Title: Estimating individual level discrete choice models and welfare measures using best worst choice experiments

Authors: Emily Lancsar[1] and Jordan Louviere[2]

1. Business School (Economics) and Institute of Health and Society, Newcastle University
2. Centre for the Study of Choice, Faculty of Business, University of Technology, Sydney

Abstract

To date discrete choice experiments have been used to estimate aggregate level models of preferences and the welfare measures derived from such aggregate models are therefore for the average or representative consumer. Evidence exists to suggest that in fact individuals have heterogeneous preferences and willingness to pay (WTP) for the same good or service. As such, various methods have been used to try to disaggregate these aggregate level choice models and aggregate WTP to allow for heterogeneity. This paper proposes a new way to allow for preferences and WTP to differ over individuals, namely by estimating a model per individual. This paper is one of the first to estimate individual level choice models and the first in a health setting. This is achieved by combining highly efficient experimental design with a best worst choice experiment (BW-DCE). BW-DCEs are a form of choice experiment designed to elicit additional information per choice set. We harness this additional information to estimate a discrete choice model for each individual in our sample in the context of treatment for cardiac arrest occurring in a public place. From these individual level preference models we estimate changes in welfare per individual. We present preliminary results and highlight key issues for discussion.

Work in progress. Please do not quote without the authors' permission

Acknowledgements

The data used to demonstrate the methods proposed in this paper come from a choice experiment designed to investigate preferences for cardiac arrest in which Cam Donaldson and Gillian Currie are involved. We also thank Cam Donaldson for comments on a draft of this paper. The experimental design used in this study was developed by Leonie Burgess and Deborah Street. The authors benefited from discussions with Tony Marley on modelling sequential Best-Worst choices.

1. Introduction

Discrete choice experiments (DCEs) are now well-established in applied economics as a method to elicit preferences and monetary values associated with goods/services and the attributes/levels that comprise such goods/services (Viney et al, 2002; Ryan & Gerard, 2003; Lancsar and Louviere, 2008). To date, DCEs have primarily been used to estimate aggregate preference models; that is, choice data from individuals in the sample of interest are pooled, and a population model is estimated. Thus, welfare measures derived from such pooled models are for an average or representative consumer. However, it is not unreasonable to expect individuals to have heterogeneous preferences for the same good/service and heterogeneous willingness to pay (WTP). In this case, a one-size-fits-all model is unlikely to be appropriate for everyone in the sample and indeed an average model may not in fact represent the preferences of any one individual.

Thus, recognition of a need to account for heterogeneity when estimating choice models from DCEs is growing. Various methods have begun to be used to try to disaggregate pooled level (representative agent) choice models to allow for both mean and variance heterogeneity. One approach is to try to capture "observable" heterogeneity with socio-demographics/covariates included as interactions with attributes or alternative specific constants. Another approach is to capture unobserved preference heterogeneity by allowing parameters to vary randomly across individuals. This involves assuming a distribution of attribute preferences and estimating the mean and standard deviation of the distribution. The assumed distribution can be continuous, in which case a random parameters specification is typically implemented as a mixed logit (MIXL) (Train & Revelt, 1998) which is popular in health economics (e.g. Kajer et al, 2008; Hole, 2008; Hall et al, 2006; King et al, 2007) or hierarchical Bayes model; or it can be discrete implemented as a latent class model (Swait and Adamowics, 2001; Hole, 2008). A key issue in modelling preference heterogeneity in this way is a priori assumptions must be made about the appropriate form of the preference distribution or the number and possibly type of classes/segments of consumers in the sample population. Additionly, there may be heterogeneity in scale or equivalently, the error variance; approaches to capture these differences include heteroscedastic error variance models (DeShazo and Fermo; 2002, Swait and Adamowics, 2001) or the new G-MNL model (Fiebig et al, 2008).

Similarly, where Hicksian welfare measures are estimated from DCEs, the standard approach has been to do so using a representative agent model that produces a single measure of welfare gain, or willingness to pay (WTP), for a representative consumer. Less attention has been paid to trying to disaggregate welfare measures estimated from such aggregate or representative agent models to account for heterogeneity in WTP. A key component in estimating measures of welfare gain is the marginal utility of income which is essentially used to measure changes in expected utility arising from a change in the good/service in monetary terms. This is usually assumed constant over respondents however differences in income are likely to be a key driver in differences in individuals' valuations of goods/services or their associated attributes. Thus, some studies have investigated ways to allow for WTP heterogeneity by allowing the marginal utility of income to vary over individuals or income groups (Karlstrom 2000; Lancsar and Donaldson, 2004). Another approach has been to simulate WTP per individual using an estimated MIXL model, taking draws from estimated distributions of random parameters, the WTP values are simulated and the average WTP calculated as the mean over the replications (e.g. Lancsar et al, 2007a).

The purpose of this paper is to describe and discuss work in progress on a new way to account for heterogeneity in preferences and welfare measure derived from discrete choice models. This new approach involves estimating separate choice models for each sampled individual and using these models to estimate measures of welfare change for each individual. In contrast to previously cited approaches that try to decompose aggregate level preference models in a top town manner, estimation of models for individuals can be viewed as a bottom up approach to capturing heterogeneity.

As Louviere, et al (forthcoming) note, best worst discrete choice experiments (BWDCE) can be combined with efficient experimental designs for DCEs (Street and Burgess, 2008) to estimate choice models for individuals. BWDCEs are a type of discrete choice experiment designed to elicit more choice information per choice set. We use this extra information to estimate discrete choice models for each individual in a sample who evaluated treatments for cardiac arrest occurring in a public place. This approach allows individuals to have different preferences, which facilitates empirical estimation of the preference distribution instead of assuming a distribution. As noted below, estimation of individual preference models highlights that individuals can and do use a variety of decision rules when completing choice experiments which may not involve compensatory decision making. To date one published paper has estimated individual level models using BWDCEs (Louviere et al, forthcoming). The current paper uses the same data collection method as Louviere et al (forthcoming) but suggests a new method of analysis, sequential best worst analysis.

From these individual level models we calculate Hicksian measures of welfare change (or WTP) per individual arising from changes in the treatment options (rather than simulating them as done in a mixed logit approach) As far as we are aware, this is a first in the context of DCEs.

The rest of the paper is organised as follows. Section 2 discusses BWDCEs and how they can be used to estimate individual level models in general and presents a new way to estimate them in particular. Section 3 discusses the Hicksian method of welfare analysis. Section 4 describes the empirical study used to illustrate these new approaches. Preliminary results for individual models and individual welfare analysis are presented in Section 5, followed by discussion of the results and the method in Section 6. Section 7 concludes.

2. Estimating individual level models using BWDCE

A new way to investigate heterogeneity in discrete choice models and welfare measures derived from such models is to estimate a model per individual, allowing direct estimation of preference parameters and WTP for each individual. A previous impediment to estimating individual level models was a general belief that many choice sets per person must be observed to obtain sufficient choice data, rendering individual-level models impractical in field typical applications of DCEs. Asking individuals to evaluate more choice sets would produce more choice data and improved statistical precision, however this would be at the expense of greater task demands and respondent burden, which may lead to less statistical efficiency (Louviere, et al forthcoming). BWDCEs represent a compromise whereby more questions are asked about the options in each choice set instead of using more choice sets. While there is no free lunch, the marginal cost of asking more questions (in terms of task demands) is likely to be less than the marginal cost of adding more choice sets, especially as the number of additional sets required is likely to be large.

BWDCEs are a type of best-worst scaling (BWS) of which there are three types. Each type involves asking respondents to choose the best and worst from a set of three or more items. The first type is called BWS object scaling (Finn and Louviere, 1992; Marley and Louviere, 2005), which involves asking respondents to choose the best and worst object or issue (which are not decomposed into attributes) from sets. The objective is to measure the objects or issues on some latent scale. For example, the importance of various issues (such as the environment, health care, taxation, etc) to citizens; or perhaps the best area to target government spending (such as education, defence, health, etc). A second type is best worst attribute scaling, or what has more recently been termed BWS profile scaling (Marley, Flynn and Louviere, 2008), which has been used in health economics to measure attribute levels on common scales, such as McIntosh &

Louviere (2002), Flynn et al (2006), Flynn et al (forthcoming) and Lancsar et al (2007b). This case involves respondents evaluating single profiles one-at-a-time and choosing the best and worst attributes (levels) in each profile. The third type is the BWDCE case described and applied in this paper, and in Louviere et al (forthcoming) as well as Lancsar and Louviere (2005) where the latter presented the results of a pilot of the study described in Section 4 below. A key advantage of the BWDCE case is that it can be used to estimate individual level models.

BWDCEs are a type of DCE designed to elicit more information per choice set by asking more questions per choice set. More specifically, instead of asking respondents to choose the best (most preferred) option in each choice set, BWDCE tasks ask respondents to also choose the worst (least preferred) option in each choice set. This gives a semi order of preference over options; repeating best and worst questions for smaller subsets of the original choice set provides a complete order. For example, if a choice set has 5 options {A, B, C, D, E} and suppose A is most and E is least preferred, asking respondents to choose the best and worst options provides a semi order in that we know that A>B, A>C, A>D, A>E, B>E, C>E, D>E, but we do not know the order of B, C or D. If respondents now choose the best and worst of the remaining 3 alternatives we obtain a complete order. For example, if B is second best and D is second worst, the implied order is A>B>C>D>E.

There are several ways to use the additional preference information to estimate pooled and/or individual level models. For example, BWDCEs can be viewed as a way to obtain an implied preference order that differs from asking for direct rankings. That is, it may be that choosing extremes is cognitively easier than a full ranking (Louviere et al, forthcoming). This ordering can be used to expand the available data by taking advantage of the fact that each choice set contains several combinations of options not explicitly evaluated by respondents. Specifically, a choice set of size J can be expanded into 2^J (minus one null and J singleton) implied subsets of options. In the case described above with 5 options, there are 2^5-6 or 26 implied subsets. The implied order obtained from BWDCE questions can be used to infer choices in the implied subsets in order to expand the amount of data that can be used to estimate conditional logit (clogit) models (assuming respondents are consistent). Or, equivalently, weights can be calculated from the implied subsets and orders with such weights used to expand the data by estimating weighted clogit models. These approaches are explained in more detail in Louviere et al (forthcoming), who use them to estimate individual level models. The weighting approach was also used to estimate pooled choice models in Lancsar and Louviere (2005).

Alternatively, the implied preference orders obtained from BWDCE could be used to estimate rank order logit models (Beggs, Cardell & Hausman, 1981), which is illustrated for comparative purposes in the empirical application below. That is, the probability of any ranking of options can be expressed as a product of conditional logit formulas. For example, for the preference order A>B>C>D>E, the probability of this ranking is the logit probability of choosing A from the set {A B C D E} times the probability of choosing B from the remaining options {B C D E}, times the probability of choosing option C from the remaining options {C D E}, times the probability of choosing alternative D from the remaining {D E}:

$$\Pr(ranking\ \ A,B,C,D,E) = \frac{e^{V_A}}{\sum\limits_{j=A,B,C,D,E} e^{V_j}} * \frac{e^{V_B}}{\sum\limits_{j=B,C,D,E} e^{V_j}} * \frac{e^{V_C}}{\sum\limits_{j=C,D,E} e^{V_j}} * \frac{e^{V_D}}{\sum\limits_{j=D,E} e^{V_j}} \qquad (1)$$

Where V is the indirect utility function estimated in the choice model. This also is known as an "exploded logit" model (Chapman & Staelin, 1982) because the ranking probability is written as the product of first choice probabilities for successively remaining options. In this way, BWDCEs can be used to obtain an implied ranking that can be modelled as if "best" is chosen from successively smaller choice sets.

This paper presents another way to estimate individual level models that combines BWDCE as a data collection method and as a method of analysis that directly models the way data are collected (and without expanding the data beyond that explicitly provided by respondents). That is, we model the sequential choice of best, then the worst of the remaining options, then the best of the remaining options and the worst of the remaining options, etc, in each choice set. Similar to rank ordered logit, the probability of a best worst preference ordering can be expressed as a product of conditional logit formulas. That is, the probability of the preference order illustrated above is the logit probability of choosing A as best from the set {A B C D E} times the probability of choosing E as worst from the remaining options {B C D E}, times the probability of choosing B as best from the remaining options {B C D}, times the probability of choosing D as worst from {C D} which can be written as:

$$\Pr(best - worst\ ording\ \ A,B,C,D,E) = \frac{e^{V_A}}{\sum\limits_{j=A,B,C,D,E} e^{V_j}} * \frac{e^{-V_E}}{\sum\limits_{j=B,C,D,E} e^{V_j}} * \frac{e^{V_B}}{\sum\limits_{j=B,C,D} e^{V_j}} * \frac{e^{-V_D}}{\sum\limits_{j=C,D} e^{V_j}} \qquad (2)$$

Equation (2) requires the assumption that the scale (utility) of worst is the negative of the scale of best. As with rank order/exploded logit, McFadden's conditional logit model is nested as a special case.

$$P(Y_i = 1) = \frac{e^{V_{i1}}}{\sum_{j=1}^{n} e^{V_{ij}}}, \quad j = 1,...,n \tag{3}$$

3. Estimating individual level Hicksian Welfare measures

Estimating individual level models offers several advantages as noted below; however, it also provides an ability to undertake welfare analysis at the individual level instead of estimating measures of welfare change for a representative or average respondent.

In particular, for each individual level model, individual measures of the change in welfare for a change in the attributes and/or alternatives can be estimated directly using the same framework as used for pooled models. That is, using the Hicksian Compensating Variation due to Small and Rosen (1981) (see Lancsar and Savage 2004):

$$CV = \frac{1}{\lambda}\left[\ln \sum_{j=1}^{n} e^{V_j^0} - \ln \sum_{j=1}^{n} e^{V_j^1}\right] \tag{4}$$

where λ is the marginal utility of income; V_j^0 and V_j^1 are values of the indirect utility function for each choice option j before and after a policy change, respectively; and J is the number of options in the choice set. The Hicksian CV essentially values a change in expected utility due to a change in attribute(s), by weighting this change by the marginal utility of income, or its proxy. It takes account of uncertainty about which option in a choice model respondents will choose and/or whether respondents substitute among options following a change in the desirability of one or more alternatives.

4. Empirical Study

This section presents a brief overview of the empirical study used to estimate individual models and welfare measures in the context of preferences for treatment for cardiac arrest occurring in a public place, that is, outside a hospital and/or home. This study used focus groups in April 2004 to develop attributes and levels and the broader study design, as well as two pilots. The first pilot in July 2004 collected data from 64 respondents sampled from the general public of Calgary, Canada, in a face to face mode in mini labs (respondents completed questionnaires in a central venue); results of which were published in Lancsar et al (2007b). After the first pilot it was decided to conduct the full study over the internet. As such, an online questionnaire was created and piloted in May 2007 with a sample of 100 respondents in Calgary and Edmonton, Canada, recruited from an internet panel. Following analysis of the pilot, the full study data were collected over the web in July 2007 from 898 individuals again sampled from Calgary and Edmonton, Canada recruited from an internet panel. Respondents evaluated 16 choice sets with 5 options per set. Each option was described by 3 attributes, 1 with 2 levels and 2 with 4 levels (see Table 1). An optimally efficient Street and Burgess design was used to construct four orthogonal main effects plans (OMEPs) consisting of 16 choice sets, which allowed independent estimation of all main effects and all two-way interactions. The two-way interactions are estimated by combining the 4 OMEPs, allowing estimation of non linear indirect utility functions.

Insert Table 1: Attributes and levels

The first option in each choice set was the status quo treatment for cardiac arrest in a public place, namely to wait for an ambulance to arrive. Status quo attributes and levels are also in Table 1. It is important to include a status quo as we were interested in which option would be chosen, not just a preference order. If a status quo is not included respondents also need to be asked if in fact they would choose their most preferred option to avoid problems associated with a forced choice and to allow estimation of a demand model. The importance of a status quo for policy and welfare analysis was discussed by Lancsar and Louviere (2008) with various ways to model the status quo presented in Ryan & Skatun (2004) and King et al (2007).

The other 4 options per choice set described different versions of what is called ‘public access defibrillation’, that is having an automated external defibrillator (AED) available in various public places (indoor shopping malls, gyms/sports centres, public libraries and senior centres) to be used while waiting for an ambulance to arrive. These options include use of an AED and usual EMS, not instead of EMS. Respondents were asked to consider the treatments they would prefer if they were having a cardiac arrest in a public place and to assume that any other possible choice factors not included, such as location (which was not significant in the pilot) were constant across choice options.

In each choice set respondents were asked four questions: 1) to choose the best option; 2) to choose the worst option; 3) to choose the best of the remaining 3 options; and 4) to choose the worst of the remaining 2 options. Thus, a full ranking of the 5 options was obtained in each of the 16 choice sets.

5. Results

This section presents the results of the individual choice modelling analysis and the associated welfare analysis. A number of modelling approaches were explored in the estimation of the individual choice models and below we present the results for the sequential approach, equation (2) in Section 2. As a comparison we include results obtained by treating BWDCE as a way to obtain an implied ranking and using this ranking to estimate a rank order or exploded logit model as outlined in equation (1) in Section 2. Summary results of individual level welfare analysis are then presented.

Convergence status, perfect predictability and an investigation of decision rules

Individual level utility models were estimated as a function of the three attributes: survival, provider and price. An alternative specific constant (ASC) was included to capture choice of any of the AED alternatives relative to the status quo. This is a nested dummy term designed to induce correlation in errors across the designed options to mimic the substitution patterns of a nested logit model similar to King et al (2007):

$$U_{ij} = ASC_i + X'_{isj}\beta + \varepsilon_{isj} \quad (5)$$

where ASC is the alternative-specific constant and X denotes the attributes (survival, price and provider). Subscript i denotes individuals, j denotes choice options and s denotes choice sets. Estimating equation (5)[1] using the sequential BWDCE method of equation (2) produced the best convergence results of the methods investigated. In particular, coding price as continuous and survival and provider with effects codes resulted in convergence of 815 individual models, while models for 83 respondent or 9 percent of the sample, did not converge. It is worth noting that coding survival as a continuous variable produced higher model convergence rates, as might be expected due to estimating less terms. In particular, models for 840 of the respondents converged,

[1] Individual level models containing all two-way interactions were also estimated but as they are not supported by the experimental design at the individual level they are not reported here.

leaving 58 individuals, or 6 percent of the sample, that did not. These are compared to the convergence results from exploded logit in which only 779 models converged and 119 did not.

There are two main reasons why models may not converge. The first is collinarity However, the use of OMEPs for the DCE means this cannot be the problem. The second is due to perfect predictability. This can occur if at least one choice is perfectly predicted/explained by at least one attribute level. In addition to creating problems for convergence, perfect predictability can also result in convergence with large beta coefficients/standard errors. In addition to the 83 models that did not converge with the first sequential best worst model, a further 166 converged, but with large beta coefficients (that is, estimates larger than +/-6 which essentially represents infinity for dummy/effects coded variables and/or a price estimate below -0.016). This may be due to the likelihood ending up in the flat part of the logit (sigmoid) distribution. In fact almost all models with large estimates exhibited convergence problems, and these estimates generally were not significant.

A solution to perfect predictability used with other forms of data involves combining categories of independent variables in order to create variation. For example, in a revealed preference setting if a binary (yes/no) dependent variable were regressed on a number of dummy variables for, say, employment status, perfect predictability can occur if all professionals choose the yes option meaning there is perfect prediction between the dummy and the outcome variable. Here groups of employment status, say professions and semi-professionals, could be pooled to obtain variation. This would be more difficult to do at the individual level and would also forgo the ability to explore preferences over attribute levels, a key objective of the experiment. There was sufficient variation when data were pooled across individuals to ensure convergence. We return to the pooled level analysis below. We also note that multinomial probit analysis of the convergence status of all individuals in the sample revealed that those whose models did not converge were significantly more likely to be better educated (masters educated relative to only primary or secondary), be in better health status (excellent relative to good self reported health), be a student (relative to being retired) or have a mortgage (relative to renters).

Exploring types of decision rules

Exploring the data for individuals whose models did not converge in more detail revealed that many seemed to use deterministic rather than compensatory choice rules. Of the 83 people who did not converge 65 (or, 78 percent) could be fit with a deterministic decision rule. Examples of such rules are always choosing the same alternative (for example, status quo) or always choosing based on one level of one attribute. For example, always choosing an option with the best (worst) level on the chance of survival attribute. Such preferences might be viewed as lexicographic, however strictly speaking lexicographic preferences involve also making decisions based on the next most preferred attribute if the most preferred attribute level is the same across all options in a choice set.

Summary results of the individual level models

Summary results for the 815 models estimated using sequential best worst choice models are in Table 2.

Insert table 2: Summary statistics of the n=815 individual level sequential BWDCE models

Summarised over all respondents, each coefficient has the expected sign: higher survival chances are preferred to lower, trained providers are preferred to non-trained providers and lower prices are preferred to higher prices. There is a very large range of estimates, which indicates that some models that converged yielded very large estimates (but they were generally not significant). The results reveal significant heterogeneity as evidenced by the large standard deviations.

For comparison, a summary of the exploded logit results are in Table 3. The estimates exhibit the same expected signs as those associated with the sequential best worst method but the magnitude of the estimates differs. As noted above, exploded logit exhibited poorer convergence; and yielded a wider range of individual estimates, suggesting that more individual models converged with very large estimates.

Insert table 3: Summary statistics of the n=779 individual level exploded logit models

As previously noted, individual models converging with very large coefficients could be due to problems of (close to) perfect predictability between one or more predictor variables and choice outcome(s). To examine this possibility, Table 4 presents summary statistics for individual-level sequential best worst choice models after censoring, or removing individuals with very large estimates (as defined above).

Insert Table 4: Summary statistics of the n=649 individual level sequential BWDCE models with censoring.

Interpretation of the average of the individual level estimates has not changed. That is, respondents preferred options with higher survival chances, lower prices and trained instead of non-trained providers. However, Table 4 reveals that the magnitude of the average estimates have changed, as has the range of the estimates and their standard deviations. Considerable heterogeneity remains, as evidenced by the range of the estimates.

A visual representation of the empirical preference heterogeneity results for each attribute is in Figure 1 for both censored and uncensored results.

Insert Figure 1: Distribution of individual level preference coefficients

Both distributions indicate considerable preference heterogeneity across individuals. It is interesting to note that some distributions do not appear consistent with normality; for example, the provider distribution is skewed to the right because most respondents preferred trained to non-trained providers, and a few respondents have positive price estimates. The ASC estimates represent a preference for a new treatment over the status quo; Tables 2, 3 and 4 indicate that this estimate was positive for some and negative for others, reflecting the fact that some people had strong preferences for or against the status quo. Here a normal distribution is likely to fit quite well.

Comparison to pooled level analysis

The individual level data were pooled over the 898 respondents and several models estimated, including a pooled sequential best worst model (equation (2)) and an exploded logit model (equation (1)). In addition, just the first choice data were used to estimate clogit and MIXL models (the MIXL models were estimated with random main effects, some with full error correlation matrices). The pooled level models included 2-way attribute interactions and socio-demographic covariates. As the pooled data models are not the focus of this paper, we do not report the results to save space. However, we note that all main effects and two-way interactions were significant at the 1 percent level with the expected signs in all pooled models. The average of the individual level models exhibited larger effects than in the pooled level or representative consumer models but only slightly so in the case of the censored individual level results. These results are available from the authors on request.

Individual level Hicksian welfare analysis

For each individual level model we estimated the Hicksian compensating variation associated with a number of potentially interesting policy changes using equation (4). The negative of the price estimate was used to proxy the marginal utility of income because there is no variation in income within an individual and a such it is not identified in the econometric model. We estimated the CV (or WTP) for doubling the chance of survival from 6 in 100 to 12 in 100, and for a trained compared with an untrained provider using the censored individual models. Summary statistics are in Table 5.

Insert Table 5: Summary of individual level welfare analysis

Preliminary results indicate that an unweighted average across the 649 respondents was $188 Canadian dollars for trained compared with untrained providers and an unweighted average of $5618 Canadian dollars for doubling survival chances (reducing the risk of death). There is clearly wide variation in the welfare measures as indicated by the large standard deviation and much smaller median values.

6. Discussion

This paper presented a way to model preference heterogeneity in DCEs and conduct welfare analyses from DCEs that differs sharply from current top-down modelling methods. Namely, we used a bottom-up approach by which a separate choice model is estimated for each individual and aggregated across individuals instead of estimating a one-size-fits-all model and making distributional assumptions to decompose the individual differences. The bottom up approach models preference heterogeneity directly because each person has their own preference parameters and their own scale or error variance.

To estimate individual models and welfare measures we used optimally efficient experimental designs (Street and Burgess, 2008) and a new type of DCE, what we called a “BWDCE”. Within this BWDCE approach we also suggested a new way to analyse BWDCE data, sequential best worst analysis, which attempts to model the process that respondents used to make their choices by recognising that the way in which the responses are collected should inform the analysis of such responses. Compared with previous analysis methods used to estimate models from BWDCE data, the sequential approach that we proposed imposes fewer assumptions on the data and the statistical results reveal that there was much better model convergence compared to using BWDCE to obtain an implied ranking and estimating an exploded logit model. It also does not involve expanding the data beyond that provided by respondents.

Preliminary results suggest considerable individual preference heterogeneity regarding the attributes of treatment for cardiac arrest in a public place and for the status quo versus public access defibrillation options. Visualisation of the empirical attribute preference estimates distribution also suggests that some are non–normal and skewed. Thus, the ability to estimate individual level models can also inform the appropriate form of distribution (or types of classes) assumed in pooled level analysis treatment of heterogeneity. In addition, pooled level analysis and treatment of heterogeneity more generally may benefit form more complex model forms that explicitly recognise that individuals can differ not only in preferences but also in error variability.

Estimation of individual level models revealed that not all respondents traded-off all attributes, with some individuals seeming to use a deterministic decision rule to choose treatment options (eg, seemingly lexicographic orders), which was not obvious in a pooled analysis of the data. The context is also relevant because cardiac arrest results in death unless treatment successfully restarts the heart; hence, it perhaps is not surprising that the chance of survival mattered to respondents. It is important to note that lexicographic preferences are not irrational, but can cause estimation problems at the individual level as respondents do not trade-off different attribute levels (Lancsar and Louviere, 2006). Deterministic preferences in general are not a problem, it simply means that a probabilistic model built on the assumption of compensatory behaviour may not fit such preferences well.

Deterministic preferences relate also to the issue of censoring of which there were two levels in this analysis. The first was at the level of estimation in that not all models converged and the second involved summarizing the individual level results with and without those individual models that converged with implausibly large coefficients which may be driven by deterministic preferences. This is the subject of ongoing research which should not be surprising for a new and emerging modeling approach. Of course this has implications for the interpretation of results averaged over individual level models or WTP which excludes those individuals with deterministic preferences. We do not advocate ignoring such preferences but instead current work is investigating how best to include them. Research clearly is needed to determine the best way(s) to deal with these types of individuals.

The welfare results are also preliminary. Here we have simply discussed unweighted means that include some large individual level WTP values driven by small coefficients for some individuals on the monetary variable. Ongoing work will investigate the use of other measures of central tendency as well as issues around trimming outliers. Current work is also addressing the estimation of standard errors and confidence intervals for individual estimates.

7. Conclusion

This paper presented preliminary results from the estimation of choice models and WTP for single individuals as another way to investigate heterogeneity. To do this we harnessed a new type of choice experiment, BWDCE used both to collect the choice data and to guide the analysis. In particular, we have suggested a new type of analysis that extends rank order logit to directly model respondents' sequential best worst choices. This approach directly links the theory and analysis to the choice task and without expanding the data beyond that directly provided by respondents. Preliminary results suggest that this sequential best worst choice method is preferred to other models with which it was compared. The individual level model results also suggest that not all respondents made choices by trading off all attributes and instead some seem to have used deterministic decision rules which may be obscured in more traditional pooled analysis of choice data. The individual level models also facilitated the direct calculation (rather than simulation) of individual level WTP. The analytical results presented here are "work in progress" and the subject of ongoing research.

8. References

1. Viney R, Lancsar E, Louviere J. Discrete Choice Experiments to Measure Consumer Preferences for Health and Healthcare. Expert Review of Pharmacoeconomics and Outcomes Research. 2002 August;2(4):319-26.
2. Ryan M, Gerard K. Using discrete choice experiments to value health care programmes: current practice and future research reflections. Applied Health Economics and Health Policy. 2003;2(1):55-64.
3. Lancsar E, Louviere J. Conducting Discrete Choice Experiments to Inform Healthcare Decision Making: A User's Guide. PharmacoEconomics. 2008;26(8):661-77.
4. Revelt D, Train K. Mixed logit with repeated choices of appliance efficiency levels. Review of Economic Statistics. 1998;80(4):647-57.
5. Hall J, Fiebig DG, King MT, Hossain I, Louviere JJ. What influences participation in genetic carrier testing? Results from a discrete choice experiment. Journal of Health Economics. 2006;25(3):520-37.
6. Hole AR. Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment. Health Economics. 2008;27:1078-94.
7. King MT, Hall J, Lancsar E, Fiebig D, Hossain I, Louviere J, et al. Patient preferences for managing asthma: results from a discrete choice experiment. Health Economics. 2007;16(7):703-17.
8. Kjaer T, Gyrd-Hansen D. Preference heterogeneity and choice of cardiac rehabilitation program: Results from a discrete choice experiment. Health Policy. 2008;85:124-32.
9. Swait J, Adamowicz W. The influence of task complexity on consumer choice: a latent class model of decision strategy switching. Journal of Consumer Research. 2001;28(1):135-48.

10. DeShazo JR, Fermo G. Designing choice sets for stated preference methods: the effects of complexity on choice consistency. Journal of Envirnonmental Economics and Management. 2002;44:123-43.
11. Swait J, Adamowicz W. Choice environment, market complexity, and consumer behavior: A theoretical and empirical approach for incorporating decision complexity into models of consumer choice. Organizational Behavior & Human Decision Processes. 2001;86(2):141-67.
12. Fiebig DG, Keane M, Louviere J, Wasi N. The generalised multinomial logit model. Marketing Science. Conditionally accepted.
13. Karlstrom A. Non-linear value functions in random utility econometrics. 9th IATBR travel behaviour conference; 2000; Australia; 2000.
14. Lancsar E, Donaldson C. 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; 2004 September 2004; London; 2004.
15. Lancsar EJ, Hall JP, King M, Kenny P, Louviere JJ, Fiebig DG, et al. Using discrete choice experiments to investigate subject preferences for preventive asthma medication. Respirology. 2007a;12(1):127-36.
16. Louviere J, Street A, Burgess L, Wasi N, Islam T, Marley AAJ. Modeling the choices of individual decision-makers by combining efficient choice experiment designs with extra preference information. The Journal of Choice Modelling. Forthcoming.
17. Street DJ, Burgess L, Viney R, Louviere JJ. Designing Discrete Choice Experiments for Health Care. In: Ryan M, Gerard K, Amaya-Amaya M, editors. Using Discrete Choice Experiments to Value Health and Health Care Dordrecht, The Netherlands, : Springer; 2008. p. 47 - 72
18. Finn A, Louviere J. Determining the appropriate response to evidence of public concern: the case of food safety. Journal of Public Policy and Marketing. 1992;11(1):12-25.
19. Marley A, Louviere J. Some probabilistic models of best, worst, and best–worst choices. Journal of Mathematical Psychology. 2005;49(6):464 - 80
20. Marley AAJ, Louviere JJ, Flynn T. Probabilistic Models of Set-Dependent and Attribute-Level Best-Worst Choice. Journal of Mathematical Psychology. 2008;52:281-96.
21. Flynn T, Louviere J, Peters T, Coast J. Best-Worst Scaling: What it can do for health care and how to do it. Journal of Health Economics. 2007;26
171-89.
22. Flynn TN, Louviere JJ, Peters TJ, Coast J. Estimating preferences for a dermatology consultation using Best-Worst Scaling: Comparison of various methods of analysis. BMC Medical Research Methodology forthcoming.
23. Lancsar E, Louviere J, Flynn T. Several methods to investigate relative attribute impact in stated preference experiments. Social Science & Medicine. 2007b;64(8):1738-53.
24. McIntosh E, Louviere J. Separating weight and scale value: an exploration of best-attribute scaling in health economics. Health Economics Study Group Meeting, Brunel University; 2002 January 2002; 2002.
25. Lancsar E, Louviere J. Several methods for dealing with scale confound and efficiency in stated preference data with an empirical illustration. Health Economics Study Group. Newcastle upon Tyne, UK; 2005.
26. Beggs S, Cardell S, Hausman J. Assessing the potential demand for electric cars Journal of Econometrics. 1981;16:1-19.
27. Chapman RG, Staelin R. Exploiting rank ordered choice set data wtihin the stochastic utility model. Journal of Marketing Research. 1982;XIX:288-301.
28. Small KA, Rosen HS. Applied Welfare Economics with Discrete Choice Models. Econometrica. 1981;49(1):105 - 30.
29. Lancsar E, Savage E. Deriving welfare measures from discrete choice experiments: inconsistency between current methods and random utility and welfare theory. Health Economics. 2004;13(9):901-7.

30. Ryan M, Skatun D. Modelling non-demanders in choice experiments. Health Economics. 2004;13(4):397-402.
31. Lancsar E, Louviere J. Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? Health Economics. 2006;15(8):797-811.

Tables and Figures

Table 1: Attributes and levels

Attribute	EMS + AED Level	Status Quo EMS level
Chance of Survival	6, 9, 12 & 15 out of 100	6 out of 100
Price	\$350, \$900, \$1450, \$2000 (includes EMS)	\$350
Provider	trained non traditional (e.g. security guard), non trained non traditional (e.g. bystander)	Ambulance personnel

Table 2: Summary statistics of the n=815 individual level sequential BWDCE models

Variable	Mean	Std Dev	Min	Max
ASC	-0.52	10.41	-73.97	87.14
Survival: 15/100	5.74	20.83	-36.10	212.13
Survival: 12/100	3.38	13.02	-7.95	130.46
Survival: 9/100	0.34	6.90	-14.40	64.97
Price (Canadian Dollars)	-0.002	0.013	-0.14	0.002
Trained Provider	0.71	2.60	-10.09	47.51

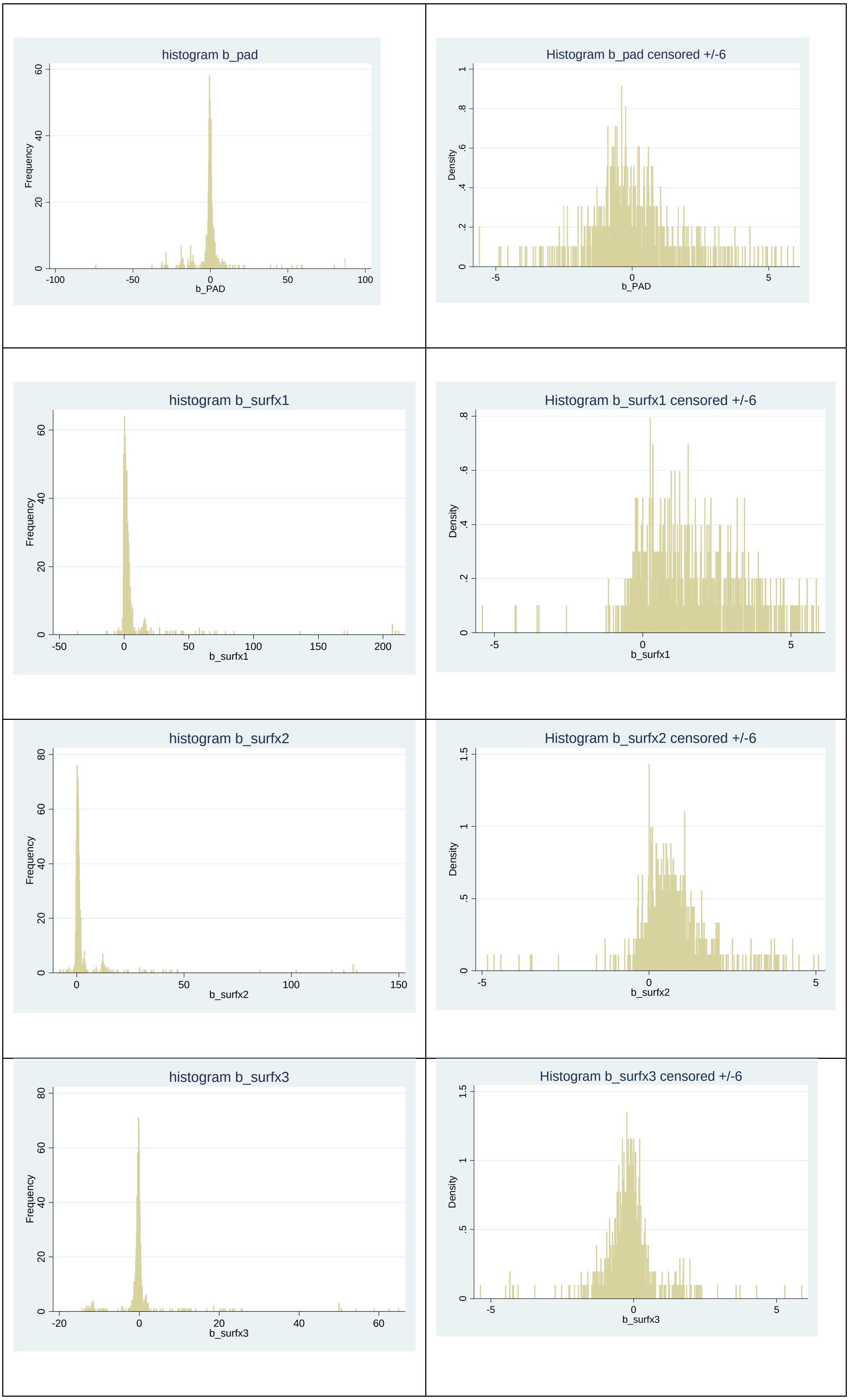
Table 3: Summary statistics of the n=779 individual level exploded logit models

Variable	Mean	Std. Dev.	Min	Max
ASC	2.236196	18.3937	-176.5	77.3275
Survival: 15/100	7.529205	16.079	-38.247	132.797
Survival: 12/100	2.484644	5.92804	-24.089	42.4877
Survival: 9/100	-2.027869	5.19513	-38.514	9.69096
Price (Canadian Dollars)	-0.004133	0.01466	-0.2869	0.00191
Trained Provider	0.31449	0.35053	4.52E-09	1

Table 4: Summary statistics of the n= individual level sequential BWDCE models with censoring.

Variable	Mean	Std Dev	Min	Max
ASC	0.14	4.62	-4.87	5.91
Survival_1	1.73	1.54	-2.56	5.93
Survival_2	0.73	0.79	-3.51	4.11
Survival_3	-0.21	0.61	-2.75	2.94
Price (Canadian Dollars)	-0.0006	0.0018	-0.01	0.002
Provider	0.43	0.63	-0.70	4.73

Figure 1: Distribution of individual level preference coefficients (uncensored & censored)



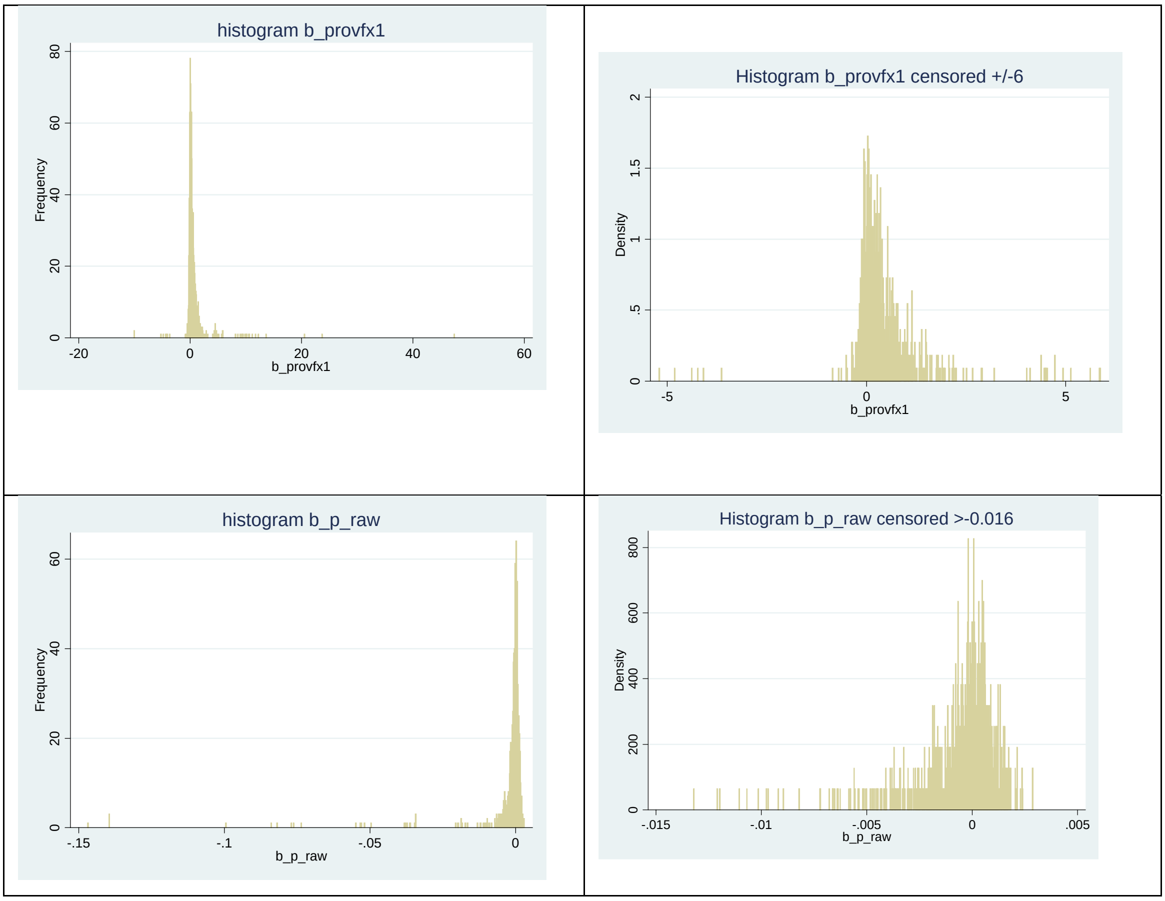


Table 5: Summary of individual level welfare analysis

	Trained rather than untrained provider	Double the chance of survival from 6 to 12 in 100
Mean	$188	$5618
Median	$-7	9
Std Dev	30773	116990
Obs	649	649