the value-based,

Formatted: Indent: First line: 0.63 cm

Formatted: Indent: First line: 0.63 cm

Formatted: Indent: First line: 0.63 cm

Formatted: Indent: First line: 0.63 cm

compensatory model [4]. This takes the form of a 'weighted-sum' calculation, multiplying the personalised numerical ratings for the performance of each option on each criterion by the relative weight assigned to the criterion by the person, and then summing these weighted scores to get an overall preference-sensitive score for each option. The different scales on which the various criteria are rated can produce scores that do not appropriately reflect the assigned weights, and we address this through 'idealisation' of the Ratings: the best- rated option on a criterion is re-assigned a rating of 1 and the other options percentaged in relation to it.

The opinion produced by the DST is not a *medical* opinion and will normally be *preliminary, a starting point,* to a deliberative decision–making phase. The tool includes an explicit disclaimer to the effect that it cannot reflect information unavailable to the person at the time of their engagement with it (e.g. a future test result, or a physical examination that may establish contraindication/s to particular option/s).

Results

We translate two selected DCEs into personalised DSTs by way of an MCDA-based add-on that uses the Annalisa implementation of the technique [5].

The DCE for Heavy Menstrual Bleeding [6] has two options - Intrauterine System (LNG-IUS) and Endometrial Ablation (EA). Each of its seven criteria has two levels, each being associated with one or other of the options as its performance Rating. In five cases the levels are binary (the attribute is present or absent), in two they are empirical probabilities.). The DCE for Prostate Cancer Screening [7] has three options (two screening programmes and no screening) and 6 criteria. The criteria become five after we drop the out-of-pocket cost attribute to minimise loss of generality across settings. Each criterion has only one level in the form of a probability. Results are obtained for three age groups, but in this illustrative translation, only those for 50-59-year-olds are presented, leaving those for the 40-49 and 60-69-year-olds aside.

In each case the sole task for the individual engaging with the DST is to indicate their relative importance Weighting for each criterion, doing so on a 0-100% scale, where 0 indicates of no importance and 100% of extreme importance. The five responses are summed and percentaged to give and display the set of provisional criterion weights that add to 100%. In this interactive tool, they may be changed by the cursor on inspection of their graphical display. More complex weight elicitation procedures, such as DCEs, or swing weights, may have greater normative appeal than Visual Analog Scales, but lack either individual applicability or practicality in the typical time and resource scale applicable in clinical practice.

The following screen captures are those for a hypothetical individual responding to each of the tools. After seeing their own result displayed, the respondent is shown the equivalent screen based on the Weightings in the underlying DCE, and can thereby compare themselves with the 'average person'.

Formatted: Indent: First line: 0.63 cm

Formatted: Indent: First line: 0.63 cm

The woman in Figure 1 rated the importance of the seven criteria (out of 10) at 6, 7, 4, 6, 5, 8 and 9 (reading left to right). Expressed as percentages, these become her displayed Weightings of 13, 16, 9, 13, 11, 18 and 20 and (after idealisation of the Ratings) produce the Scores of 36 for LNG and 84 for EA. In comparison, the average group percentage Weightings from the DCE (not shown) were 11, 12, 3, 1, 14, 19 and 40, producing Scores of 16 for LNG and 88 for EA. EA is the opinion of the DST in both cases, but by a much smaller margin for this individual woman, mainly as a result of her lower weight for Hormone avoidance. In this example, we use the program option of displaying the Ratings as chances of avoiding the criterion rather than experiencing it. (In Annalisa, the attributes can be oriented either positively or negatively, but all must be in the same direction and the Ratings made compatible with this orientation.)

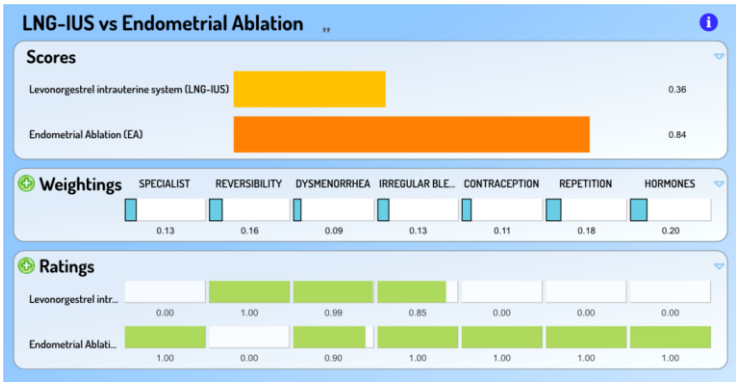
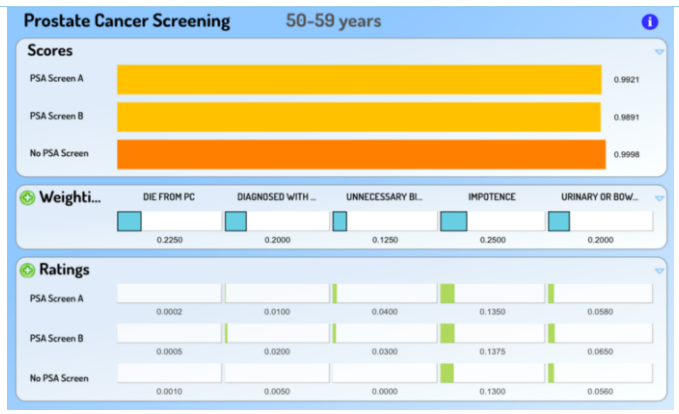


Figure 1: Screen capture from Heavy Menstrual Bleeding Decision Support Tool



Formatted: Centered

Formatted: Font: Bold

Formatted: Font: 20 pt

Formatted: Centered

Figure 2: Screen capture from Prostate Cancer Screening Decision Support Tool

The man in Figure 2 rated the importance of the five criteria (out of 10) at 9, 8, 5, 10, and 8 (reading from left to right). Expressed as percentages these become his displayed Weightings of 22.5, 20, 12.5, 25, and 20 and (after idealisation of Ratings) produce the Scores of 99.21 for PSA A, of 98.91 for PSA B and of 99.98 for No Screening. The group average percentage Weightings from the DCE (not shown), were 31.4, 1.6, 13.4, 11.0, and 42.5, producing scores of 99.30 for PSA A, 99.06 for PSA B and 99.98 No Screening. No Screening is optimal in both cases and almost dominant. In this example, we use the software option of displaying the Ratings as chances of not avoiding.

The MCDA-based decision support tool incorporating the data from the DCE for Heavy Menstrual Bleeding is available at ale.rsyd.dk with Survey ID 1496. The MCDA-based decision support tool incorporating the data from the DCE for Prostate Cancer Screening is available at ale.rsyd.dk with Survey ID 1495. The tools are on open access so only anonymous and non-confidential data should be entered.

No position is taken here on the absolute quality of the underlying DCEs - or *a fortiori* the present DST add-ons. ~~For instance, we believe a DST should always contain a 'no change'/'watchful waiting' type option, and an 'Impact on Others' type criterion will be relevant in many cases.~~ However, these two DCEs are representative enough and of sufficient quality to justify and support the present exercise.~~. The bridging DSTs are consequently far from perfect normatively, but any evaluation of them should be empirical, not normative and respect their purpose and provenance~~. ~~Whether a DCE has~~The ~~'external validity' is a key issue, and the~~ rare published studies suggest most DCEs have moderate 'external validity', if any. Soekhai et al. report only 2% undertook a check of this sort [1].

Beyond its basic provision of a personalised preliminary opinion, the Annalisa display can make a substantive contribution to better communication to both clinicians and person/patient. It displays the DCE Ratings data in a way that permits those for each option to be compared side-by-side, along with the Weightings and resulting Scores.

Discussion

A scan of recent DCE studies reveals frequent claims that the group-level results can somehow contribute to or facilitate better *clinical* decision making. Often, however, there are only ambiguous indications of *how* this could happen, often accompanied by a worrying lack of precision in the use of 'patients' preferences' and 'patient's preferences' [8]. As just one example, Johnson et al. [9] note that the US FDA and the European Medicines Agency are 'considering ways to incorporate patients' perspectives into the regulatory framework for drug and device approvals', but also suggest that 'the data generated from DCEs can provide evidence of the perceived value of treatment features (e.g. convenience and acceptable levels of risk) from a patient's perspective and can be used as supplementary and/or complementary information for physicians to

consider when prescribing therapies‘ (p.388). Placing the apostrophe after the ‘s’ is correct in the first quote, which concerns regulatory decisions, but not in the second.

Most DCE authors are clear that it is the individual’s preferences which are central to decision making in their context, including those responsible for the PSA study.

> “Our results can be used by both clinicians and patients to facilitate informed discussions of relevant benefit and downsides of PSA screening for an individual man. Future research should examine whether feeding back this information from DCEs compared to other values clarification methods helps men in their decision making in this complex area.” [7] (p.3133-4)

We hope that our proposal for an enhanced DCE, going well beyond simple feedback of the results, will be seen as a promising route to value-clarifying decision support. This will avoid the danger of individual preferences being subjected to an epidemiological version of the ecological fallacy, treated as the dependent variable in a regression equation based on sociodemographic characteristics [10].

Conclusion

We provide generic proof of a method for translating ‘real’ DCEs into personalised Decision Support Tools for clinical use, along with two substantive examples. Research into the impact of DCEs with and without such add-ons is an obvious next step. However, use of the technique should not await the results of such evaluation, since the add-on tool can enable each and any stakeholder in the decisional context to access the useful information in any extant DCE in a decision-relevant and preference-sensitive, way.

Funding

The software used was installed at https:// ale.rsyd.dk as part of 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 Annalisa software, but does not benefit from non-commercial use.

References

[1] V.Soekhai, E.W. de Bekker-Grob, A.R. Ellis, C.M. Vass, Discrete Choice Experiments in Health Economics: Past, Present and Future, *Pharmacoeconomics* (2018), Nov 3 (E-pub ahead of print)

[2] E.W. de Bekker-Grob, M. Ryan, K. Gerard, Discrete choice experiments in health economics: a review of the literature, *Health Economics* **172** (2012), 145–72.

[3] C.M. Vass, K. Payne, Using Discrete Choice Experiments to Inform the Benefit-Risk Assessment of Medicines: Are We Ready Yet? *Pharmacoeconomics* **35** (2017), 859–66.

[4] K. Marsh, M. Goetghebeur, P. Thokala, R. Baltussen. *Multi-Criteria Decision Analysis to Support Healthcare Decisions*. Springer International Publishing, 2015.

[5] J. Dowie, M.K. Kaltoft, G. Salkeld, M. Cunich, Towards generic online multicriteria decision support in patient-centred health care, *Health Expectations* **18** (2015), 689–702.

[6] M.J. van den BrinK, P. Beelen, M.C. Herman, N.J.J. Claassen, M.Y. Bongers, P.M. Geomini, et al. Women’s preferences for the levonorgestrel intrauterine system versus endometrial ablation for heavy menstrual bleeding, *European Journal of Obstetrics & Gynecology and Reproductive Biology* **228** (2018), 143–7

[7] K. Howard, G.P. Salkeld, M.I. Patel, G.J. Mann, M.P.Pignone. Men’s preferences and trade-offs for

prostate cancer screening: a discrete choice experiment, *Health Expectations* **18** (2015), 3123–35.
[8] M.K. Kaltoft, J.B. Nielsen, G. Salkeld, J. Dowie, Can a Discrete Choice Experiment contribute to person-centred healthcare? *European Journal for Person Centered Healthcare* **3** (2015), 431–7.
[9] P. Johnson, T. Bancroft, R. Barron, J. Legg, X. Li, H. Watson, et al. Discrete Choice Experiment to Estimate Breast Cancer Patients' Preferences and Willingness to Pay for Prophylactic Granulocyte Colony-stimulating Factors, *Value in Health* **17** (2014), 380-9
[10] M.K. Kaltoft, J.B. Nielsen, G. Salkeld, J. Dowie, Preferences cannot be treated as epidemiological characteristics in person-centred care : a riposte. *European Journal for Person Centered Healthcare* **4** (2016), 6–9.