# A comparison of discrete choice experiments and constant sum paired comparisons for the elicitation of societal preferences

**ABSTRACT**

**Background**: Stated preference elicitations of societal welfare require respondents to consider interpersonal trade-offs and how they feel about others in specific conditions.

**Objective**: To compare discrete choice experiment (DCE) and constant-sum paired comparison (CSPC) questionnaires for eliciting societal preferences for healthcare resource allocation.

**Methods**: Respondents were asked to allocate a fixed budget between two patient groups in 10 scenarios, including one repeated task. DCE allocated the entire budget to one group, while CSPC allocated budget percentages between groups. Questionnaires were compared in terms of completion rates, consistency, non-trading and attribute importance. CSPC distributional preferences are also reported.

**Results**: A significantly greater proportion completed the DCE compared to the CSPC questionnaire. There was no significant difference in the proportions that rated the tasks somewhat or extremely difficult. Consistency was higher with DCE. The incidence of non-trading was low and not significantly different between surveys. Fewer than 10% of all CSPC allocations explicitly equalised outcomes or resources while 18% maximised the allocation to one group; 11% of respondents maximised in >50% of tasks. Attribute coefficients were consistent with expectations, but relative importance rankings differed between the surveys.

**Conclusions**: Although the significantly lower completion rate suggests the CSPC questionnaire was less acceptable in some aspect(s), it revealed an unexpected willingness among respondents to maximise allocations. CSPC tasks can also generate more strongly-ordered and statistically efficient preference data than DCE. Richer and more statistically efficient preference data with CSPC must be weighed against better response rates and consistency (‘respondent efficiency’) with DCE.

# Introduction

Stated preference (SP) methods can be used to measure preferences for efficiency and equity in healthcare. Descriptive SP approaches, such as rating scales and simple ranking exercises, can assess the direction of preferences but not the strength of preferences. Choice-based SP approaches encourage respondents to trade-off between alternatives on the basis of attributes and levels and allow the measurement of strength of preference as well direction of preference.(1,2) But whereas conventional elicitations of individual welfare ask respondents to judge how they would feel about being in a certain condition or health state, elicitations of societal welfare require respondents to consider interpersonal trade-offs and how they would feel about *others* a certain condition. As such, Menzel and others argue that societal preferences should be measured using an instrument that allows explicit consideration of the trade-offs inherent in allocating healthcare resources and advocate the person trade-off (PTO) approach.(3–5)

The PTO asks respondents how many persons in state Y they would consider equivalent in terms of social value to a fixed number of persons in state X. The ratio of persons in state Y to persons in state X represents relative social values.(3) While some argue that PTO has an intuitive appeal for the elicitation of societal equity weights (2,3), Green notes that respondents often find the task complex, difficult, and occasionally offensive.(6) This is supported by Damschroder *et al*, who found that between 21 and 91 percent of PTO respondents were unwilling to make person trade-offs despite clear differences between programs.(7) This unwillingness to trade-off persons may reflect Mullen’s argument that “theoretical validity does not always coincide with acceptability, people’s comprehension and even people’s value systems.”(8)

With this argument in mind, two alternative choice-based SP methods are compared here: discrete choice experiments (DCEs) and indirect constant-sum paired comparisons (CSPC). The details of the two instruments will be outlined in the following methods section, along with specific details of the survey methods and the criteria used in the comparisons. The results section will highlight the performance of each instrument in terms of these criteria and finally the concluding section will discuss the relative strengths and weaknesses of each.

# Methods

Discrete choice experiments (DCEs) ask respondents to identify their preferred choice from a set of two or more alternatives based on specific attributes and levels, and encourages respondents to consider the trade-offs between attributes. DCEs are consistent with random utility theory (RUT), which assumes that the latent utility of alternative *i* (U*i*) to a respondent is derived from an observed, systematic component (v*i*) and an unobserved component (ε*i*):

U*i* = v*i* + ε*i*

Although the respondent is assumed to be rational, randomness is introduced from the perspective of an observer because they are limited to inferring utility from v*i* rather than directly observing U*i*.(1) In a DCE, the difference in latent utility between alternatives is assumed to be proportional to the probability of choosing each alternative. However, the preference data derived from a single DCE task are ‘weakly ordered’ – although the alternative that was chosen can be ranked relative to the alternative that was not chosen, the strength of that preference cannot be determined from that single choice. More responses, to different choices from the same respondent, or to the same choice from different respondents, are required, making DCE relatively statistically inefficient.(9) In their favour, DCEs are conceptually straightforward tasks and are relatively easy for respondents to grasp. DCEs are increasingly common in health economics applications and have previously been used in elicitations of societal values for priority setting.(10–12) An example DCE is shown in Appendix A.

Constant-sum paired comparison (CSPC) tasks ask respondents to allocate a fixed budget between two alternatives, described as in a DCE.[[1]](#footnote-1) The budget is usually presented as points or percentages as realistic monetary sums for healthcare programs are likely to be unfamiliar to respondents and may compromise their ability to make realistic allocations, while trivial sums risk respondents not taking the task seriously.(2,8) Ryan (2) finds that the approach has an apparent simplicity and intuitive appeal in its explicit consideration of trade-offs and strength of preferences, while Louviere, Hensher and Swait (1) argue that CSPC is consistent with RUT and can yield cardinal utility measures if it can assumed that differences in the budget allocation reflect differences in latent utility. As these differences reflect strength of preference, CSPC tasks tend to produce more strongly ordered preference data than DCE, giving them an advantage in terms of statistical efficiency over DCE.(1) Schwappach and Strasmann also argue that the ability of respondents to avoid extreme distributions and absolute discrimination against specific alternatives makes “allocation of points” tasks particularly useful in eliciting preferences for the allocations of societal healthcare resources.(13) Relative to DCE, CSPC has had only limited use in health economics applications, but it has been successfully used to elicit societal preferences for efficiency and equity in healthcare.(13–17) Ratcliffe (14) reported that only 2.3 percent of CSPC respondents exhibited non-trading behaviour (i.e. dominant or strictly egalitarian preferences), and only 14 percent of respondents rated the CSPC elicitation as ‘very difficult’, while Schwappach and Strasmann (13) found that CSPC could be successfully administered to a sample of the general population via the internet with fewer than 10 percent non-traders. An example of a CSPC task is shown in Appendix A.

## Survey Methods

Individuals were invited to participate via a mass email to students at The University of Sheffield, Sheffield, UK, posters and electronic announcements to students, staff and faculty at Dalhousie University, Halifax, Canada, and posters and electronic announcements to physicians and staff in the Capital District Health Authority, Halifax, Canada. All aspects of the surveys and the analysis were approved by The University of Sheffield, School of Health and Related Research Ethics Committee and the Capital District Health Authority Research Ethics Board.

Questionnaires were administered via the internet. Face-to-face administration of SP elicitations has significant benefits, including the ability to thoroughly explain the objectives of the survey and to provide timely feedback to respondents, but it is also costly, time consuming and can often lead to small or selective samples.(18) Relative to less personal elicitation formats, there is also evidence that face-to-face interviews tend to increase ‘social desirability’ or ‘yea saying’ biases, where respondents offer the answer they perceive to be socially ‘correct’ or that will please the interviewer, rather than their true preference.(19,20) There is also evidence that web-based administration can minimise social desirability bias in eliciting socially sensitive information.(21)

Potential respondents were randomised to either the DCE or CSPC questionnaire using a random number algorithm. Sixty percent of potential respondents were assigned the CSPC questionnaire in order to compensate for a lower expected completion rate; the remaining 40 percent were assigned the DCE questionnaire. As each potential respondent followed the link and was assigned a questionnaire, a record was written to an SQL database indicating the assigned survey. This database was used as the denominator in calculating the overall response rate for each survey. Each respondent only saw one survey design.

Respondents were asked to imagine themselves as a societal decision maker responsible for allocating a fixed budget between two alternative healthcare programs. They were told that both programs had the same overall cost and that the budget was large enough fund one program or the other, but not both. The precise budget and the cost of the programs were not specified. Drawing on the results of an empirical ethics review (22), the programs were described in terms of the average age of the patients, their health-related quality of life before and after treatment and the number of patients in each group. To facilitate the use of quality-adjusted life years (QALYs) as a measure of aggregate outcomes, life expectancy without/before treatment and life years gained with/after treatment were also included despite ambiguous evidence around societal preferences for duration and the QALY itself.(22–24) Each of these six attributes had three possible levels (see Table 1). To avoid situations where respondents were required to make numerical calculations, the total number of QALYs gained with each program was calculated[[2]](#footnote-2) and presented as an additional attribute in each alternative, along with a brief conceptual overview of the QALY. This was intended to avoid potential calculation errors and to allow each respondent to base their decisions on the same information.(25) The concept of cost-effectiveness was not presented or explained to respondents, but given that both programs had the same cost, respondents could have inferred that the program associated with the greater number of QALYs gained was the more cost-effective option. Preliminary interviews and informal focus groups were used to refine the wording and presentation of the DCE and CSPC tasks.

The DCE questionnaire asked respondents to allocate the entire budget to their preferred group, while the CSPC questionnaire asked respondents to allocate budget percentages between the two groups by moving a slider. Respondents could allocate 100 percent of the budget to program A or program B, or to some combination of the two. The number of patients treated and total QALYs gained changed in proportion with the budget as the respondent moved the slider (e.g. a 25 percent budget allocation meant 25 percent of the potential patients could be treated). Respondents to the DCE questionnaire did not have an option to indicate indifference or equality, but there was an option to indicate no answer. There was no opt-out or no answer option in the CSPC but respondents could indicate indifference (or an egalitarian preference) with a 50%-50% budget allocation. Respondents could drop-out at any stage and data was only collected from those respondents who completed the entire questionnaire.

An optimal fractional factorial experimental design was constructed using Kuhfeld’s SAS macros.(26) The design started with the full factorial candidate design and excluded illogical combinations where the net QALY gain with treatment was negative. Combinations where health state and life expectancy were unchanged before and after treatment were also excluded. Of the 729 combinations in the full factorial design, 135 (19%) were excluded as illogical. From this set of potential scenarios, a D-efficient 18 choice set design with 2 alternatives in each choice set was constructed and divided into two blocks of 9 choice pairs each.(27) One block was used for the DCE questionnaire and the other block was used for the CSPC questionnaire. Including one repeated task, each questionnaire presented a series of 10 choice tasks to each respondent. As each unique choice task contributes 1 degree of freedom, this design provided 9 degrees of freedom for each analysis.(28)

Following the choice tasks, respondents were asked to rate the importance of each attribute (including distributional concerns) in their choices on a 0 to 10 scale, and to rate the difficulty of understanding the tasks and of answering the tasks on 7-point scales ranging from extremely easy to extremely difficult. Respondents were also asked to indicate their gender and age group and to voluntarily identify themselves as a governmental decision maker or academic expert, a physician, and/or a frequent healthcare user (≥12 contacts in the past 12 months). These categories were not mutually exclusive. Respondents not self-identifying as one or more of these stakeholder groups were assumed to be members of the general public.

## Survey Comparisons

Responses from the two surveys were compared on a number of dimensions, intended to measure the acceptability of the instruments to respondents and their ability to elicit valid preference data. These included the completion rate, the respondent-rated ease of understanding and answering the questionnaires, internal consistency, and the incidence of non-trading behaviour. Although the estimation of respondent preferences was not a primary objective of these questionnaires, an analysis of the choice responses was also conducted in order to compare the preference information derived from the two questionnaires.

Differences in the survey completion rates and stakeholder and gender proportions were tested using a two-sample Z-test of proportions. Age group proportions were tested using a test of independence. On the assumption that the randomisation algorithm assigned an equal proportion of each age, gender and stakeholder subgroup to each questionnaire, differences in these proportions were taken to indicate a differential drop-out rate among these groups. The proportions of respondents who indicated that they found the questionnaire ‘somewhat difficult’ or ‘extremely difficult’ to understand or to answer were also compared using a two-sample Z-test of proportions. Internal consistency was measured using a repeated task included in each questionnaire: the two alternatives from task 3 were reversed and re-presented as task 8. Internal consistency requires that respondents should prefer the same program in the repeated task as in the original. To compare the two instruments on a common basis, CSPC responses were transformed to discrete choices based on the alternative to which a respondent allocated the majority of the budget. Equal budget allocations were also allowed. Consistency in responses between the original and the repeated task was assessed using McNemar’s test of independence in paired data.(29) P-values were adjusted for simultaneous comparisons using Hommel’s method,(30) although p-values to identify the significance of parameters in the DCE and CSPC models were *not* adjusted in order to allow for the broadest possible inclusion of model parameters.(31)

Dominant preferences were defined as “individuals who always choose the scenario where *x*1 is greater than , no matter what the level of the other attributes.”(32) Although such preferences are valid, they violate the assumptions of compensatory decision making and an additive utility function that underlie choice-based stated preference methods.(9) Lexicographic preferences cannot be represented by a utility function, and as no trading takes place, marginal rates of substitution have no meaning.(9,32)

To test for dominant preferences, a set of flags was created for each alternative in each choice task. The flags indicated whether or not an alternative presented the greater, or dominant, level of each attribute. For example, if one alternative presented younger patients than the other, that alternative was flagged as dominant in the age attribute; the attribute flag for the paired alternative was set to zero. Similarly, if one alternative was associated with greater life year gains than the other, that alternative was flagged as best in the life years gained attribute. There were a total of seven flags for each alternative: age, initial utility, initial life expectancy, final utility, life years gained, (potential) number of patients treated and (potential) number of QALYs. CSPC responses were transformed to discrete choices and the flags were set based on the potential number of patients that could be treated and the potential number of QALYs gained if 100% of the budget were allocated to each alternative. The correlation between choice and each attribute flag was taken as a measure of the degree of that attribute’s dominance in each respondent’s choices. As the outcome of interest was the correlation between pairs of nominal variables (i.e. choice and attribute flag), Kendall’s tau was used as the measure of correlation.(33) A respondent that always chose the alternative with, for example, the younger patients, would have a choice-attribute correlation coefficient of 1.0 in the age attribute. Which end of the attribute scale the respondent considered ‘best’ is not critical, as in this example correlation would -1.0 if in fact they preferred older patients to younger patients and always chose the alternative with the older patients, although it is important to recognise that this only holds where preferences are monotonically increasing or decreasing over the attribute.

As the optimal fractional factorial design used in the surveys presented only a small subset of potential attribute combinations, it is not possible to say with certainty that observed cases of perfect choice-attribute correlation would hold across all possible scenarios.(32) Therefore, to support the identification of non-trading behaviour, each respondent’s self-rated attribute importance scores were converted to rankings, allowing for ties as individuals could potentially give each attribute the same score, and compared to their choice-attribute correlations. Individuals with a perfect 1.0 choice correlation with a particular attribute, who also rated that attribute as most important, were assumed to be confirmed as non-traders. The proportion of non-traders was compared across the two questionnaires using a two-sample Z-test.

Given the limited degrees of freedom available in the design, the models assumed monotonic preferences and only linear main effects were estimated. The analyses were performed with R 2.13.1 using the mlogit 0.2-1, censReg 0.5-6 and plm 1.2-7 packages. Responses to the repeated task were excluded from the analyses, as were ‘no answer’ responses from the DCE questionnaire. The QALYs gained attribute, as a linear combination of initial and final utility, life years gained and number of patients treated, was also excluded from the analyses in order to avoid multicollinearity.

The DCE choice tasks were analysed using a multinomial logistic (MNL) model, which defines the probability of choosing alternative *i* over alternative *j* as:

** [1]

Where v*i* and v*j* are the observable components of utility associated with alternatives *i* and *j*. For each alternative, observed utility was defined as:

[2]

Where the parameters are the levels of each attribute in alternative, the β’s represent the marginal change in observed utility associated with a one-unit change in each of the attributes, and ε is a Gumbel (extreme value type 1) error term.(9) This model is not able to account for the panel structure of the data.

CSPC responses were analysed using a double-bounded random effects tobit model to account for the censored dependent variable and the panel structure of the data:(34)

[3]

Where is the budget allocated to Program B less the budget allocated to Program A by respondent *i* in choice *t*. If 100% of the budget was allocated to Program B, if 100% was allocated to Program A, if the budget was allocated 50%-50%, . The parameters are the differences in attribute levels between Program B and Program A, the β’s represent the marginal change in the budget difference associated with a one-unit change in each of the attributes, is an individual-specific error term and is a stochastic error term.

Consistent with a RUT interpretation of the DCE and CSPC, model coefficients were assumed to represent the marginal change in latent utility (or in the difference in latent utility) associated with a one-unit change in each attribute.(35) These coefficients were compared on the basis of relative attribute importance, or the relative contribution of each attribute to overall utility.(36) In the DCE, each attribute’s contribution to latent utility was calculated based on the most preferred and least preferred level of that attribute:

[4]

where is the utility associated with the *most* preferred level of , is the utility associated with the *least* preferred level of , and is the net difference in latent utility. This attribute-specific contribution was then divided by the difference in overall utility/budget between the ‘best’ scenario, based on the most preferred level of each attribute, and the ‘worst’ scenario, based on the least preferred level of each attribute:

*Relative Importance of xi* =

Where *v* is defined as in equation [2] above. The larger this relative contribution, the more important an attribute was to latent utility. The calculation was essentially the same for the CSPC using equation [3], where the *xi*’s represented the minimum and maximum differences in levels and *v* was the difference in the budget allocation. Marginal rates of substitution between the different attributes and life years gained were also calculated, but given the non-representativeness of the convenience sample these results are not emphasised.

Finally, as the ability to explicitly elicit distributional preferences is a key feature of the CSPC, the distribution of budget allocations is also presented. Consistent with the identification of non-trading behaviour discussed above, respondents demonstrating strictly egalitarian preferences in the CSPC questionnaire (i.e. an equal allocation of resources across all tasks) were considered non-traders if they also ranked the distribution of resources as the most important factor in their choices.

# Results

A total of 604 individuals began a questionnaire: 348 (58%) were randomised to the CSPC questionnaire and 256 (42%) were randomised to the DCE questionnaire. Completion rates and respondent characteristics are shown in Table 2. Asignificantly greater proportion of individuals completed the DCE compared with the CSPC questionnaire. There were no significant differences in the age group or gender distributions, or in the proportion of respondents who identified themselves as doctors or frequent healthcare users, but a significantly lower proportion of respondents identified themselves as decision makers in the CSPC questionnaire.

## Respondent-rated difficulty

As shown in Table 3, there was no significant difference between the two surveys in the proportion that rated the tasks “somewhat difficult” or “extremely difficult” to understand among all respondents who submitted a completed questionnaire. A greater proportion of decision makers found the CSPC “somewhat difficult” or “extremely difficult” to understand compared to the DCE, but this difference was not statistically significant. There were no notable differences in the difficulty of understanding in the other stakeholder subgroups. Table 4 shows no notable or statistically significant differences within respondent subgroups in the perceived difficulty of answering the DCE or CSPC.

## Consistency in repeated tasks

In the DCE survey, 148 out of 154 respondents (96%) preferred the same program (including 3 ‘no answers’) in the original and the repeated task (p=0.22; adjusted-p=1.00). After converting budget allocations to discrete choices, 119 out of 150 CSPC respondents (79%) allocated the majority of the budget to the same program or preferred an equal allocation of resources in both tasks (p=0.06; adjusted-p=0.96). The mean budget allocation to program B in the original task was 27 percent versus 19 percent in the repeated task (difference=-8%, p<0.001; adjusted-p<0.001). The mode budget difference was zero.

## Dominant preferences and non-trading behaviour

Figure 1 suggests only moderate correlation between choice and dominant attributes in either survey. Excluding three individuals who always chose ‘no answer’ in the DCE questionnaire, the proportion of respondents who demonstrated potentially dominant preferences – that is, had at least one choice-attribute correlation coefficient of 1.0 – was 13 percent in the DCE questionnaire (20/151) and 18 percent in the CSPC questionnaire (27/150); this difference was not significant (p=0.33, adjusted-p=1.00). Among this potentially non-trading subset, 10 DCE respondents (7%) and 13 CSPC respondents (9%) with perfect choice-attribute correlations also ranked that perfectly correlated attribute as the most important attribute in their choices, confirming their status as non-traders. This difference was not statistically significant (p=0.65, adjusted-p=1.00).

Due to an error in the SQL database, attribute importance ratings were not recorded for the ‘number of patients’ attribute. As three additional DCE respondents had a perfect correlation in this attribute that could not be confirmed against their attribute rating, it is possible that up to 13 DCE respondents (9 percent) may have been confirmed as non-traders.

## Attribute weights

Table 6 presents the model coefficients, coefficients of variation, p-values, marginal rates of substitution in terms of life years gained, and the relative importance weights and rankings for the DCE and CSPC models. All coefficients were significant at the 95% level with the exception of initial life expectancy in the MNL model. This attribute was included in the final model in order to present the broadest possible model. Excluding it worsened the log-likelihood ratio slightly and had little effect on AIC. Figure 1 illustrates the relative attribute importance weights from the two models.

## CSPC distributional preferences

Figure 2 shows that 18 percent of all CSPC responses maximised the budget allocation to program A (0 percent to Program B in Figure 2) or program B (100 percent to Program B in Figure 2), while 7 percent of responses equalised the budget between the two programs (50% to Program B in Figure 2). At the individual level, 10 percent of respondents maximised the budget to one program or the other in 50 percent or more of their choices. One respondent equalised the budgets in the majority of their choices and ranked the distribution of resources as the most important factor in their choices, but as they did not equalise budgets in 100 percent of their choices they were not categorised as a non-trader. No respondents equalised patients or QALYs in the majority of their choices.

# Discussion

The respondent-rated difficulty of the two surveys was strikingly similar, with only a small minority rating the tasks as difficult to understand but a strong majority rating the tasks as difficult to answer. The directions of these ratings are somewhat reassuring as they suggest that respondents were able to understand the tasks and that the key challenge was ethical rather than cognitive. The proportions reporting the questionnaires to be difficult to understand or to answer were similar to proportions reported by Green in a DCE of societal preferences.(10) As difficulty ratings were only collected from respondents that completed the questionnaires, however, the difficulty ratings are most likely biased downwards as individuals that found the surveys exceedingly difficult would most likely have dropped-out before completion.

The completion rate in the CSPC survey was significantly lower than the DCE survey, suggesting that it was less acceptable in some respects. This response rate was similar to that reported by Ratcliffe (38%) in her application of CSPC.(14) This lower completion rate despite similar difficulty ratings supports the idea of a completion bias in these ratings. It is interesting to note that as a group, decision makers expressed the greatest difficulty in answering both questionnaires and had significantly lower completion rates in the CSPC survey compared to other stakeholder subgroups. This was somewhat surprising given that CSPC is arguably a more realistic reflection of real-world decision making tasks. Keeney notes that a desire to calculate a ‘correct’ value trade-off despite the absence of any externally correct judgements is a common error in decision making (37), and such a phenomenon may be particularly strong among self-identified decision makers. Qualitative work to confirm the reasons for this difficulty and the lower CSPC completion rates would be of interest.

Respondents to the DCE questionnaire were more consistent in preferring the same alternative in the repeated task compared to respondents to the CSPC questionnaire. It is worth noting that the mode difference between the original and repeated actual budget allocation was zero, and that the mean and median budget differences were less than 8 percent. The absolute size of a meaningful difference is not clear in cases where this difference does not affect which program receives the majority of resources. In retrospect, the task that was randomly chosen as the repeated task in the DCE may have contributed to the strong consistency observed in that survey, as 95 percent of respondents preferred the same alternative in the original DCE task, whereas only 77 percent preferred the same alternative in the original CSPC task. A less ‘obvious’ choice may have been a better test of consistency in the DCE questionnaire.

The proportion of dominant preferences in both surveys, considering perfect choice-attribute correlations as well as the strictly egalitarian budget equalisers, was less than the 20-90 percent proportions reported by Damschroder (7) using a PTO. Both surveys had a greater incidence of dominant preferences than the 2.3 percent reported by Ratcliffe (14), but less than the 13 percent of confirmed non-traders reported Schwappach (15), both of which used CSPC. It is worth noting that the non-traders in Schwappach’s CSPC elicitation were predominantly egalitarian – a preference not observed here. The low incidence of non-trading behaviour suggests that respondents were not resorting to simplifying heuristics in the face of challenging ethical decisions.

The directions of the coefficients and the marginal rates of substitution in the two models were consistent with each other and with previous reviews of societal preferences (22–24) with the exception of initial life expectancy, where DCE respondents preferred patients with lower initial life expectancy, while contrary to expectations, CSPC respondents preferred patients with greater initial life expectancy. It is important to note, however, that although the coefficient on initial life expectancy in the DCE model was more consistent with expectations, it was not statistically significant.

Rankings of the relative importance weights the surveys appeared to focus respondents’ attention on different attributes. DCE respondents gave relatively more weight to patient characteristics such as final health state and patient age, although the number of life years gained was also important. CSPC respondents gave more weight to quantitative program attributes, particularly the number of patients treated and the number of life years gained. It is possible that these respondents took the fact that the number of patients treated and total QALYs gained changed as they moved the slider as a cue to focus on these attributes.

The reasonably high consistency, along with a relative low incidence of confirmed non-trading, suggests both instruments are eliciting valid preference data. CSPC, though, has the advantage of being able to explicitly elicit information about distributional preferences, particularly preferences for equality in the allocation of resources. The CSPC survey reported here demonstrated an unexpected willingness among respondents to maximise budget allocations to a single group. This challenges previous studies that found a general aversion to such extreme distributions (14,15) and highlights the value of explicitly testing for such preferences. To the extent that CSPC budget allocations reflect strength-of-preference, CSPC tasks will also tend to generate more strongly-ordered and statistically efficient preference data. However, it is clear from the significantly lower response rate that respondents, and particularly decision makers, found the CSPC less acceptable in some respect(s). Ultimately, potentially richer and statistically more efficient preference data with CSPC must be weighed against better completion rates and consistency (‘respondent efficiency’) with DCE.

***4,909 words***

**REFERENCES**

1. Louviere J, Hensher DA, Swait J. Conjoint Preference Elicitation Methods in the Broader Context of Random Utility Theory Preference Elicitation Methods. In: Gustafsson A, Herrmann A, Huber F, editors. Conjoint measurement: methods and applications. Berlin: Springer; 2000. p. 167–98.

2. Ryan M, Scott DA, Reeves C, Bate A, Teijlingen ER van, Russell EM, et al. Eliciting public preferences for healthcare: a systematic review of techniques. Health technology assessment (Winchester, England). 2001;5(5):1–186.

3. Nord E. The Person-Trade-Off Approach to Valuing Health-Care Programs. Medical Decision Making. 1995 Sep;15(3):201–8.

4. Prades JLP. Is the person trade-off a valid method for allocating health care resources? Health economics. 1997 Feb;6(1):71–81.

5. Menzel P. How should what economists call “social values” be measured. The Journal of Ethics. 1999 Jul;3(3):249.

6. Green C. On the societal value of health care: what do we know about the person trade-off technique? Health economics. 2001 Apr;10(3):233–43.

7. Damschroder LJ, Roberts TR, Zikmund-Fisher BJ, Ubel PA. Why people refuse to make tradeoffs in person tradeoff elicitations: a matter of perspective? Medical decision making: an international journal of the Society for Medical Decision Making. 2007 Jun;27(3):266–80.

8. Mullen PM. Public involvement in health care priority setting: an overview of methods for eliciting values. Health Expectations. 1999;2(4):222–34.

9. Louviere JJ, Hensher DA, Swait JD. Stated choice methods: analysis and applications. Cambridge, U.K.: Cambridge University Press; 2000. 402 p.

10. Green C, Gerard K. Exploring the social value of health-care interventions: a stated preference discrete choice experiment. Health economics. 2009 Aug;18(8):951–76.

11. Koopmanschap MA, Stolk EA, Koolman X. Dear policy maker: Have you made up your mind? A discrete choice experiment among policy makers and other health professionals. International Journal of Technology Assessment in Health Care. 2010;26(02):198.

12. Lancsar E, Wildman J, Donaldson C, Ryan M, Baker R. Deriving distributional weights for QALYs through discrete choice experiments. Journal of Health Economics. 2011 Mar;30(2):466–78.

13. Schwappach DL, Strasmann TJ. “Quick and dirty numbers”? The reliability of a stated-preference technique for the measurement of preferences for resource allocation. Journal of health economics. 2006 May;25(3):432–48.

14. Ratcliffe J. Public preferences for the allocation of donor liver grafts for transplantation. Health economics. 2000 Mar;9(2):137–48.

15. Schwappach DL. Does it matter who you are or what you gain? An experimental study of preferences for resource allocation. Health economics. 2003 Apr;12(4):255–67.

16. Desser AS, Gyrd-Hansen D, Olsen JA, Grepperud S, Kristiansen IS. Societal views on orphan drugs: cross sectional survey of Norwegians aged 40 to 67. BMJ. 2010 Sep;341(sep22 3):c4715–c4715.

17. Chan HM, Cheung GMY, Yip AKW. Selection criteria for recipients of scarce donor livers: a public opinion survey in Hong Kong. Hong Kong Medical Journal. 2006 01;12.

18. Damschroder LJ, Baron J, Hershey JC, Asch DA, Jepson C, Ubel PA. The Validity of Person Tradeoff Measurements: Randomized Trial of Computer Elicitation Versus Face-to-Face Interview. Medical Decision Making. 2004 Mar 1;24(2):170–80.

19. Arrow K, Solow R, Portney PR, Leamer EE, Radner R, Schuman H. Report of the NOAA panel on contingent valuation. Federal register. 1993;58(10):4601–14.

20. Leggett CG, Kleckner NS, Boyle KJ, Dufield JW, Mitchell RC. Social Desirability Bias in Contingent Valuation Surveys Administered Through In-Person Interviews. Land Economics. 2003 Nov 1;79(4):561–75.

21. Kreuter F, Presser S, Tourangeau R. Social Desirability Bias in CATI, IVR, and Web Surveys. Public Opinion Quarterly. 2008 Dec 1;72(5):847–65.

22. Skedgel C. Societal Preferences in the Allocation of Healthcare Resources: An Empirical Ethics Approach. Medical Decision Making. 2011;31(1):E71.

23. Schwappach DL. Resource allocation, social values and the QALY: a review of the debate and empirical evidence. Health expectations: an international journal of public participation in health care and health policy. 2002 Sep;5(3):210–22.

24. Dolan P, Shaw R, Tsuchiya A, Williams A. QALY maximisation and people’s preferences: a methodological review of the literature. Health economics. 2005 Feb;14(2):197–208.

25. Mathews KE, Freeman ML, Desvousges WH. How and How Much? [Internet]. In: Valuing Environmental Amenities Using Stated Choice Studies. Springer Netherlands; 2007. p. 111–33. Available from: http://dx.doi.org/10.1007/1-4020-5313-4\_5

26. Kuhfeld WF. Marketing Research Methods in SAS [Internet]. Marketing Research Methods in SAS. 2010 Oct [cited 2010 Dec 1];Available from: http://support.sas.com/resources/papers/tnote/tnote\_marketresearch.html

27. Carlsson F, Martinsson P. Design techniques for stated preference methods in health economics. Health Econ. 2003 Apr;12(4):281–94.

28. Hensher DA, Greene WH, Rose JM. Applied choice analysis : a primer. Cambridge: Cambridge University Press; 2005. 717 p.

29. Fleiss JL, Levin BA, Paik MC. Statistical methods for rates and proportions. 3rd ed. Hoboken N.J.: J. Wiley; 2003. 798 p.

30. Wright SP. Adjusted P-Values for Simultaneous Inference. Biometrics. 1992 Dec 1;48(4):1005–13.

31. Hosmer DW, Lemeshow S. Applied logistic regression. New York: Wiley; 2000. 375 p.

32. Scott A. Identifying and analysing dominant preferences in discrete choice experiments: An application in health care. Journal of Economic Psychology. 2002;23(3):383–98.

33. Abdi, Herve. The Kendall Rank Correlation Coefficient. In: Salkind, Neil J., Rasmussen, Kristin, editors. The Encyclopedia of Measurement and Statistics. Thousand Oaks, Calif.: SAGE Publications; 2007. p. 1416.

34. Baltagi BH. Econometric Analysis of Panel Data. 4th ed. Chichester, UK: Wiley; 2008. 366 p.

35. Long J. Regression models for categorical and limited dependent variables. Thousand Oaks: Sage Publications; 1997.

36. Orme B. Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research. Madison: Research Publishers LLC; 2006.

37. Keeney RL. Common Mistakes in Making Value Trade-Offs. Operations Research. 2002 Nov 1;50(6):935–45.

**Table 1: Attributes and levels**

|  |  |
| --- | --- |
| Attributes | Levels |
| Age | 10, 40, 70 |
| Initial utility | 0.1, 0.5, 0.9 |
| Initial life expectancy | 1 month, 5 years, 10 years |
| Final utility | 0.1, 0.5, 0.9 |
| Life years gained | 1 year, 5 years, 10 years |
| Number of patients | 500, 2500, 5000 |

**Table 2: Respondent characteristics by questionnaire**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | DCE (%) | CSPC (%) | p-value | Adjusted *p* |
| Completion | 154/256 (60%) | 150/348 (43%) | <0.001 | <0.001 |
| ‘Decision maker’ | 33 (21%) | 18 (12%) | 0.04 | 1.00 |
| ‘Doctor’ | 35 (23%) | 35 (23%) | 1.00 | 1.00 |
| ‘Frequent user’ | 14 (9%) | 18 (12%) | 0.52 | 1.00 |
| Female | 113 (74%) | 107 (71%) | 0.77 | 1.00 |
| Mean age\* | 31.5 | 33.2 | 0.65 | 1.00 |

*\* Mean age calculated using mid-point of age groups; p-value for test of independence*

**Table 3: Respondents rating the questionnaires “somewhat difficult” or “extremely difficult” to understand**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | DCE (%) | CSPC (%) | p-value | Adjusted *p* |
| All respondents | 19/154 (12%) | 19/150 (13%) | 1.00 | 1.00 |
| ‘Decision maker’ | 5/33 (15%) | 5/18 (28%) | 0.47 | 1.00 |
| ‘Doctor’ | 6/35 (17%) | 5/35 (17%) | 1.00 | 1.00 |
| ‘Frequent user’ | 1/14 (7%) | 2/18 (11%) | 1.00 | 1.00 |

**Table 4: Respondents rating the questionnaires “somewhat difficult” or “extremely difficult” to answer**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | DCE (%) | CSPC (%) | p-value | Adjusted *p* |
| All respondents | 100/154 (65%) | 99/150 (66%) | 0.94 | 1.00 |
| ‘Decision maker’ | 25/33 (76%) | 14/18 (78%) | 1.00 | 1.00 |
| ‘Doctor’ | 20/35 (57%) | 21/35 (60%) | 1.00 | 1.00 |
| ‘Frequent user’ | 7/14 (50%) | 11/18 (61%) | 0.79 | 1.00 |

**Table 5: Incidence of dominant preferences and non-traders**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | DCE (%) | CSPC (%) | p-value | Adjusted *p* |
| Dominant preferences | 20/151 (13%) | 27/150 (18%) | 0.328 | 1.00 |
| Confirmed non-traders | 10/151 (7%) | 13/150 (9%) | 0.652 | 1.00 |

**Table 6: Model coefficients, coefficients of variation and marginal rates of substitution**



**Figure 1: Choice-attribute correlation by survey**



*Mean correlation between the dominant level of an attribute in an alternative and the choice of that alternative*

**Figure 2: Relative attribute importance by survey**



*Relative importance weights showing each attribute’s standardised contribution to overall utility (DCE) or budget difference (CSPC)*

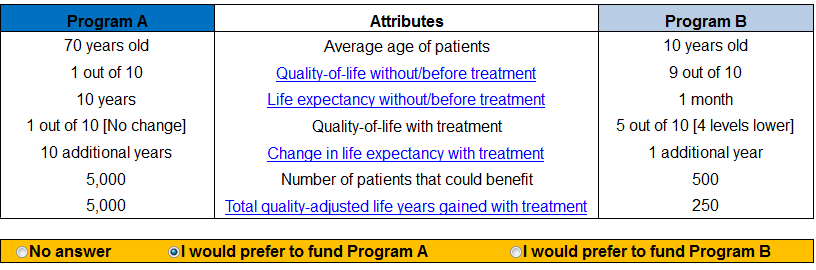
**Figure 2: Histogram of budget allocations**



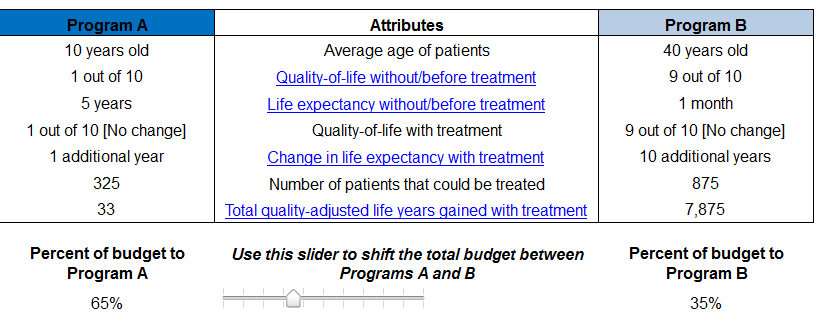
*Histogram showing the distribution of budget allocations to Program B across all CSPC tasks. 100 represents a 100% allocation to Program B; 0 represents a 100% allocation to Program A; 50 represents an equal 50%-50% allocation to both programs.*

**Appendix**

**Sample discrete choice experiment**

****

**Sample CSPC task**

****

1. CSPC is similar to constant-sum scaling (CSS), which presents a single alternative and asks respondents to allocate a fixed budget or points between attributes.(8) Both CSPC and CSS are sometimes referred to as ‘budget pie’ tasks. [↑](#footnote-ref-1)
2. Total QALYs gained = (Ufinal-Uinitial) x LYs gained x Patients [↑](#footnote-ref-2)