doi: 10.1111/hex.12098

ORIGINAL RESEARCH PAPER

# Choosing vs. allocating: discrete choice experiments and constant-sum paired comparisons for the elicitation of societal preferences

Chris D. Skedgel MDE,*† Allan J. Wailoo PhD‡ and Ron L. Akehurst BSc, Hon MFPHM‡

*PhD Student, ‡Professor, School of Health & Related Research, The University of Sheffield, Sheffield, UK, †Research Health Economist, Atlantic Clinical Cancer Research Unit, Capital Health, Halifax, NS, Canada

**Correspondence**
Chris Skedgel, MDE
PhD Student
Centre for Clinical Research
Room 242
5790 University Avenue
Halifax
NS B3H 1V7
Canada
E-mail: c.skedgel@sheffield.ac.uk

**Accepted for publication**
23 May 2013

**Keywords:** constant-sum paired comparisons, discrete choice experiments, societal preferences, stated preference methods

## Abstract

**Background** There is growing evidence of a reluctance to allocate health care solely on the basis of maximizing quality-adjusted life years (QALYs). Stated preference methods can be used to elicit preferences for efficiency vs. equity in the allocation of health-care resources.

**Objective** To compare discrete choice experiment (DCE) and constant-sum paired comparison (CSPC) methods for eliciting societal preferences.

**Methods** Over a series of choice pairs, DCE respondents allocated a fixed budget to one preferred group and CSPC respondents allocated budget percentages between the groups. Questionnaires were compared in terms of completion rates, preference consistency, dominant preferences and derived attribute importance.

**Results** There was no significant difference in the proportions that rated the questionnaires somewhat or extremely difficult, but a significantly greater proportion completed the DCE compared to the CSPC. Preference consistency was also higher in the DCE. The incidence of dominant preferences, including for aggregate QALYs, was low and not significantly different between questionnaires. Similarly, no CSCP respondents equalized budgets or outcomes in every task. Final health state was the most important attribute in both questionnaires, but the rankings diverged for the other attributes. Notably, the total patients' treated attribute was important in the CSPC but insignificant in the DCE, perhaps reflecting a 'prominence effect'.

**Conclusions** Despite lower completion rates and preference consistency, CSPC may offer advantages over DCE in eliciting preferences over the distribution of resources and/or outcomes as well as attribute levels, avoiding extreme 'all-or-nothing' distributions and possibly aligning respondent attention more closely with a societal perspective.

 

## Introduction

The conventional quality-adjusted life year (QALY) maximization approach to health-care priority setting explicitly assumes that within a fixed budget, the only factors that are relevant to the societal value of health care are the absolute gain in health-related quality of life, the duration of the benefit and the number of patients receiving treatment. An increase in any of these factors is associated with a proportional increase in value.[1] There is growing evidence, however, of a reluctance to allocate health care solely on the basis of maximizing QALYs and a willingness to sacrifice efficiency in exchange for equity or 'distributive justice' goals.[1–3]

Choice-based stated preference (SP) methods can be used to measure the strength and direction of preferences for efficiency and equity in health care on a cardinal scale by asking respondents to trade-off between alternatives on the basis of different attributes and levels.[4,5] But whereas conventional elicitations of welfare ask individuals 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* being in a certain condition.[6]

Reviews by Ryan *et al.*[5] and Mullen[7] identify a number of methods for eliciting societal preferences, but two specific choice-based SP methods are compared here. First, discrete choice experiments (DCEs) ask respondents to choose a single preferred option from two or more alternatives, and over a series of choice tasks can elicit preferences for the different attribute levels that define each alternative. DCEs are reasonably straightforward tasks, conceptually similar to many decisions respondents make on a daily basis and are increasingly common in health economic applications.[5,8,9] They have been used in a number of recent elicitations of societal values for priority setting in health care.[10–15]

Second, constant-sum paired comparisons (CSPCs) ask respondents to allocate a fixed budget between two alternatives.* In contrast to the 'pick one' nature of a DCE, respondents can choose to allocate the entire budget to alternative A, to alternative B, or to some combination of the two, including an equal split. Schwappach[16] argues that the approach is unique in connecting budget constraints, health outcomes and patient characteristics, and Ryan[5] suggests that the approach has an apparent simplicity and intuitive appeal. CSPC has the additional advantage of being able to identify preferences for specific resource allocations. For example, respondents may express a preference for equality in the allocation of resources or, depending on the attributes included in the task, in outcomes. CSPC has had more limited use than DCE in health economic applications, but has been successfully used to elicit societal preferences for efficiency and equity in health care.[16–21] Sample DCE and CSPC tasks are shown in Appendix S1.

The response characteristics of CSPC have not, to our knowledge, been directly compared to DCE. To this end, the two methods were compared as part of a pilot study intended to identify the preferred technique for a larger elicitation of preferences over the allocation of societal health-care resources. The following section outlines the experimental design and data collection methods used in the elicitation, as well as the criteria by which the two techniques were compared. These criteria included completion rates, respondent-rated difficulty, preference consistency and the incidence of a dominant preference for a specific choice attribute. The estimation of preferences was not a primary objective of this comparison, but a simple analysis of the choice responses was conducted confirm the general feasibility of the

*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.[7] Both CSPC and CSS are sometimes referred to as 'budget pie' tasks.

 

methods and to identify potential response biases in the two formats.

## Methods

### Experimental design

An empirical ethics review of factors relevant to the allocation of health-care resources between patient groups[22] identified four attributes that had evidence of public support as well as defensible ethical justification: patient age, health-related quality of life before and after treatment, and the number of patients in each group. Life expectancy without/before treatment and life years gained with/after treatment were also included to facilitate the calculation of quality-adjusted life years (QALYs) as a summary measure of health gain, despite ambiguous evidence around societal preferences for duration of health benefit and the QALY itself.[1,2,22]

A D-efficient optimal fractional factorial experimental design with 18 choice sets of two alternatives each was developed using SAS macros.[23] The design started with a full factorial candidate design of 6 attributes with three levels each (see Table 1) and excluded combinations where the net QALY gain with treatment was negative, or health state and life expectancy were unchanged before and after treatment. The final 18 choice sets were divided into two blocks of 9 choice sets each, with no duplication between the two blocks. Choice set 3 from each design block was repeated to test preference consistency, resulting in a total of 10 choice tasks presented–always in the same order–to each respondent. The survey was simplified by using block 1 for the DCE questionnaire and block 2 for the CSPC questionnaire, rather than randomly assigning respondents to different blocks within each questionnaire. Although this simplified questionnaire administration, it violates the principles of optimal experimental design and limits its overall efficiency.[24]

**Table 1** Attributes and levels

| Attributes | Levels |
|---|---|
| Average age of patients (Age) | 10, 40, 70 |
| Quality of life without/before treatment (Initial utility) | 0.1, 0.5, 0.9 |
| Life expectancy without/before treatment (Initial life expectancy) | 1 month, 5 years, 10 years |
| Quality of life with treatment (Final utility) | 0.1, 0.5, 0.9 |
| Change in life expectancy with treatment (Life years gained) | 1 year, 5 years, 10 years |
| Number of patients that could be treated (Patients treated) | 500, 2500, 5000 |

Aggregate QALYs gained was calculated for each choice scenario as [final utility 9 (initial life expectancy + life years gained) − initial utility 9 initial life expectancy] 9 total patients treated. The value of aggregate QALYs gained in the experimental design ranged from 54 to 45 373, with a mean of 10 591.

### Data collection

Individuals were invited complete an online questionnaire via a mass email to students at the University of Sheffield, Sheffield, UK, and posters and electronic announcements to students, staff and faculty at Dalhousie University, Halifax, Canada, and physicians and staff at Capital District Health Authority, Halifax, Canada. The invitations, questionnaires and analysis methods were approved by The University of Sheffield, School of Health and Related Research Ethics Committee and the Capital District Health Authority Research Ethics Board.

Potential respondents were randomized to the DCE or CSPC questionnaire using a random number algorithm. Sixty percent of potential respondents were assigned the CSPC questionnaire to compensate for a lower expected completion rate due to its greater cognitive demand, with the remaining 40% assigned the DCE. As each potential respondent followed a link to participate, the assigned questionnaire was recorded and used as the denominator in calculating completion rates.

Respondents were asked to imagine themselves as a societal decision maker responsible for allocating a fixed budget between two health-care programs. They were told that both programs had the same 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. Aggregate QALYs gained† was included as part of each alternative to test respondent consistency with the principles of QALY maximization, and a brief description of each attribute was provided (see Appendix S1). The concept of cost-effectiveness was not introduced, but given that both programs had the same cost, some respondents may have inferred that the program associated with the greatest aggregate QALYs gained was also the more cost-effective option.

The DCE tasks asked respondents to allocate the entire budget to a single alternative and did not include an option to indicate indifference between the alternatives. The CSPC tasks asked respondents to allocate budget percentages between the two groups by moving a slider. The number of patients treated and aggregate QALYs gained in the CSPC alternatives changed in proportion with the budget as the respondent moved the slider (e.g. a 25% budget allocation meant 25% of the potential patients could be treated). Unlike the DCE, respondents could express indifference between the alternatives with a 50–50% budget allocation. Respondents could dropout at any stage, and data were only collected from those respondents who completed the entire questionnaire. See the Appendix S1 for sample DCE and CSPC tasks.

Following the choice tasks, respondents were asked to rate the importance of each attribute, and the importance of an equal distribution of resources, to their choices on a 0–10 scale (see Appendix S1) and to separately 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 health-care user (≥12 contacts in the past 12 months). These categories were not mutually exclusive, and each respondent may have been in more than one category or none at all.

Preliminary interviews and informal focus groups were used to refine the wording and presentation of the DCE and CSPC tasks as well as the difficulty and importance ratings.

## Survey comparisons

### Completion rates

Differences in completion rates and stakeholder and gender proportions were tested using a two-sample $Z$-test of proportions. Age group proportions were tested using a $\chi^2$ test of independence. On the assumption that the randomization 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 dropout rate among these groups. Throughout the comparisons, multiple $P$-values were adjusted for simultaneous comparisons using Hommel's method,[25,26] and the analyses were conducted with R, version 2.15.2 (R Foundation for Statistical Computing, Vienna, Austria).[27]

### Respondent-rated difficulty

The proportions of DCE and CSPC respondents who indicated that they found the questionnaire 'somewhat difficult' or 'extremely difficult' to understand or to answer were compared using a two-sample $Z$-test of proportions.

### Preference consistency

Preference consistency was measured by including a repeated task in each questionnaire–the two alternatives presented as task 3 of each block were reversed and represented as task 8. Consistency required that respondents should prefer the same alternative in the repeated task as in the original. To compare the two methods on a common basis, CSPC responses were

†Aggregate QALYs gained = [final utility × (initial life expectancy + life years gained) − initial utility × initial life expectancy] × total patients treated.

transformed to discrete choices based on which alternative was allocated the majority of the budget. Equal budget allocations were allowed, but an equal allocation had to be made in both tasks to be classified as consistent. The proportion of consistent respondents in the two questionnaires was compared using a two-sample $Z$-test of proportions.

### Dominant preferences

A respondent was considered to have a dominant preference for a particular attribute if they always choose the alternative with the higher (or lower) level of that attribute, regardless of the levels of the other attributes.[28] Although such preferences are not inconsistent with the axioms of choice theory,[29,30] they do violate the assumptions that underlie choice-based stated preference methods: compensatory decision making, which implies that a deterioration in one attribute can be compensated for by an improvement in another attribute, and an additive utility function. Such preferences cannot be represented by an indifference curve, and as no trading takes place, marginal rates of substitution have no meaning.[28,31]

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 'best' or more preferred level of each attribute, based on expectations from the empirical ethics review.[22] For example, based on public support and an ethical justification for prioritizing more severely ill patients, if one alternative presented patients in a more severe initial health state, that alternative was flagged as 'best' (from the perspective of the respondent) in the initial utility attribute; the attribute flag for the paired alternative was set to zero. A respondent who invariably chose the alternative with more severely ill patients would have a correlation coefficient of 1.0 between the choice flag and the initial utility flag (perfect choice–attribute correlation). Which end of the attribute scale each respondent considered 'best' was not critical as this only affected the sign on the correlation coefficient, although this only holds where preferences are monotonically increasing or decreasing over the attribute, as was assumed here.

Constant-sum paired comparison responses were transformed to discrete choices on the basis of which program was allocated the majority of the budget, and the attribute flags were set based on the potential number of patients treated and QALYs gained if 100% of the budget were allocated to that alternative. CSPC alternatives that received a 50% budget allocation were flagged as 'not chosen' (i.e. both alternatives were assigned a choice flag of zero) as neither alternative was considered to be prioritized, but the impact of counting such allocations as prioritizing the preferred attribute level–under the rationale that respondents gave at least equal priority to the preferred attribute level–was also tested. As the outcome of interest was the correlation between pairs of nominal variables (i.e. choice and attribute flag), Kendall's tau[32] was used as the measure of correlation and was estimated using the ltm package.[33]

As respondents saw only a subset of possible scenarios, it is not possible to say that an observed dominant preference, defined as perfect correlation between a respondent's choices and the level of a particular attribute, would necessarily hold across all possible scenarios.[28] To support the identification of dominant preferences, therefore, each respondent's self-rated attribute importance scores were converted to rankings, and individuals with a perfect choice–attribute correlation who also rated that attribute as most important were considered to have a dominant preference for that attribute. The proportion of dominant preferences was compared across the two questionnaires using a two-sample $Z$-test.

Given that program cost was the same in both alternatives, a dominant preference for greater aggregate QALYs was, in effect, a preference for the more cost-effective alternative, consistent with the principles of QALY maximization.[1] However, in holding costs constant, it is not possible to distinguish between a

QALY-maximizing preference for the more cost-effective alternative and a dominant preference for aggregate QALYs regardless of cost. Therefore, a dominant preference for aggregate QALYs is necessary but not sufficient to confirm support for QALY maximization. The mean number of QALY-maximizing choices made by respondents to the two questionnaires was compared using a two-sample *t*-test.

### CSPC budget allocation preferences

As the ability to elicit preferences for specific resource allocations is a key feature of the CSPC, the distribution of allocation choices is also presented. Based on the attributes presented in each task, CSPC respondents had the opportunity to maximize the budget to one program or the other or to split the budget between the two programs in ways that could equalize resources, patients treated or aggregate QALYs gained.

In the case of an equal allocation of resources, it was not possible in any single CSPC task to distinguish between indifference due to equality in the latent utility of the alternatives and a preference for equality regardless of the attribute levels. As with the confirmation of dominant preferences discussed above, therefore, respondents were only categorized as having a dominant preference for equality if they chose an equal 50–50% allocation across all tasks and also ranked an equal distribution of resources as the most important factor in their choices.

### Econometric analysis

The simple choice models assumed monotonic preferences and linear main effects, and responses to the repeated task were excluded from the analyses to avoid double-counting. All respondents were included in the analysis, including those with confirmed dominant preferences.[30]

The response variable in the CSPC analysis was the budget allocated to alternative B less the budget allocated to alternative A. If 100% of the budget was allocated to alternative B, ΔBudget = +100; if 100% was allocated to Program A, ΔBudget = −100; and if the budget was allocated 50–50%, ΔBudget = 0. The DCE response variable was a 0, 1 flag indicating whether or not Program B was chosen. Responses were modelled as a function of the differences in the raw (i.e. continuous) attribute levels, calculated as above. Aggregate QALYs gained were excluded to avoid collinearity with the other attributes. CSPC responses were modelled using a double-bounded random effects Tobit to account for the censored dependent variable and the panel structure of the data,[34] while DCE responses were modelled using a binary random effects probit to be as comparable as possible with the CSPC model. To allow for the broadest possible inclusion of explanatory parameters in the models, a significance threshold of 0.10 was adopted and *P*-values were not adjusted for multiple comparisons.[35] The analyses were performed using the censReg and pglm packages.[36,37]

Coefficients from the DCE and CSPC models represent the change in latent utility, or the change in the *difference* in latent utility, respectively, associated with a 1-unit change in each attribute. Due to this difference in interpretation, as well as differences in the variance scale between the two models,[38] these coefficients are not directly comparable. Instead, they were compared on the basis of the relative contribution of each attribute to overall utility,[39] calculated based on the most preferred and least preferred level of each attribute, $x$:

$$\Delta vx = (\beta x)^{\max} - (\beta x)^{\min}$$

where $(\beta x)^{\max}$ is the utility associated with the *most* preferred level of attribute $x$, $(\beta x)^{\min}$ is the utility associated with the *least* preferred level of attribute $x$ and $\Delta v(x)$ is the net difference. This attribute-specific contribution to utility was divided by the difference in overall utility 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:

$$\text{Relative Importance of } x = \frac{\Delta v(x)}{v^{\max} - v^{\min}}$$

Marginal rates of substitution (MRS) were also calculated as the ratio of the coefficient of each attribute to the coefficient on life years gained and represented the average willingness to sacrifice individual life year gains for a 1-unit change in the level of a particular attribute.[40] A negative sign indicates that respondents preferred a lower level of an attribute, while a positive sign indicates a preference for a higher level.

## Results

### Completion rates

A total of 604 individuals chose to participate: 348 (58%) were randomized to the CSPC, and 256 (42%) were randomized to the DCE. Completion rates and respondent characteristics are shown in Table 2. A significantly greater proportion of individuals randomized to the DCE completed a questionnaire compared with individuals randomized to the CSPC ($P < 0.001$). 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 health-care users. A substantially lower proportion of respondents identified themselves as decision makers among completed CSPC questionnaires, although this difference was not significant after adjusting for multiple comparisons.

### Respondent-rated difficulty

As shown in Table 3, there was no significant difference between the two questionnaires 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 difficulty of answering the DCE and CSPC.

### Consistency in the repeated task

Ninety-six percent of respondents to the DCE (148/154) preferred the same program (including three consistent 'no answers') in the original and the repeated task, while 79% (119/150) of CSPC respondents allocated the majority of the budget to the same program or preferred an equal allocation of resources in both tasks. The difference in the proportion of consistent respondents between the two questionnaires was statistically significant (difference = 17%; $P < 0.001$).

### Dominant preferences

Excluding three individuals who always chose 'no answer' in the DCE, the proportion of

**Table 2** Completion rates and respondent characteristics by questionnaire

| | DCE (%) | CSPC (%) | *P*-value | Adjusted *P* |
|---|---|---|---|---|
| Overall completion rate | 154/256 (60%) | 150/348 (43%) | <0.001 | <0.001 |
| Self-identified stakeholders, *N* (%) | | | | |
| 'Decision maker' | 33 (21%) | 18 (12%) | 0.04 | 0.20 |
| 'Doctors' | 35 (23%) | 35 (23%) | 1.00 | 1.00 |
| 'Frequent user' | 14 (9%) | 18 (12%) | 0.52 | 1.00 |
| Demographics, *N* (%) | | | | |
| Female | 113 (74%) | 107 (71%) | 0.77 | 1.00 |
| Mean age* | 31.5 | 33.2 | 0.65 | 1.00 |

*Mean age was calculated using the mid-point of age groups; *P*-value for $\chi^2$ 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 |

respondents who always chose the alternative with the preferred level of a particular attribute was 9% in the DCE (14/151) and 5% in the CSPC (7/150); the difference was not significant ($P = 0.27$, adjusted- $P = 0.34$). The most frequent perfectly correlated attribute in the DCE and CSPC was final health state (9/14) and life years gained (6/7), respectively. Among respondents with at least one perfectly correlated attribute, seven DCE (5%) and three CSPC (2%) respondents also ranked that attribute as the most important factor in their choices, taken as confirmation of a dominant preference. Again, this difference was not statistically significant ($P = 0.34$, adjusted-$P = 0.34$). Counting equal budget allocations in the CSPC as prioritizing the dominant attribute increased the proportion of respondents with perfect choice–attribute correlations and dominant preferences to 11 and 5%, respectively, but the differences between the CSPC and DCE were still not significant.

Three additional DCE respondents had a perfect correlation between choice and total patients treated, but due to an error in the database attribute importance, ratings were not recorded for the 'number of patients treated' attribute. These respondents were not counted among the individuals with confirmed dominant preferences, but it is possible that up to 10 DCE respondents (7%) may have had a dominant preference, but this would not change the insignificant difference between the DCE and CSPC.

With specific reference to aggregate QALY gains, only one respondent, from the DCE questionnaire, chose the QALY-maximizing alternative in every task. This individual also ranked QALYs as the most important attribute, confirming a dominant preference for aggregate QALYs. On average, DCE respondents chose the QALY-maximizing alternative in 6.3 of 10 tasks, compared to 5.4 tasks of 10 among CSPC respondents ($P < 0.001$); both were slightly but significantly greater than the five choices of 10 that would be expected by chance alone, given that one alternative in each choice pair maximized QALYs gained (adjusted-$P < 0.001$ in both comparisons).

## CSPC distributional preferences

Figure 1 shows that the modal CSPC allocation (18% of all responses) maximized the budget to one program or the other, while the next most common allocation (7%) equalized the budget between the two programs. Two percent of CSPC respondents (3/150) maximized the budget in every task, and 11% (16/150) maximized the budget in five or more of their 10 choices. No respondents equalized budgets, patients or QALYs in more than five of their choices.

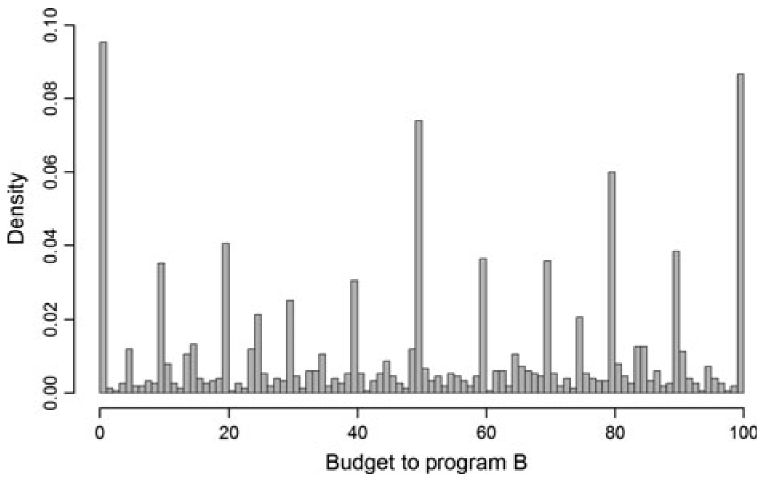


**Figure 1** Histogram of budget allocations.

**Table 5** Model coefficients, marginal rates of substitution and importance ranks

| Attribute | Estimate | Std. Error | Coeff of Var | Unadjusted *P*-value | MRS(LYg) | Rel. Importance | Rank |
|---|---|---|---|---|---|---|---|
| DCE | | | | | | | |
| Constant | 0.158 | 0.048 | 0.303 | <0.001*** | | | |
| Life years gained | 0.086 | 0.011 | 0.133 | <0.001*** | 1.00 | 0.321 | 2 |
| Age | −0.008 | 0.001 | 0.147 | <0.001*** | −0.09 | 0.202 | 3 |
| Initial utility | −0.058 | 0.009 | 0.154 | <0.001*** | −0.67 | 0.019 | 4 |
| Initial life expectancy | | | | | | | |
| Final utility | 0.138 | 0.012 | 0.086 | <0.001*** | 1.61 | 0.458 | 1 |
| Patients treated/1000 | | | | | | | |
| CSPC | | | | | | | |
| Constant | −32.255 | 3.891 | 0.121 | <0.001*** | | | |
| Δ Life years gained | 7.298 | 0.420 | 0.057 | <0.001*** | 1.00 | 0.101 | 4 |
| Δ Age | −0.137 | 0.053 | 0.390 | 0.010* | −0.02 | 0.013 | 6 |
| Δ Initial utility | −18.480 | 3.750 | 0.203 | <0.001*** | −2.53 | 0.228 | 2 |
| Δ Initial life expectancy | 2.995 | 0.411 | 0.137 | <0.001*** | 0.41 | 0.046 | 5 |
| Δ Final utility | 40.101 | 3.583 | 0.089 | <0.001*** | 5.49 | 0.496 | 1 |
| Δ Patients treated / 1000 | 16.705 | 1.544 | 0.092 | <0.001*** | 2.29 | 0.116 | 3 |

MRS(LYg), marginal rate of substitution using life years gained as the numeraire.
Significance codes: *** <0.001, ** <0.01, * <0.05, + <0.10.

## Econometric analysis

Table 5 presents the model coefficients and significance, marginal rates of substitution (MRS) in terms of life years gained, and relative importance weights and rankings for the DCE and CSPC models. Although respondents to the DCE and CSPC did not see the same choice sets, there were only weak correlations between the block indicator and the specific attribute levels, ranging from −0.14 to 0.16, suggesting no systematic bias in the attribute levels presented in the two questionnaires.

Initial life expectancy and the number of patients treated were not significant at a 0.10 threshold in the initial DCE probit model, and it was re-estimated excluding these attributes. All attributes were significant in the CSPC model. The direction of preferences was consistent between the two models: negative coefficients on age and initial utility suggest that younger and more severe patients were preferred, while positive coefficients on individual life years gained and final utility suggest respondents preferred greater life year gains and better final health states. CSPC respondents also preferred larger patients groups and, unexpectedly, patients with greater initial life expectancy, while these attributes were not significant in the DCE.

Marginal rates of substitution and relative importance rankings, illustrated in Fig. 2, suggest that final health state was the single most important attribute in both models, although the rankings diverged for the other attributes. After final utility, the DCE emphasized life years gained and patient age, while the CSPC emphasized initial utility and then patients treated and life years gained.

## Discussion

The respondent-rated difficulty of the two surveys was strikingly similar, with a small minority rating the tasks as difficult to understand and a strong majority rating the tasks as difficult to answer. These ratings are reassuring as they suggest that respondents were able to understand the tasks and that the key challenge was substantive rather than cognitive. These ratings are comparable to difficulty ratings reported by Green and Gerard in a societal DCE,[10] but as ratings were only collected from respondents that completed the questionnaires, they are most likely biased downwards as individuals that found the surveys exceedingly difficult are likely to have dropped out before completion.

The completion rate in the CSPC was significantly lower than the DCE, suggesting that it was less acceptable in some respects, although the CSPC response rate was similar to that reported by Ratcliffe (38%) in her application of CSPC.[18] The 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 lower completion rates in the CSPC survey compared to other stakeholder subgroups. 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,[41] and such a phenomenon may be particularly strong among decision-maker respondents to the CSPC, which requires a more explicit consideration of trade-offs.

Respondents to the DCE questionnaire were more consistent in preferring the same alternative in the repeated task compared to respondents to the CSPC questionnaire. The choice set that was used as the repeated task in the DCE may have contributed to this consistency, as 95% of respondents preferred the same alternative in the original DCE task, compared to only 77% in the original CSPC task. An

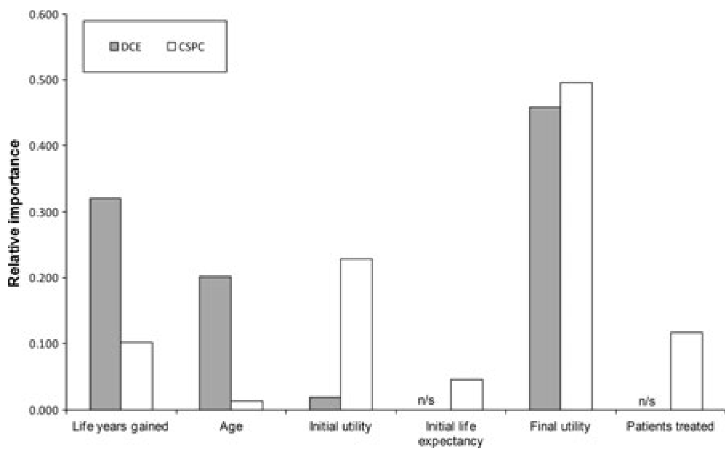


**Figure 2** Relative attribute importance by survey.

ideal test of consistency would have two alternatives with roughly equal choice probabilities, but it was not possible to predict these probabilities prior to the elicitations.

The proportion of dominant preferences in the DCE (5%) and the CSPC (2%), including preferences for aggregate QALYs gained and equality in CSPC budget allocations, were less than the 45% reported by Scott using a DCE,[28] and the 13% reported Schwappach[16] with a CSPC, but greater than the 2.3% reported by Ratcliffe,[18] also with a CSPC. Overall, respondents to the more cognitively demanding CSPC appeared no more likely to resort to this heuristic than respondents to the DCE. Although the difference between the DCE and CSPC was not statistically significant, the higher observed rate of dominant preferences observed in the DCE may support a notion that the more competitive nature of the DCE task tended to focus attention on a single attribute to a greater degree than the more reflective CSPC.[42]

The low incidence of dominant preferences for aggregate QALYs in both questionnaires offers little support for strict QALY maximization and suggests that respondents gave more weight to patient characteristics than aggregate outcomes. This is particularly noteworthy in the CSPC, where respondents may reasonably have been expected to take the fact that aggregate QALYs gained changed as they moved the slider as a cue to focus on this attribute. As above, this suggests that respondents to the more complex CSPC were no more likely than DCE respondents to resort to QALY maximization as a simplifying heuristic.

It is also worth noting that the predominant preference in Schwappach's CSPC elicitation[16] was for equality in resources–a preference not observed here. The unexpected willingness among CSPC respondents to maximize allocations to a single group challenges previous studies that found a general aversion to extreme distributions[16,18] and highlights the value of explicitly testing for such preferences. Although equal budgets were relatively frequent among the CSPC responses, the relative lack of strict equalizing behaviour observed here seems noteworthy, as in the absence of any obvious rationale for a particular budget allocation, respondents may have been expected to use an equalizing allocation (of resources, patients or outcomes) as a heuristic for a 'fair' allocation. Instead, consistent with the underlying theory, it appeared that respondents chose allocations based on the relative utility of the paired alternatives.

In the econometric analysis, respondents to both questionnaires had a strong preference for patients who would finish treatment in the best final health state, with relative importance weights close to 50%. But whereas the next most important attribute in the CSPC was initial health state, where respondents preferred patients in more severe initial health states, DCE respondents assigned greater weight to individual life year gains. Note that if all respondents were strict QALY maximizers, it would imply that attribute levels had no impact on choices, and the regression coefficients, attribute importance and rates of substitution would have no interpretation.[28]

Interestingly, CSPC respondents also had a preference for larger patient groups, while the number of patients was not statistically significant in the DCE. This result may be consistent with a 'prominence effect', which suggests that respondents become more sensitive to a quantity when it is harder for them to ignore.[43,44] As mentioned, the number of patients treated changed as respondents moved the budget slider, potentially highlighting this attribute. In the light of qualitative evidence that suggests SP respondents may reduce abstract, macrolevel allocation problems to more comprehensible two-person analogies,[45,46] such an effect may be an advantage in 'nudging' respondents back towards a macrolevel perspective.[44,47] These results must be interpreted cautiously, though, given the relatively small sample sizes and the less than optimal application of the experimental design. Although there appeared to be no systematic bias in the attribute levels of the two blocks of the experimental design, it is not possible to say conclusively that observed differences between the two questionnaires were not the result of respondents seeing different choice sets. In addition, there was little evidence to support a prominence effect with respect to aggregate QALYs in respondent choices, particularly as CSPC respondents were relatively less likely to choose QALY-maximizing alternatives.

Although the DCE had better completion rates and preference consistency, this 'respondent efficiency'[48] must be weighed against the richer preference data from the CSPC tasks. Despite the greater complexity of the task, the ability of CSPC to identify preferences for the distribution of resources and/or outcomes, to avoid uncomfortable extreme distributions and possibly to align respondent attention more closely with a societal perspective, appears to offer advantages over the more straightforward DCE. Arguably, the ability to avoid absolute discrimination against less-preferred groups makes CSPC particularly appealing in eliciting preferences in the context of health care, where respondents may be likely to have 'protected values'[49] or 'rights-based' preferences,[28] over which they may be reluctant to make or accept absolute trade-offs.[17]

## Funding

This work is a preliminary phase of a larger research program partly funded by the Canadian Centre for Applied Research in Cancer Control.

## Acknowledgement

The authors would like to acknowledge Dr. TallalYounis, Capital Health, Nova Scotia, for his assistance as site collaborator and the helpful comments of three anonymous reviewers.

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Appendix S1.** Sample DCE, CSPC and attribute rating tasks.

## References

1 Dolan P, Shaw R, Tsuchiya A, Williams A. QALY maximisation and people's preferences: a methodological review of the literature. *Health Economics*, 2005; **14**: 197–208.

2 Schwappach DL. Resource allocation, social values and the QALY: a review of the debate and empirical evidence. *Health Expectations*, 2002; **5**: 210–222.

3 Baker R, Bateman I, Donaldson C *et al*. Weighting and valuing quality-adjusted life-years using stated preference methods: preliminary results from the Social Value of a QALY Project. *Health Technology Assessment*, 2010; **14**: 1–162.

4 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 (eds). *Conjoint Measurement: Methods and Applications*. Berlin: Springer, 2000a. 167–198.

5 Ryan M, Scott DA, Reeves C *et al*. Eliciting public preferences for healthcare: a systematic review of techniques. *Health Technology Assessment (Winchester, England)*, 2001; **5**: 1–186.

6 Menzel P. How should what economists call "social values" be measured. *The Journal of Ethics*, 1999; **3**: 249.

7 Mullen PM. Public involvement in health care priority setting: an overview of methods for eliciting values. *Health Expectations*, 1999; **2**: 222–234.

8 Ali S, Ronaldson S. Ordinal preference elicitation methods in health economics and health services research: using discrete choice experiments and ranking methods. *British Medical Bulletin*, 2012; **103**: 21–44.

9 De Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. *Health Economics*, 2010; **21**: 145–172.

10 Green C, Gerard K. Exploring the social value of health-care interventions: a stated preference discrete choice experiment. *Health Economics*, 2009; **18**: 951–976.

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**: 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; **30**: 466–478.

13 Norman R, Hall J, Street D, Viney R. Efficiency and equity: a stated preference approach. *Health Economics*, 2013; **22**: 568–581.

14 Ratcliffe J, Bekker HL, Dolan P, Edlin R. Examining the attitudes and preferences of health care decision-makers in relation to access, equity and cost-effectiveness: a discrete choice experiment. *Health Policy*, 2009; **90**: 45–57.
15 Shah K, Tsuchiya A, Hole AR, Wailoo A. Valuing health at the end of life: a stated preference discrete choice experiment [Internet]. NICE Decision Support Unit; 2012 Dec p. 56. Available at: http://www.nicedsu.org.uk/DSU%20End%20of%20Life%20full%20report%20-%20version%203%20_Dec%202012_.pdf, accessed 5 June 2013.
16 Schwappach DL. Does it matter who you are or what you gain? An experimental study of preferences for resource allocation. *Health Economics*, 2003; **12**: 255–267.
17 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; **25**: 432–448.
18 Ratcliffe J. Public preferences for the allocation of donor liver grafts for transplantation. *Health Economics*, 2000; **9**: 137–148.
19 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; **22**: c4715–c4715.
20 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; **12**: 40–46.
21 Linley WG, Hughes DA. Societal views on nice, cancer drugs fund and value-based pricing criteria for prioritising medicines: a cross-sectional survey of 4118 adults in Great Britain. *Health Economics*, 2012; doi: 10.1002/hec.2872 [Epub ahead of print]
22 Skedgel C. Societal preferences in the allocation of healthcare resources: an empirical ethics approach. *Medical Decision Making*., 2011; **31**: E71.
23 Kuhfeld WF. Marketing research methods in SAS [Internet]. Marketing Research Methods in SAS. 2010 [cited 2010 Dec 1]. Available at: http://support.sas.com/resources/papers/tnote/tnote_marketresearch.html, accessed 5 June 2013.
24 Carlsson F, Martinsson P. Design techniques for stated preference methods in health economics. *Health Economics*, 2003; **12**: 281–294.
25 Wright SP. Adjusted P-values for simultaneous inference. *Biometrics*, 1992; **48**: 1005–1013.
26 Shaffer JP. Multiple hypothesis testing. *Annual Review of Psychology*, 1995; **46**: 561–584.
27 R Core Team. R: a language and environment for statistical computing [Internet]. Vienna, Austria; 2012. Available at: http://www.R-project.org/, accessed 5 June 2013.
28 Scott A. Identifying and analysing dominant preferences in discrete choice experiments: an application in health care. *Journal of Economic Psychology*, 2002; **23**: 383–398.
29 Mathews KE, Freeman ML, Desvousges WH. How and How Much? valuing environmental amenities using stated choice studies [Internet]. Springer Netherlands; 2007. p. 111–33. Available at: http://dx.doi.org/10.1007/1-4020-5313-4_5, accessed 5 June 2013.
30 Lancsar E, Louviere J. Deleting "irrational" responses from discrete choice experiments: a case of investigating or imposing preferences? *Health Economics*, 2006; **15**: 797–811.
31 Louviere JJ, Hensher DA, Swait JD. *Stated Choice Methods: Analysis and Applications*. Cambridge: Cambridge University Press, 2000b.
32 Herve A. The kendall rank correlation coefficient. In: Salkind NJ, Rasmussen K (eds) *The Encyclopedia of Measurement and Statistics*. Thousand Oaks, CA, SAGE Publications, 2007: 1416.
33 Rizopoulos D. ltm: an R package for latent variable modelling item response theory analyses. *Journal of Statistical Software*, 2006; **17**: 1–25.
34 Baltagi BH. *Econometric Analysis of Panel Data*, 4th edn. Chichester, UK: Wiley, 2008.
35 Hosmer DW, Lemeshow S. *Applied logistic regression*. New York: Wiley; 2000.
36 Henningsen A. censReg: censored regression (Tobit) models [Internet]. 2012. Available at: http://CRAN.R-project.org/package = censReg, accessed 5 June 2013.
37 Croissant Y. Pglm: panel generalized linear model [Internet]. 2010. Available at: http://CRAN.R-project.org/package = pglm, accessed 5 June 2013.
38 Swait J, Louviere J. The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of Marketing Research*, 1993; **30**: 305–314.
39 Orme B. *Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research*. Madison: Research Publishers LLC, 2006.
40 Lancsar E, Louviere J, Flynn T. Several methods to investigate relative attribute impact in stated preference experiments. *Social Science & Medicine (1982)*, 2007; **64**: 1738–1753.
41 Keeney RL. Common mistakes in making value trade-offs. *Operations Research*, 2002; **50**: 935–945.
42 Huber J. What we have learned from 20 years of conjoint research? [Internet]. Sawtooth Software; 2009. Available at: http://www.sawtoothsoftware.com/download/techpap/whatlrnd.pdf, accessed 5 June 2013.
43 Baron J, Greene J. Determinants of insensitivity to quantity in valuation of public goods: contribution,

warm glow, budget constraints, availability, and prominence. *Journal of Experimental Psychology*, 1996; **2**: 107–125.

44 Fischer GW, Carmon Z, Ariely D, Zauberman G. Goal-based construction of preferences: task goals and the prominence effect. *Management Science*, 1999; **45**: 1057–1075.

45 Giacomini M, Hurley J, DeJean D. Fair reckoning: a qualitative investigation of responses to an economic health resource allocation survey. Health Expectations [Internet]. 2012 Mar [cited 2012 Apr 30]; Available at: http://doi.wiley.com/10.1111/j.1369-7625.2011.00751.x, accessed 5 June 2013.

46 Ryan M. Rationalising the "irrational": a think aloud study of discrete choice experiment responses. *Health Economics*, 2009; **18**: 321.

47 McQuillin B, Sugden R. Reconciling normative and behavioural economics: the problems to be solved. *Social Choice and Welfare*, 2012; **38**: 553–567.

48 Severin V. Comparing statistical and respondent efficiency in choice experiments. [Unpublished PhD dissertation]. Sydney: University of Sydney, 2001.

49 Bartels DM, Medin DL. Are morally motivated decision makers insensitive to the consequences of their choices? *Psychological Science*, 2007; **18**: 24–28.