Several methods to investigate relative attribute impact in stated preference experiments

Abstract

There is growing use of discrete choice experiments (DCEs) to investigate preferences for products and programs and for the attributes that make up such products and programs. However, a fundamental issue overlooked in the interpretation of many choice experiments is that attribute parameters estimated from DCE response data are confounded with the underlying subjective scale of the utilities, and strictly speaking cannot be interpreted as the relative 'weight' or 'impact' of the attributes, as is frequently done in the health economics literature. As such, relative attribute impact cannot be compared using attribute parameter size and significance. Instead, to investigate the relative impact of each attribute requires commensurable measurement units; that is, a common, comparable, scale. We present and demonstrate empirically a menu of five methods that allow such comparisons: 1) partial log likelihood analysis; 2) the marginal rate of substitution for non-linear models; 3) Hicksian welfare measures; 4) probability analysis; and 5) best worst attribute scaling. We discuss the advantages and disadvantages of each method and suggest circumstances in which each is appropriate.

1. Introduction

A common objective of discrete choice experiments (DCEs) is to compare the relative impact of attributes of the product or program under investigation. For example, is test accuracy relatively more important to patients than time spent waiting for results when choosing diagnostic tests? Most studies compare relative impacts of attributes by comparing the size and significance of estimated parameters for attributes of interest. Unfortunately, these parameters are not directly comparable because the attribute parameter estimates in discrete choice models are confounded with the underlying subjective utility scale. That is, parameter estimates combine the relative impact or importance of an attribute *and* the utility scale values associated with its levels. Thus, utility estimates for attribute levels cannot be interpreted as indicating relative importance of an attribute.

In particular, the estimated utility of each attribute level is measured on an interval scale, but the origins and units of each attribute's utility scale differ. Apart from obvious differences in underlying physical attribute units like price in dollars, time in minutes/hours etc, qualitative attributes have no physical referents. For example, attribute levels for 'provider of care' might be nurse, doctor, etc. Thus, distances between the levels of different attributes need not have the same meaning. So, utility scale locations, or utility differences between levels of different attributes, generally do not have equal scale units. One can equate the origins of each scale, but not the scale units; hence, direct comparisons of ranges of utility estimates are meaningless without transforming them in a theoretically acceptable way, or modifying a choice experiment. Put simply, one cannot determine whether the magnitudes of the parameter estimates for an attribute's levels, and hence the resulting range of parameter estimates for these levels, is due to the 'impact' of that attribute or the position of each attribute level on the underlying utility scale. To assess relative attribute impacts one needs to measure each on a common, comparable, scale.

The purpose of this paper is to focus attention on the confound between attribute impact and attribute level scale utilities in DCEs, and to outline and discuss five ways to compare relative attribute impacts: 1) partial log likelihood analysis; 2) marginal rates of substitution (MRS); 3) Hicksian welfare measures; 4) probability analysis; and 5) best worst attribute scaling (BWAS). The first four methods deal with the issue of relative attribute impact within a traditional DCE. We demonstrate these in an empirical application, which to our knowledge is the first health-related DCE to include two-way attribute interactions in a non-linear indirect utility function (IUF). The BWAS method is a modified DCE.

The rest of the paper is organised as follows. The next section discusses the theoretical background for the confound between attribute impact and level scale. Section 3 outlines a menu of five methods to investigate the relative impact of attributes that are illustrated in two empirical applications in Section 4. Section 5 discusses advantages and disadvantages of each method and circumstances in which each may be appropriate. Section 6 concludes.

2. Confound between attribute impact and scale

Attribute parameters estimated in choice experiments combine the impact of an attribute *and* the underlying latent utility scale on which its levels are measured. This "confound" of impact and scale has long been recognised in utility theory and psychology (Anderson, 1970; Keeney & Raiffa, 1976; Louviere, 1988b; Lynch, 1985), but is less widely recognised by those who apply conjoint elicitation procedures (see McIntosh & Louviere (2002) for an exception). The following issues relate to the confound:

1. The importance or "impact" of an attribute on an individual's choice may a) be constant across the range of an attribute, implying it is independent of the levels, or b) vary systematically with the attribute levels. Anderson (1970, 1982) discusses the distinctions and a more general treatment of the concept of attribute weight is in Shanteau (1980). Notions of attribute weight are widespread, associated with many ad hoc schemes in everyday life where people use "weighting schemes" to compute overall indices like "attractiveness" or "utility" for sets of attributes, like restaurant quality ratings.
2. The attribute level scales discussed in this paper differ from the more commonly known concept of "scale factors" in the discrete choice literature. All utility scale parameter estimates in choice-based random utility models are confounded with a scale factor that is inversely related to the variance of the error term (Train, 2003), and can differ in each data source. To avoid confusion we refer to the latter as the "variance scale factor" and we term the scale under discussion in this paper as "level scale".
3. Attribute impacts are NOT the same as attribute level scales. A level scale value is the estimated position of an attribute level on an underlying latent dimension like "utility". Many psychologists and social scientists (eg, Fishbein & Ajzen (1975)) try to measure weights and scale level values independently, combining them via some integration rule or function, but such measures must satisfy mathematical operators to be valid.

Consider a doctors appointment described by three attributes (appointment length, choice of doctor, location). Suppose one uses a weighted averaging rule to combine the attributes into an overall utility index for appointments such that each attribute is assigned a weight that multiplies the scale values associated with each attribute level, with the resulting products summed into an index. For such a rule to be mathematically meaningful, each weight must be measured on a ratio scale. In the case of a weighted averaging model, one might have the following expression for an index that is a function of the three attributes:

$$U = w\_len * sc\_len + w\_dr * sc\_dr + (1 - w\_len - w\_dr) * sc\_loc \quad (1)$$

where w_len, w_dr and are weights associated with length of consultation and choice of doctor, with the weight for location implied by the restriction that the weights sum to 1; sc_len, sc_dr and sc_loc are scale values for each attribute defined as follows:

$$sc\_len = \alpha_{len} + \beta_{len} * len\_level$$

$$sc\_dr = \alpha_{dr} + \beta_{dr} * dr\_level$$

$$sc\_loc = \alpha_{loc} + \beta_{loc} * loc\_level$$

Allowing each attribute to have two levels, there are eight (2^3) possible appointments. If the scale values and weights have the values in Table 1, we can substitute them into equation (1) to obtain the (hypothetical) total appointment utilities in Table 2.

Table 1

Table 2

This appointment example involves known weights and level scales associated with each attribute. However, if we estimate these utilities from DCE data, we only recover a constant term (equal to 0.68, representing the utility of an appointment defined by the lowest level of each attribute) and a utility difference in the two levels of each attribute: 0.14 for length (calculated as 0.2-0.06); 0.03 for choice of doctor; and 0.15 for location. Such analysis does not separate the weights and level scale values, which is true in general for all conjoint elicitation procedures.

An important consequence of the weight-scale confound is that the "effect" (or lack thereof) of an attribute across its levels can be due to a large (small) weight relative to other attributes, or due to large (small) differences in scale values associated with the levels, or some combination of both. In fact, this applied to our artificial appointments example. The 'large' utility difference between the two levels for appointment length and location arise for different reasons. The first is due to a relatively large scale difference whilst the second is due to a relatively large weight. One cannot determine which case applies in a DCE without additional information. Because attribute impacts and level scales are confounded, inter-dimensional utility comparisons combine these two effects. To investigate the relative impact of an attribute in a traditional DCE, further analysis must be undertaken to place attributes on a common, comparable scale; alternatively, a modified choice experiment can be implemented. Both strategies are explored in Section 3.

Further, weight/scale confounds pose issues for generalising choice models because results may be "level dependent". That is, effects estimated from a DCE depend on levels varied, and different sets/ranges of levels may yield different results (e.g. see Louviere & Islam (2004); Ohler, Le, Louviere, & Swait (2000); Ryan & Wordsworth, (2000)). So, strictly speaking, conclusions about attribute effects should be qualified to be "relative", not absolute, with stronger conclusions reserved for results that generalise across different levels and subsets of attributes.

3. Methods to investigate relative impact of attributes

We outline five methods that place attributes on common and commensurable scales.

Partial log likelihood

One way to compare the relative 'impacts' of product/program attributes is to investigate the explanatory power of each attribute (or attribute level) by calculating how much each attribute contributes to the overall log likelihood of a choice model (Crouch & Louviere, 2004). This involves systematically re-estimating a choice model, omitting each attribute one at a time and recording the associated log likelihood. The contribution of each attribute is the difference between the full and reduced model log likelihoods. That is, the difference in model log likelihoods for an attribute (with all its levels) in and out of a model. Thus, attributes that are more 'important' in explaining choices will contribute more to the total log likelihood, as indicated by their partial log likelihoods. This approach is analogous to calculating partial r-squares for each attribute in traditional rating and ranking tasks in conjoint analysis (Louviere, 1988a). It also is related to statistical tests for the additional explanatory power of variables included/excluded from choice models (a model selection problem).

This method requires that the data for the analysis be orthogonal because if the data are multi-colinear, the impact of a removed attribute can be associated with another attribute, which would understate the importance of the removed attribute. Orthogonal experimental designs are necessary but not sufficient to ensure an orthogonal data set; if a design is blocked into versions, the versions must have equal sample sizes. If version sample sizes are unequal, one must re-weight the versions to ensure orthogonality.

Marginal rates of substitution

Often, marginal rates of substitution (MRS) are used to measure the rate at which individuals trade off one attribute for another (Gyrd-Hansen & Søgaard, 2001; Ryan, 1999; Scott, 2001). Following standard consumer theory, the MRS are calculated by partially differentiating an IUF with respect to the first attribute and then with respect to the second attribute, then calculating the ratio:

$$MRS_{X_1,X_2} = \frac{\partial V / \partial X_1}{\partial V / \partial X_2} \tag{2}$$

where V is an IUF estimated from a DCE and X_1 and X_2 are attributes of the good/service. The numerator (denominator) is interpreted as the marginal utility of attribute 1 (2). If time or price is used as the numeraire, the denominator denotes the marginal disutility of time or price; if price is used, we refer to the calculation as the 'implicit price' of each attribute.

On the face of it, this suggests that MRS calculated from DCEs measure the relative impact of attributes because they put the effect of each attribute on a common scale. However, the estimated common scale depends on the functional form of the IUF estimated. With linearly additive "main effects only models" traditionally estimated from DCEs in health economics, MRS provide the same ordering of relative attribute impact as comparing the size and significance of the raw attribute coefficients. That is, until recently DCEs reported in the health economics literature have estimated linearly additive, main effects only IUFs of the form:

$$V_j = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n \qquad (3)$$

In this case the MRS between two attributes is simply the ratio of the two attribute estimates, and the relative 'impact' is determined solely by the estimates. However, utility need not be linearly additive, in which case the MRS may be useful for measuring relative attribute impact as it will not simply depend on the size and significance of the estimates. For example, in the first empirical illustration in this paper we estimate a non linear IUF of the form:

$$\begin{aligned} V_j = {} & \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \beta_{12} X_1 X_2 + \ldots + \beta_{1n} X_1 X_n \\ & + \ldots + \beta_{n1} X_n X_1 + \ldots + \beta_{nn-1} X_n X_{n-1} \end{aligned} \qquad (4)$$

Equation (4) includes all main effects and all two-way attribute interactions for attributes X_1 to X_n. It is important to note that in such cases the MRS is not simply the ratio of two attribute parameter estimates because each attribute enters the utility function linearly and multiplicatively. For example, to calculate the marginal utility of X_1 we partially differentiate equation (4) with respect to X_1, including both the main effect of X_1 and terms where X_1 is interacted with other attributes. In fact, as discussed in Section 4, besides including all two-way attribute interactions described in equation (4), our empirical illustration also includes a non linear functional form for one main effect in the form of a quadratic term, which also must be taken into account when calculating MRS. Given attributes can have positive *and negative* impacts on utility and it is the size of the impact rather than the direction that is of interest; we use the absolute value of the MRS.

Hicksian Welfare Measures

The method of calculating Hicksian compensating variation (CV) in discrete choice random utility models is due to Small & Rosen (1981) and Williams (1971) and was introduced to the health economics literature to calculate welfare measures from DCEs by Lancsar (2002) and Lancsar & Savage (2004). In addition to calculating welfare measures for entire products/programs, the CV method can be used to measure relative impacts of each attribute (or level) in a common monetary metric by calculating willingness to pay or accept compensation for changes in a given attribute. Both forms of welfare measures are calculated using the utility estimates and attribute levels in the following expression:

$$CV = -\frac{1}{\lambda}\left[\ln\sum_{j=1}^{J} e^{V_j^0} - \ln\sum_{j=1}^{J} e^{V_j^1}\right] \tag{5}$$

where λ is the marginal utility of income; V_j^0 and V_j^1 are the value of the IUF for each choice option j before and after the policy change, respectively; and J is the number of options in the choice set.

The Hicksian CV essentially values the change in expected utility due to a change in the attribute(s) by weighting this change by the constant marginal utility of income (implications of using varying compared to constant marginal utility of income have been investigated in health economics by Lancsar & Donaldson (2004, 2005), and are due to Karlstrom (2000); McFadden (1999). This approach takes account of the uncertainty in the model about which alternative respondents will choose and/or whether respondents substitute between alternatives following a change in the desirability of one or more of the choice alternatives. Thus, the monetary values are calculated taking account of the probability that each alternative will be chosen by the average respondent.

For example, consider a choice between treatments A, B and C. If a policy change to be valued is an improvement in treatment A, the CV calculates willingness to pay (WTP) associated with this improvement taking account of the probability with which A is chosen before and after the policy change as the improvement could induce people who previously chose B and C to substitute to A. The clearest way to see the impact of the probability of choosing each alternative on the resulting WTP is if no one chooses A before or after the improvement the welfare gain associated with the improvement must be zero.

In addition to valuing changes in entire products/programs, one also can define the monetary equivalent associated with each level using the CV approach. We demonstrate this in the empirical application reported below by calculating the CV for a move from a base case where all attributes are set to their mean values (in the case of effects coding this will be zero (Louviere, Hensher, & Swait, 2000) and the IUF contains only alternative specific constants (ASCs)) to a case where the IUF includes ASCs plus each attribute level included one-at-a-time. The CV approach involves calculation of willingness to pay or accept, so we use the absolute value to estimate the size instead of the direction of the impact.

Probability analysis

Another way to measure the relative impact of each attribute (or level) is to calculate the probability of choosing an alternative given a particular attribute (level). The probability that respondents will choose each alternative in a choice set is now starting to be calculated in health economics (Hall, Kenny, King, Louviere, Viney, & Yeoh, 2002), and we show how this can be extended to measure relative attribute impacts.

In the context of a conditional logit model, the probability with which each alternative in the choice set is chosen is:

$$\pi_{i1} = \frac{e^{V_1}}{\sum_{j \in J} e^{V_j}} \qquad (6)$$

where π_{i1} is the probability that alternative 1 is chosen from a choice set containing J alternatives, V_1 is the estimated utility associated with alternative 1, V_j is the utility associated with each of the J alternatives in the choice set. If instead a non-closed form model is used like multinomial-probit or mixed logit, one would need to simulate the choice probabilities to approximate the integration over choice situations/respondents.

Predicted probabilities are used to predict market shares in marketing applications. In health economics the obvious analogy is to predict uptake or choice shares for the sample that provided the data. To predict beyond the sample requires recalibration of the experimental results if market data are available. Equation (6) also can be used to measure relative attribute impacts by first calculating the probability of choosing each alternative in a base case where all attributes are set to their mean values. As noted above, in the case of effects coding, the IUF contains only ASCs that represent the underlying preference for each alternative when all attributes are at zero. The probability of choosing a particular alternative based on its ASC plus the attribute (or level) of interest then can be calculated. Next, the percent change in the probability of choosing a particular alternative is calculated to measure the effect of each attribute over and above the base case. Systematically repeating this procedure over all attributes produces an implied ordering of the relative impact of each attribute with respect to its impact on the probability of choosing a particular alternative.

Best worst attribute scaling

BWAS involves a different choice task and was first used by Finn & Louviere (1992); it is sometimes called "maximum difference scaling" (Cohen, 2003; Cohen & Neira, 2003; Szeinback, Barnes, McGhan, Murawski, & Corey, 1999). It was introduced to health care by McIntosh & Louviere (2002), and the underlying theoretical properties were formally proven by Marley & Louviere (2005), and illustrated in health care by Flynn, Louviere, Peters, & Coast (in press). Respondents in BWAS tasks are presented with a series of experimentally designed alternatives one at a time and are asked to pick the best and worst attribute on offer in each alternative, based on the combination of levels that describe a particular alternative. Thus, in BWAS tasks choices are made within, rather than between alternatives. In a design with K attributes where L_k represents the number of levels of attribute k, the total number of best-worst pairs available to be chosen (when order does matter) is $2\sum_{i=1}^{K-1}\left[L_i \sum_{k=i+1}^{K} L_k\right]$, and is K(K-1) for any given alternative described by a combination of attribute levels. Theoretically, the pair of attribute levels chosen maximises the difference in the underlying attribute level utilities in that alternative. The BWAS model assumes that the relative choice probability of a given pair is proportional to the distance between the two attribute levels on the latent utility scale. So, BWAS is a difference model where one estimates utilities relative to a single attribute level instead of relative to an entire alternative (or the sample mean). Thus $\sum_{k=i}^{K} L_k - 1$ attribute levels are estimated relative to a remaining (base) level, placing the attribute levels on a common scale instead of the $\sum_{k=i}^{K}(L_k - 1)$ in a traditional DCE.

Estimating all attribute levels on a common scale allows BWAS to calculate mean utilities (again, on a common scale) that measure average attribute impacts. If effects coding is used, partial utility measures for attribute levels are simply deviations in utility from an attribute's impact and sum to zero. If dummy variables are used, attribute impacts are calculated by taking averages of the level estimates. Thus, BWAS estimates relative impacts of each attribute placing them on a common scale, which overcomes the inability to estimate impacts directly in traditional DCE model estimates.

4. Empirical applications

This section presents two empirical studies. The first demonstrates the first four methods outlined above in the context of a choice experiment and the second illustrates BWAS.

4.1 Empirical application one

4.1.1 Data

We demonstrate the first four methods using data from a choice experiment designed to investigate preferences of a sample of 64 people drawn from the general public in Calgary, Alberta, Canada, for treatment of cardiac arrest occurring in a public place. Treatment options were described by the five attributes in Table 3. A D-optimal design was used to construct 512 scenarios using the Burgess and Street approach (Burgess & Street, 2004; Street, Burgess, & Louviere, 2005). The experiment was blocked into 32 versions of 16 choice sets by randomly assigning choice sets to versions without replacement. Each version was viewed by an equal number of respondents, ensuring an orthogonal dataset, as discussed earlier. The design (and resulting dataset) allowed independent estimation of all main effects and all two-way attribute interactions, thereby allowing us to undertake partial log likelihood analysis and to estimate non-linear multiplicative IUFs. Some health related DCEs included interactions between single attributes and socio demographic characteristics, but to our knowledge this is the first DCE in health to include attribute-by-attribute interactions.

Table 3

Each choice set contained four treatment options. The first was the status quo treatment for cardiac arrest occurring in a public place; namely waiting for an ambulance to arrive. The three other treatment options described 'public access defibrillation' options (labelled PAD A, B and C); that is, having automated external defibrillators available in public places that can be used to restart the heart while waiting for an ambulance to arrive.

In each choice set respondents were asked to choose: 1) the best treatment; 2) the worst treatment; and 3) the best of the remaining 2 options. A standard first choice discrete choice model was estimated using the most preferred alternative per choice set as the dependent variable.

4.1.2 Results

We initially estimated a discrete choice model (DCM) with main effects only. All attributes were effects coded to visualise the results by plotting the estimated coefficients against the attribute levels to infer possible more parsimonious reduced form utility expressions. These results suggested that the effects of the survival levels were non-linear. Hence, we re-estimated a model specifying survival using linear and quadratic effects. We mean-centered the eight level price attribute (which equates the mean with the intercept), and effects coded the other attributes. We then included all 2-way attribute interactions in the IUF. Estimation results for the main effects only (Model 1) and main effects plus all 2-way attribute interactions (Model 2) are in Table 4.

Table 4

All main effects are statistically significant at the 1 percent level, except for location of care in both Models 1 and 2. Model 2 results suggest that few interactions are significant; the interaction of survival with each of: provider, from of payment, and library location are statistically significant. Model 2 results were used to illustrate the first four ways methods outlined in Section 3.

Partial log likelihood analysis

The results of the partial log likelihood analysis for Model 2 are presented in Table 5. We estimated 25 models in which we systematically included/removed each attribute level. Log likelihood values associated with each model are in column 2. For each attribute level, the change in log likelihood is in column 3, the relative effect is calculated as the percent change in log likelihood in column 4, the cumulative effect is in column 5 and the implied ordering of attribute level 'impacts' is in column 6.

Table 5

Not surprisingly, attributes with relatively large 'impacts' also are significantly different from zero. Interestingly, location is not 'impactful' individually, as it has a negligible impact on the log likelihood, but it is relatively 'impactful' when interacted with survival, which has the seventh largest impact on the log likelihood, highlighting the importance of testing interactions.

Survival accounted for 78 percent of the log likelihood; price, provider and form of payment collectively added another 14 percent. Adding interaction terms increased the log likelihood marginally. Variables that are not significant are included in the analysis because the lack of significance is taken into account in estimating partial log likelihoods. However, we exclude these attributes from further examination of relative impacts because the results suggest that they do not differ from zero.

Marginal rates of substitution

MRS between price and other attributes are in Table 6.

Table 6

When calculating MRS we took account of the non-linear IUF. By way of example, the attribute 'chance of survival' was decomposed into a linear and a quadratic term in the estimated IUF, and the interaction of these terms with 'provider of care', 'form of payment' and 'Library location' also were significant. Thus, the MRS between price and survival is no longer a ratio of estimated main effects parameters. Instead, it is obtained by partially differentiating the IUF with respect to survival and then with respect to price and setting survival to one percent to give:

$$MRS_{S,P} = \frac{\beta_{S\_lin} + 2 * \beta_{S\_quad} + \beta_{S\_lin\_pr} + \beta_{S\_lin\_f} + \beta_{S\_lin\_lib}}{\beta_P} \tag{7}$$

where β_P , β_{S_lin} and, β_{S_quad} are the price estimate, the linear survival estimate and the quadratic survival estimate, respectively, and $\beta_{S\_lin\_pr}$, $\beta_{S\_lin\_f}$, and $\beta_{S\_lin\_lib}$ are estimates of the interaction of survival with provider, form of payment and library location, respectively.

Hicksian Welfare Measures

The results of the welfare analysis using equation (5) are in Table 7. The welfare measures were calculated taking into account the non-linear, multiplicative nature of the estimated IUF. That is, the Vs in equation (5) include significant main effects, two-way attribute interaction terms and non-linear effects.

Table 7

Probability analysis

Results of the probability analysis are in Table 8 where predicted probabilities include both main effects and interactions. Predicted probabilities for the base case across the four alternatives are in row 3. The percent change in the probability from the base case to the case including each attribute one at a time are in columns 6 to 9, and the implied order of attribute 'impact' is in column 10.

Table 8

Comparison of the relative impact of attributes

The results of the four methods used to measure relative attribute impacts are summarised in Table 9.

Table 9

The chance of survival attribute was consistently ordered the most ‘impactful’ across all methods, with location consistently the least. In contrast, relative impacts of attributes provider and form of payment were less consistent across methods. This may reflect the view that individuals have fully formed preferences about the attributes they do and do not like but there is less certainty in preferences for attributes in between. In fact, the MRS between price and provider and price and form of payment differ by only $1 and were similar in the welfare analysis.

4.2 Empirical application two

BWAS uses a different choice task from DCEs, so we illustrate it using a second example. We conducted a simulation study to estimate relative attribute impacts and utility level estimates. A second aim was to compare these estimates with their true values (which are known in a simulation study) in terms of R-squared values in ordinary least squares regressions. However, we note that all choice model estimates are perfectly confounded with the unobservable random utility variance scale factor; hence, the BWAS estimates are a linear function of their true values.

4.2.1 Data

Five thousand Monte Carlo BWAS simulations were performed using the paired method of analysis (Flynn et al., in press). Six attributes were simulated, three with two levels and three with four. All were qualitative/categorical although one of the two-level attributes can be conceptualised as a price variable defined by constant plus slope parameter with no loss of generality. To illustrate the ability of BWAS to estimate impact and utility level estimates, the systematic components of utilities were chosen such that:

- Three attributes (attributes C, D and E) had similar weights but very different attribute level scale values or distances between levels (an attribute with two levels with almost identical utilities, one with two different levels and one with four very different levels);
- Two attributes (B and C) had similar level scale values but very different weights;
- One attribute (F) had the largest range of scale values but was not the most valued attribute overall;
- One attribute (B) had the largest impact but comparatively small level scale values.

To add random utility components, we drew from an Extreme Value Type I (gumbel) distribution (with mean adjusted to be zero). The experiment was conducted once with a small variance scale factor (0.25, equivalent to an EV1 beta parameter of four) and once with a relatively large one (1, equivalent to an EV1 beta parameter of one). Exploratory work suggested that variance scale factors within this range are sufficient to ensure that choices made are neither dominated by the systematic utility nor so random so as to make parameter estimates insignificantly different from zero.

We simulated 150 people in each of the 5000 simulations. In a traditional DCE, utilities of three levels can be estimated for each of the four-level attributes, with one level estimated for each two-level attribute. In BWAS four parameters can be estimated for a four level attribute (either a utility parameter for each level if dummy variables are used, or a utility parameter for three levels plus an overall attribute impact, or mean utility if effects codes are used). We used the second approach because the attribute utility impacts can be read directly from regression output, rather than calculating it as the mean of the attribute level estimates. Therefore, since 18-1=17 utility parameters can be estimated, five attribute impacts (6-1) and 12 level scale values (3+3+3+1+1+1) utilities were estimated using effects codes with the impact of attribute A and the lowest level of every attribute omitted. The natural log of the choice frequency for each unique best-worst pair is the dependent variable. Weighted least squares (WLS) with weights given by the choice frequencies (adjusted for the unbalanced design as detailed in (Flynn et al., in press) was used to estimate the model parameters.

4.2.2 Results

True and estimated systematic impacts and level scales and attribute rankings are in Table 10. True utilities (impacts and levels) are in columns 2 and 3, whilst ranking of attribute impacts is in column 4. Columns 5 and 6, and 8 and 9 contain the estimated attribute impact and partial utility measure for the levels, for each variance scale. For each attribute the lowest level was omitted from the regression model and its utility was calculated by multiplying the sum of the other level scale utilities by minus one. This illustrates the benefit of effects coding in BWAS, namely it provides estimates that are naturally mean centred (in this case around the attribute impact). The percentage of simulations for which each attribute was correctly ranked is in columns 7 and 10.

Table 10

Previous applications have obtained good results for sample sizes less than 100 (Coast, Salisbury, de Berker, Noble, Horrocks, Peters et al., 2006; Szeinbach, Barnes, McGhan, Murawski, & Corey, 1999), and exploratory work suggested that sample sizes of 150 and above were usually sufficient to guarantee that sampling variation was small enough to be consistent with the properties of BWAS. Table 10 shows that it performed well in the presence of both small and large variances in random utility components. As might be expected, it was attributes with similar impacts that were sometimes incorrectly ranked. Nevertheless the incorrect rankings were almost always only for adjacent ranks: the third largest attribute was rarely ranked fifth or vice versa.

In terms of recovering the true utilities, there is generally good agreement between true and estimated values (after taking into account the effects of the variance scale factor on the estimated values). The mean R-squared value from a regression of the estimates on true utilities was 96 percent for a variance scale of 1 and 98 percent for a variance scale of 0.25.

5. Discussion

We outlined and illustrated five ways to measure relative attribute impacts in stated preference studies. Some of these methods, or variations of them, have been used in the health economics literature, although not for the purpose of this paper. Some, such as the Hicksian CV and BWAS, only recently were introduced to health economics, (see Lancsar & Savage (2004) for the former and Flynn et al. (in press); McIntosh & Louviere (2002) for the latter). Others, such as probability analysis and MRS elicited from non-linear models are variations of methods currently used in health economics, while partial log likelihood analysis is novel to the literature.

The comparison in Table 9 highlights that orderings of relative attribute impacts were similar across methods. Although the method that is most appropriate to investigate the issue of relative attribute impact in part will depend on study objectives, each has advantages and disadvantages. If one includes interactions and a continuous attribute in a DCE, the MRS between a continuous attribute as numeraire and all other attributes provides a way to measure relative attribute impacts. However, as illustrated, calculations are more complex for non-linear and/or non-additive IUFs. Of course, one also may want to measure the relative impact of the attribute used as the common base (such as price or time), which cannot be done using MRS. Further, if only main effects are included in DCEs, MRS will provide the same order of impact as the raw attribute coefficients, suggesting that one may wish to consider another way to measure relative impact.

Hicksian CV provides a viable alternative to measure the relative impact of numeraire attributes like price/time because the marginal utility of income can be used to convert the impact of other attributes into monetary terms, rather than using one of the attributes. One might wish to consider this approach in cases where calculating welfare measures is a study objective independent of investigating relative impact.

Probability analysis is another way to measure relative attribute impacts. This approach also can be used to predict uptake (e.g. of a new screening program) or demand (e.g. for a new medication), which is relevant in many policy and commercial settings. However, as noted above, predicting market shares beyond a DCE sample requires recalibration with market data.

If one only wishes to measure overall attribute effects relative to one another, and there is no interest in policy measures like MRS, CV or uptake, the partial log likelihood approach provides a way to do this. The appeal of this approach lies in the fact that it does not require one attribute to be used as a common base, nor any attributes be quantitative. It also measures the impact of each attribute across its levels in a simple and intuitive way by estimating the relative contribution of each level to the explanatory power of the model.

Each of the four methods illustrated in the first empirical application involve straightforward additional analyses using the results of a standard DCE. They do not require a different experiment to be designed. Each puts the attributes (levels) on a common and therefore comparable scale, thereby facilitating statements about relative impacts of attributes of a good/service. They also provide information of interest over and above a comparison of relative impact.

If one is interested not only in placing attributes on a common scale but also in defining that scale, BWAS is appropriate. The simulation experiment illustrated how

BWAS allows direct estimation of relative attribute impacts in addition to partial utility estimates for attribute levels if effects codes are used. Simulation results suggest that random utility components must be large for BWAS to rank attributes incorrectly and even then attribute impact magnitudes and relative orderings are rarely affected. Furthermore, the relationship between estimated impacts and scales and true values remained strongly linear even when error components were large.

Decomposition of impact and scales is useful to investigate the effects of respondent-level covariates on utilities. For example, by understanding whether observed differences in utility between men and women are due to differences in attribute impacts or scale values, policy-makers can better tailor services to suit individuals; that is, policies to improve attribute levels among a target patient group may differ in scope/practicality from those to improve perceived attribute impacts. BWAS may be useful in taking such policies further by estimating respondent-level utilities. Greater individualisation of care necessitates better understanding of how much patients value attributes generally and how they value changes in the levels presented and BWAS provides a way to do this.

However, BWAS may necessitate designing a different or separate experiment. Due to the nature of the study in Section 4.1, we could not incorporate a BWAS task in the DCE, but it is possible to do so. Nevertheless, there are issues around combining BWAS and DCE data. For example, task differences imply that random components should differ in both size and nature (different variance scale factors), and such issues should be the subject of future research. Also, if one uses an independent BWAS task, it may not be possible to estimate some policy relevant measures such as welfare measures and predicted choices.

The use of independent BWAS designs also may have implications for statistical efficiency. Recent research suggests that multinomial logit efficiency is maximised when the number of attribute differences between alternatives is maximised in generic experiments (labelled designs can be handled within the existing theory by insuring that all 2-way interactions with the labels can be estimated) (Burgess & Street, 2004; Street et al., 2005). Presently, the efficiency of BWAS designs is unknown, and so further research is needed on this topic. That said, it may be that unfamiliarity with choosing between alternatives in some areas of healthcare may imply that BWAS tasks are less cognitively demanding than traditional DCE tasks, potentially leading to smaller random utility components and more precise utility estimates.

We showed how a BWAS task enables us to estimate an attribute's impact and its levels on the same scale. Attribute impact is related to the concept of attribute importance explored in Section 2, and current research into the theoretical properties of BWAS aims to set out necessary and sufficient conditions for the two to be equivalent.

Respondents make repeated choices in DCEs and BWAS, resulting in panel data. A limitation of our study is that the models reported in Table 4 do not take the potential for correlation among the error terms arising from the panel nature of the data into account. If the errors are correlated, this will impact the standard errors and asymptotic t-ratios as well as partial log likelihoods and probability results. Of course, the way DCEs are implemented, including types of tasks and task instructions impacts these correlations. Furthermore, one can take steps to minimise these correlations, such as administering only one choice set per person, but increasing the sample that receives the DCE. This also can be addressed by including random intercepts or estimating more complex models that allow error correlations, such as random coefficients models. However, such models require behavioural assumptions that may not hold in practice (Louviere, Street, Carson, Ainslie, Deshazo, Cameron et al.; 2002). An alternative approach that also can be considered is to develop ways to model single individuals, which avoids correlated errors across individuals (Louviere, Burgess, Street, & Marley, 2004).

Future research is required on appropriate ways to capture respondent heterogeneity within the BWAS framework. This might include random effects models, but discrete distributions of parameters implied by clustering, mixture and archetypal taxonomic methods that have been used in previous BWAS studies suggest that models for discrete parameter distributions also may be appropriate. Finally, research is underway that explores the use of BWAS to estimate individual-level parameters.

6. Conclusion

We discussed the fact that despite common practice, relative attribute impacts in DCEs cannot be inferred directly from parameter estimates due to confounds between the attribute impacts and utility scales on which attribute levels are positioned. We presented a menu of five methods that can be used to compare relative attribute impacts: partial log likelihood analysis; MRS in the context of non linear models; Hicksian welfare measures; probability analysis; and BWAS. The first four methods deal with issues of relative attribute impact in traditional DCEs by placing the effects on common and comparable scales. The fifth method, BWAS, uses a modified choice task to investigate relative impacts by decomposing an attribute's impact and scale.

We also illustrated estimation of a non linear IUF that included all two-way attribute interactions, which we believe is the first use of this type of IUF in the health economics literature. We also demonstrated how to derive MRS from such a non linear utility specification.

Finally, we discussed when it may be appropriate to use each of the five methods. Each has certain advantages, so choice of method in part will depend on other study objectives. Indeed, the methods should not be seen as mutually exclusive, but instead it is likely that there are many circumstances in which it would be advantageous to use a combination of methods.

References

Anderson, N. H. (1970). Functional measurement and psychophysical judgement. *Psychological Review,* 77(1970), 153-170.

Anderson, N. H. (1982). Methods of Information Integration Theory. New York: Academic Press.

Burgess, L., & Street, D. J. (2004). Optimal designs for choice experiments with asymmetric attributes. *Journal of Statistical Planning and Inference,* In press.

Coast, J., Salisbury, C., de Berker, D., Noble, A., Horrocks, S., Peters, T., et al. (2006). Preferences for aspects of a dermatology consultation. *British Journal of Dermatology,* 155, 387-392.

Cohen, S. (2003). Maximum Difference Scaling: Improved Measures of Importance and Preference for Segmentation. Sawtooth Software Conference Proceedings. Sequim, WA.

Cohen, S., & Neira, L. (2003). Measuring preference for product benefits across countries: Overcoming scale usage bias with Maximum Difference Scaling. ESOMAR Latin America Conference Proceedings. Amsterdam, The Netherlands.

Crouch, G. I., & Louviere, J. J. (2004). The Determinants of Conventions Site Selection: A logistic Choice Model from Experimental Data. *Journal of Travel Research,* 43, 118-130.

Finn, A., & Louviere, J. (1992). Determining the appropriate response to evidence of public concern: the case of food safety. *Journal of Public Policy and Marketing,* 11(1), 12-25.

Fishbein, M., & Ajzen, I. (1975). Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research. Reading, MA: Addison-Wesley.

Flynn, T., Louviere, J., Peters, T., & Coast, J. (in press). Best-Worst Scaling: What it can do for health care and how to do it. *Journal of Health Economics*.

Gyrd-Hansen, D., & Søgaard, J. (2001). Analysing public preferences for cancer screening programmes. *Health Economics,* 10(7), 617-634.

Hall, J., Kenny, P., King, M., Louviere, J., Viney, R., & Yeoh, A. (2002). Using stated preference discrete choice modelling to evaluate the introduction of varicella vaccination. *Health Economics,* 11, 457-465.

Karlstrom, A. (2000). Non-linear value functions in random utility econometrics. 9th IATBR travel behaviour conference. Australia.

Keeney, R. C., & Raiffa, H. (1976). Decisions with multiple objectives: preferences and value tradeoffs. New York: Wiley.

Lancsar, E. (2002). Deriving welfare measures from stated preference discrete choice modelling experiments. CHERE Discussion Paper No 48. Centre for Health Economics Research and Evaluation, University of Technology, Sydney.

Lancsar, E., & Donaldson, C. (2004). The importance of the indirect utility function and marginal utility of income in the derivation of welfare measures from discrete choice experiments. European Conference of Health Economists. London.

Lancsar, E., & Donaldson, C. (2005). A method to derive welfare measures from DCEs with non constant marginal utility of income. International Health Economics Association Conference. Barcelona, Spain.

Lancsar, E., & Savage, E. (2004). Deriving welfare measures from discrete choice experiments: inconsistency between current methods and random utility and welfare theory. *Health Economics Letters,* 13(9), 901-907.

Louviere, J. (1988a). Analysing Decision Making Metric Conjoint Analysis. Newbury Park: Sage Publications.

Louviere, J. (1988b). Conjoint analysis modelling of stated preferences: a review of theory methods recent developments and external validity. *Journal of Transport Economics and Policy,* 22(1), 93-119.

Louviere, J., Street, D., Carson, R., Ainslie, A., Deshazo, J. R., Cameron, T., et al. (2002). Dissecting the Random Component of Utility. *Marketing Letters,* 13(3), 177-193.

Louviere, J. J., Burgess, L., Street, D. J., & Marley, A. A. J. (2004). Modeling the choices of single individuals by combining efficient choice experiment designs with extra preference information. CenSoC working paper series 04-005 Sydney: Centre for the Study of Choice, Faculty of Business, University of Technology.

Louviere, J. J., Hensher, D. A., & Swait, J. D. (2000). Stated Choice Methods Analysis and Application. Cambridge: Cambridge University Press.

Louviere, J. J., & Islam, T. (2004). A Comparison of Importance Weights/Measures Derived from Choice-Based Conjoint, Constant Sum Scales and Best-Worst Scaling. CenSoC Working Paper: University of Technology, Sydney.

Lynch, J. (1985). Uniqueness issues in the decompositional modeling of multiattribute overall evaluations. *Journal of Marketing Research,* 22, 1-19.

Marley, A., & Louviere, J. (2005). Some probabilistic models of Best, Worse and Best-Worst choices. *Journal of Mathematical Psychology,* in press.

McFadden, D. (1999). Computing willingness to pay in random utility models. In J. Moore, R. R & J. Melvin (Eds.), Trade, Theory and Econometrics: Essays in Honour of John S. Chipman pp. 253-274). London: Routledge.

McIntosh, E., & Louviere, J. (2002). Separating weight and scale value: an exploration of best-attribute scaling in health economics. Health Economics Study Group Meeting, Brunel University.

Ohler, T., Le, A., Louviere, J. J., & Swait, J. (2000). Attribute Range Effects in Binary Response Tasks. *Marketing Letters,* 11(3), 249-260.

Ryan, M. (1999). Using conjoint analysis to take account of patient preferences and go beyond health outcomes: an application to in vitro fertilisation. *Social Science and Medicine,* 48(4), 535-546.

Ryan, M., & Wordsworth, S. (2000). Sensitivity of willingness to pay estimates to the level of attributes in discrete choice experiments. *Scottish Journal of Political Economy,* 47, 504-524.

Scott, A. (2001). Eliciting GPs' preferences for pecuniary and non-pecuniary job characteristics. *Journal of Health Economics,* 20, 329-347.

Shanteau, J. (1980). The Concept of Weight in Judgement and Decision Making - A Review and Some Unifying Proposals. Centre for Research and Judgement and Policy, Institute of Behavioral Research, University of Colorado, Boulder, Colorado.

Small, K. & Rosen, H. (1981). Applied Welfare Economics with Discrete Choice Models. *Econometrica*, 49(1), 105 - 130.

Street, D. J., Burgess, L., & Louviere, J. (2005). Quick and easy choice sets: Constructing optimal and nearly optimal states choice experiments. *International Journal of Research in Marketing,* 22, 459-470.

Szeinbach, S., Barnes, J., McGhan, W., Murawski, M., & Corey, R. (1999). Using conjoint analysis to evaluate health state preferences. *Drug Information Journal,* 33, 849-858.

Train, K. E. (2003). Discrete Choice Methods with Simulation. Cambridge: Cambridge University Press.
Williams, H. (1971). On the formation of travel demand models and economic evaluation measures of user benefits. *Environment and Planning A,* 9, 285-344.

Table 1: Scale value parameters and weights for attributes

Attribute	Scale value for each attribute is alpha + appropriate beta			Weights
	α	β_{low}	β_{high}	
Appointment length	0.1	0.2	0.9	Weight_len: 0.2
Doctor	0.5	0.4	0.5	Weight_dr: 0.3
Location	0.3	0.4	0.7	Weight_loc: 0.5

Table 2: Utilities of attribute levels and appointments

	Attribute levels			Scale			Scale* weights			Total Utility
Appointment	Lng	dr	loc	αlen + βlen	αdr + βdr	αloc + βloc	Length	Doctor	Location	
1	0	0	0	0.3	0.9	0.7	0.06	0.27	0.35	0.68
2	1	0	0	1	0.9	0.7	0.2	0.27	0.35	0.82
3	0	1	0	0.3	1	0.7	0.06	0.3	0.35	0.71
4	0	0	1	0.3	0.9	1	0.06	0.27	0.5	0.83
5	1	1	0	1	1	0.7	0.2	0.3	0.35	0.85
6	1	0	1	1	0.9	1	0.2	0.27	0.5	0.97
7	0	1	1	0.3	1	1	0.06	0.3	0.5	0.86
8	1	1	1	1	1	1	0.2	0.3	0.5	1

Table 3: Attributes and levels

Attributes	Levels
Chance of survival with treatment	• 6 out of 100 • 9 out of 100 • 12 out of 100 • 15 out of 100
Provider of care	• Trained responder • Non-trained responder
Location of treatment	• Shopping mall • Gym or other sports centre • Senior centre • Public library
Price	• $170 • $200 • $230 • $260 • $290 • $320 • $350 • $380
Form of payment	• Direct out of pocket payment (OPP) • A one off increase in taxation

Table 4: DCM results

	Model 1: DCM, main effects only (MNL)		Model 2: DCM, main effects + all 2-way interactions (MNL)	
Attribute	Coefficient	Standard Error	Coefficient	Standard Error
Main effects				
Out of pocket payment (OPP)	0.1563***	0.0438	0.2113***	0.0557
Tax payment	-0.1563		-0.2113	
Non-trained provider	-0.1798***	0.0448	-0.2142***	0.0556
Trained provider	0.1798		0.2142	
Location_shopping mall	0.0132	0.0989	0.0242	0.1132
Location_library	0.0002	0.0988	-0.1028	0.1210
Location_gym	0.0095	0.0992	0.0582	0.1113
Location_senior centre	-0.0229		0.0205	
Survival_linear	0.3929***	0.0289	0.4210***	0.0325
Survival_quadratic	-0.0517***	0.0145	-0.058***	0.0156
Price	-0.0037***	0.0008	-0.0035***	0.0009
Interactions				
Survival_linear x price			-0.0002	0.0004
Survival_quadratic x price			-0.00002	0.0002
Price x OPP			-0.0004	0.0007
Price x non-trained provider			-0.0003	0.0007
Survival_linear x OPP			-0.0578**	0.0281
Survival_quadratic x OPP			0.0067	0.0132
Survival_linear x non-trained provider			0.0573**	0.0286
Survival_quadratic x non-trained provider			0.0122	0.0139
OPP x non-trained provider			-0.0470	0.0485
Location_shopping mall x price			0.0003	0.0012
Location_library x price			-0.0004	0.0013
Location_gym x price			0.0011	0.0012
Survival_linear x location_shopping			0.0034	0.0496

mall				
Survival_quadratic x location_shopping mall			-0.0037*	0.0246
Survival_linear x location_library			0.1057	0.0564
Survival_quadratic x location_library			-0.0378	0.0265
Survival_linear x location_gym			-0.0404	0.0479
Survival_quadratic x location_gym			0.0126	0.0243
Location_shopping mall x OPP			-0.0033	0.0767
Location_library x OPP			-0.0103	0.0773
Location_gym x OPP			-0.0582	0.0762
Location_shopping mall x non-trained provider			0.0222	0.0780
Location_library x non-trained provider			0.0868	0.0784
Location_gym x non-trained provider			-0.0652	0.0784
Constant_PADA	-1.1160***	0.1028	-1.1699***	0.1059
Constant_PADB	-0.5614***	0.0869	-0.6084***	0.0904
Constant_PADC	-0.9755***	0.0987	-1.0281***	0.1017
Log likelihood	-1168.5779		-1155.4730	
McFadden R^2	0.1125		0.1224	

*** significant at 1%, ** significant at 5%, * significant at 10%

McFadden's R^2 is defined as $1 - (LL/LL_0)$, where LL is the value of the (simulated) log-likelihood function evaluated at the estimated parameters while LL_0 is the value of the log-likelihood function for a base model that only contains a non-random alternative-specific intercepts.

Table 5: Partial log likelihood analysis

Attribute level excluded from the analysis	Log likelihood	Partial effect - change in log likelihood	Relative Effect - % sum of change in log likelihood	Cumulative %	Order of impact
None (full model)	-1155.47302				
Survival (linear +quadratic)*	-1282.71212	-127.23910	0.78055	0.78055	1
Price*	-1163.75144	-8.27842	0.05078	0.83133	2
Non-trained provider*	-1163.21902	-7.74600	0.04752	0.87885	3
Out of pocket payment (OPP)*	-1163.01186	-7.53884	0.04625	0.92510	4
Survival (linear + quadratic) x non-trained provider*	-1159.59524	-4.12222	0.02529	0.95038	5
Survival_linear x OPP* or Survival_quadratic x OPP	-1157.86815	-2.39513	0.01469	0.96508	6
Surivival_linear x location_library* or Surivival_quadratic x location_library	-1157.56435	-2.09133	0.01283	0.97791	7
Location_library x non-trained provider	-1156.08407	-0.61105	0.00375	0.98165	8
OPP x non-trained provider	-1155.94241	-0.46939	0.00288	0.98453	9
Location_library	-1155.84212	-0.36910	0.00226	0.98680	10
Location_gym x price	-1155.84126	-0.36824	0.00226	0.98906	11
Survival (linear + quadratic) x location_gym	-1155.83660	-0.36358	0.00223	0.99129	12
Location_gym x non-trained provider	-1155.82044	-0.34742	0.00213	0.99342	13
Location_gym x out of pocket payment	-1155.76398	-0.29096	0.00178	0.99520	14
Survival (linear + quadratic) x price	-1155.69114	-0.21812	0.00134	0.99654	15
Price x OPP	-1155.63761	-0.16459	0.00101	0.99755	16
Location_gym	-1155.60886	-0.13584	0.00083	0.99838	17
Price x non-trained provider	-1155.57682	-0.10380	0.00064	0.99902	18
Location_library x price	-1155.52369	-0.05067	0.00031	0.99933	19
Location_shopping mall x non-trained provider	-1155.51346	-0.04044	0.00025	0.99958	20

Location_shopping mall x price	-1155.49661	-0.02359	0.00014	0.99972	21
Location_shopping mall	-1155.49576	-0.02274	0.00014	0.99986	22
Survival (linear + quadratic) x shopping mall	-1155.48535	-0.01233	0.00008	0.99994	23
Location_library x OPP	-1155.48186	-0.00884	0.00005	0.99999	24
Location_shopping mall x OPP	-1155.47397	-0.00095	0.00001	1.00000	25

* Significant in DCM

Table 6: Marginal rates of substitution

Attribute	MRS with P	Absolute value MRS	Order of impact
Chance of survival	-116	116	1
Non-trained provider	44	44	2
Out of pocket payment	-43	43	3
Library	-29	29	4

Table 7: Welfare measures

Attribute	CV [a]	Absolute value	Order of impact
Chance of survival	-80	80	1
Non-trained provider	-35	35	2
Trained provider	23	23	5
Tax	31	31	3
Out of Pocket Payment	-25	25	4
Library	-0.43	0.43	6

a In Canadian dollars

Table 8: Probability analysis

Attribute	Probability of choice					Percentage change in probability				Order
ASC	Wait for Ambulance	PAD A	PAD B	PAD C	Wait for Ambulance	PAD A	PAD B	PAD C	
Baseline (ASCs only)[a]	0.45201	0.1403	0.246	0.16168					
Price[b]									
Price =100	0.54003	0.11777	0.20649	0.13571	19.47%	16.06%	16.06%	16.06%	2
Chance of survival									
Linear surival + quadratic survival[a]	0.34046	0.16886	0.29608	0.1946	24.68%	20.36%	20.36%	20.36%	1
Provider									
non-trained	0.49109	0.1303	0.22846	0.15015	8.65%	7.13%	7.13%	7.13%	5
Trained	0.39969	0.1537	0.26949	0.17712	11.57%	9.55%	9.55%	9.55%	4
Form of payment									
Out of pocket	0.41435	0.14994	0.26291	0.17279	8.33%	6.87%	6.87%	6.87%	6
Tax	0.50468	0.12682	0.22236	0.14614	11.65%	9.61%	9.61%	9.61%	3
Location									
linear survival x location 2	0.426	0.14696	0.25768	0.16936	5.75%	4.75%	4.75%	4.75%	7

a Baseline sets all attributes to their mean levels

b The change in predicted probability for price calculated for a move from the mean price of $275 to $100.

Table 9: Ordering of relative impact of attributes across methods

Attribute	Partial LL	MRS	Welfare measure	Probability
Main effects				
Survival	1	1	1	1
Price	2			2
Form of payment	4	3		
• Out of pocket			4	6
• Tax			3	3
Provider	3	2		
• Non-trained			2	5
• Trained			5	4
Location				
• Mall				
• Library		4	6	7
• Gym				
• Senior centre				
Interactions[a]				
Survival x provider	5			
Survival x form of payment	6			
Survival x library	7			

a The effect of significant interactions are included in the effect of the main effect in the calculation of the MRS, welfare measures and probability analysis

Table 10: Utilities and ranking of impacts of attributes across simulations

	True Systematic Utilities & Ranks			Estimated Systematic Utilities[a] & Percentage Correctly ranked: Scale=0.25 (β =4)			Estimated Systematic Utilities[a] & Percentage Correctly ranked: Scale=1 (β =1)		
Attribute	Levels	Impact	Rank (impact)	Levels	Impact	Ranking (%)	Levels	Impact	Ranking (%)
A	-0.3	0.5	6	-0.175	-	100	-0.404	-	100
	-0.1			-0.053			-0.076		
	0.2			0.120			0.249		
	0.2			0.108			0.231		
B	-0.6	2.0	1	-0.320	0.973 (0.043)	100	-0.398	2.593 (0.098)	100
	-0.3			-0.162			-0.301		
	0.3			0.167			0.226		
	0.6			0.315			0.473		
C	-0.6	0.9	5	-0.385	0.271 (0.040)	88.60*	-0.984	0.752 (0.066)	96.32**
	-0.3			-0.198			-0.511		
	0.3			0.186			0.486		
	0.6			0.397			1.009		
D	-0.05	1.0	4	-0.019	0.330 (0.050)	82.24*	0.129	0.874 (0.089)	96.30**
	0.05			0.019			-0.129		
E	-0.6	1.1	3	-0.399	0.402 (0.049)	93.32*	-0.989	1.090 (0.080)	99.98**
	0.6			0.399			0.989		
F	-1.5	1.5	2	-0.955	0.580 (0.045)	100	-2.396	1.340 (0.063)	100
	1.5			0.955			2.396		

[a] Standard errors in brackets

*The maximum number of times an attribute was incorrect by two whole rankings was 9 (0.18% of simulations).

**3.68% of simulations ranked D (4th) and C(5th) the wrong way round whilst a further 0.02% of simulations ranked E(3rd) and D(4th) the wrong way round.