

NBER WORKING PAPER SERIES

ESTIMATING PATIENTS' PREFERENCES FOR MEDICAL DEVICES:  
DOES THE NUMBER OF PROFILE IN CHOICE EXPERIMENTS MATTER?

John Bridges  
Christine Buttorff  
Karin Groothuis-Oudshoorn

Working Paper 17482  
<http://www.nber.org/papers/w17482>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
October 2011

Funding for this research was provided by a grant from InHealth: The Institute for Health Technology Studies and by a faculty innovation award from the Johns Hopkins Bloomberg School of Public Health. In addition, the authors are grateful to Drs. John Niparko and Angela T. Lataille for their assistance in developing survey, Rick Li and team at Knowledge Networks for their assistance in data collection and Mattijs Lambooi and participants of the lolaHESG 2011 conference for their comments on an earlier draft of this manuscript. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2011 by John Bridges, Christine Buttorff, and Karin Groothuis-Oudshoorn. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Estimating Patients' Preferences for Medical Devices: Does the Number of Profile in Choice Experiments Matter?

John Bridges, Christine Buttorff, and Karin Groothuis-Oudshoorn

NBER Working Paper No. 17482

October 2011

JEL No. C91,I11,I18

**ABSTRACT**

**Background:** Most applications of choice-based conjoint analysis in health use choice tasks with only two profiles, while those in marketing routinely use three or more. This study reports on a randomized trial comparing paired with triplet profile choice formats focused on measuring patient preference for hearing aids.

**Methods:** Respondents with hearing loss were drawn from a nationally representative cohort, completed identical surveys incorporating a conjoint analysis, but were randomized to choice tasks with two or three profiles. Baseline differences between the two groups were explored using ANOVA and chi-square tests. The primary outcomes of differences in estimated preferences were explored using t-tests, likelihood ratio tests, and analysis of individual-level models estimated with ordinary least squares.

**Results:** 500 respondents were recruited. 127 had no hearing loss, 28 had profound loss and 22 declined to participate and were not analyzed. Of the remaining 323 participants, 146 individuals were randomized to the pairs and 177 to triplets. The only significant difference between the groups was time to complete the survey (11.5 and 21 minutes respectively). Pairs and triplets produced identical rankings of attribute importance but homogeneity was rejected ( $P < 0.0001$ ). Pairs led to more variation, and were systematically biased toward the null because a third (32.2%) of respondents focused on only one attribute. This is in contrast to respondents in the triplet design who traded across all attributes.

**Discussion:** The number of profiles in choice tasks affects the results of conjoint analysis studies. Here triplets are preferred to pairs as they avoid non-trading and allow for more accurate estimation of preferences models.

John Bridges  
Department of Health Policy & Management  
Johns Hopkins Bloomberg  
School of Public Health  
624 N. Broadway, Rm 451  
Baltimore, MD 21205  
and NBER  
jbridges@jhsph.edu

Karin Groothuis-Oudshoorn  
Department of Health Technology and Services Resea  
University of Twente  
7500 AE Enschede, Netherlands  
c.g.m.oudshoorn@utwente.nl

Christine Buttorff  
Johns Hopkins Bloomberg  
School of Public Health  
624 N. Broadway, Rm. 691  
Baltimore, MD 21205  
cbutterf@jhsph.edu

## 1. Introduction

The ever-increasing costs of health care have prompted most developed countries to establish evaluative bodies charged with assessing the value of drugs, therapies and devices. Traditional evaluative bodies, such as the UK's National Institutes for Health and Clinical Excellence, use cost-effectiveness of a technology compared to an arbitrary willingness-to-pay standard (e.g. Trowman et al. 2011). Emerging bodies include the Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (Institute for Quality and Efficiency in Healthcare), or IQWiG, in Germany and the Patient Centered Outcomes Research Institute (PCORI) in the US. Worldwide, these organizations are heeding increased calls for more inclusion of patients in the assessment of outcomes and in the deliberations over new medical technologies (Bridges et al. 2010, IQWiG 2008, Facey et al. 2010). To include patient input in a rigorous way, researchers and regulators have explored several stated preferences methods. These methods allow researchers to determine what aspects of treatments are most important to patients. This may go beyond side effects or cost to include wait times, travel costs or discomfort (Ryan 1999).

Hearing aids are at the intersection of these difficult issues over coverage, value of the technology and patient preferences. Hearing aids can improve the quality of life and functioning in older adults. However, the National Institute of Deafness and other Communication Disorders estimates only one in five people who could benefit from a hearing aid wear one in the US (NIDCD 2011). Part of this low level of use is a patient preference story. Hearings aids don't always work well, their owners might feel stigma for wearing them and they are expensive (Fitzpatrick and

Leblanc 2010, Bertoli 2009, Franks and Beckman 1985). The other part of the story is that Medicare, the major insurer for US adults over the age of 65, doesn't cover hearing aids (Medicare.gov 2011). Part of the goal in evaluating the value of new technologies is to aid payers in making coverage determinations, but it is someone unclear what types of studies reimbursement decisions are based upon, especially in the US. It is hoped that the increased focus of PCORI on patient-reported outcomes in determine the value of services will result in more coverage decisions using this information. Stated preference techniques provide just the sort of patient-level information needed to give a complete picture of the device's value.

In addition to offering substantive conclusions on the relative importance of different hearing aid characteristics, the main objective of this study was identify the impact that the number of profiles in a choice experiments has on the estimated preferences. Section 2 presents background information on the stated preferences methods, including conjoint analysis and discrete choice experiments. Section 3 discusses existing knowledge on the design of choice experiments. Section 4 discusses out methods, including survey design, study population and statistical analysis. Section 5 presents our results and section 6 offers some discussion and conclusions.

## **2. Stated preferences methods**

While a range of stated preference methods have been used in health (Bridges 2003), the applications of conjoint analysis methods have rapidly increased over the past decade (Ryan and Gerard 2003, Bridges et al. 2008, Marshall et al.

2010). Conjoint analysis relies on the theory constructed over the course of the last century from several disciplines (Thurstone 1927, Lancaster 1966, Luce and Tukey 1964, McFadden 1974), all focused on how to model our decision-making processes.

The model we use here is the multinomial logit model operationalized for economics by Daniel McFadden (McFadden 1974). Our utility for a given item, hearing aids in this case, is comprised of a deterministic part,  $V$  and a random component,  $\varepsilon$ .

$$U_{iq} = V_{iq} + \varepsilon_{iq}$$

An individual's ( $q$ ) utility for the  $i$ th alternative is a function of the observed and unobserved variations in preferences. Random utility allows for some error in our decision processes. In discrete choice experiments, the utility of one set of characteristics is more than another if the individual chooses that alternative, so that:

$$U_{iq} > U_{jq} \text{ for all } j \neq i.$$

Substituting in the deterministic and random utility components and then rearranging leave us with:

$$V_{iq} - V_{jq} > \varepsilon_{jq} - \varepsilon_{iq}.$$

The difference in observed tastes must be greater than that of the error terms, in order for the individual to select choice  $i$ .

While conjoint analysis is often used generically to cover a range of stated preference methods (Orme 2010, Bridges et al. 2011, Van Houtven et al. 2011), there is growing resistance to the use of this term by those who favor the term 'discrete choice experiments,' which relate only to choice-based methods grounded in theory

(Louviere et al. 2010). Despite this important distinction in nomenclature, most applications that describe themselves as conjoint analysis actually use a discrete choice format (Pereira et al. 2011, Marshall et al. 2010, van Til et al. 2009, Phillips et al. 2002) although other formats, such as graded pairs (Viscusi et al. 1991), profile valuation (Shumway 2003), and adaptive conjoint analysis (Fraenkel et al. 2010) are used.

In applying conjoint analysis and discrete choice methods, researchers observe patients' choices among hypothetical scenarios, and decompose this overall valuation into how they value each characteristic (Ryan and Gerard 2003, Bridges et al. 2008). Such methods can also be used to explore tradeoffs among attributes (Ryan 1999), which can be used to estimate willingness-to-pay (Vroomen and Zweifel 2011), welfare estimates (Lancsar and Savage 2004), and maximum acceptable risks (Van Houtven et al. 2011). These methods are different than simply asking individuals which technology they prefer. The experiments force patients to make tradeoffs between cost and comfort, for example, much as they would in real life.

### **3. Designing choice experiments**

Recently, a number of methodological guidelines have been produced to inform the design, execution and analysis of these studies (Bridges et al. 2010, Lancsar and Louviere 2008, Ryan and Farrar 2000, Viney et al. 2002). Even with these methodological guidelines, there is still ongoing work relating to the actual design of the studies. These studies focus on how much the design of the choice

tasks influences the estimation of preferences. Randomized studies have become increasingly popular to test the impact of study design on results (Griffith et al. 2009, Kinter and Bridges 2011, Fraenkel 2010). This field of research builds on the work of Hensher and colleagues who have been exploring the effects of designs on preferences (Caussade et al. 2005, Hensher 2006a, Hensher 2006b, Johnson 2006, Rose et al. 2009, Hensher et al. 2011).

If a choice task is unrealistic, too complicated or poorly explained, then respondents may resort to a simplified decision rule or heuristic when making selections. These shortcuts may elicit responses that are not consistent with their preferences. Respondents may only focus on parts of the information presented (Tversky and Kahneman 1974, Payne et al. 1993). One of the more common heuristics is to focus only on one characteristic of a choice at a time (Gillbride and Allenby 2004). Such respondents, often call non-traders or lexicographic decision makers, are often dropped from the analysis (Ryan 1999, Bishai et al. 2007, Ryan et al. 2009, Miguel et al. 2005). More recently, researchers have argued against dropping the lexicographic decision makers, given that trading only on one option may actually reflect an individual's preferences. Lancsar and Louviere (2006) argue deleting those with seemingly irrational preferences unnecessarily reduces sample size and reduces the generalizability of results, without necessarily violating the underlying theory. More importantly, Lancsar and Louviere claim that lexicographic responses might be as a consequence of the study design, implying that experimental design may distort the responses of some respondents, leading to an imperfect estimation of their preferences

Inefficiencies in the estimation of preferences in health might be of a direct consequence of researchers presenting respondents with only two choices at a time. Limiting the number of tasks and the number of profiles in each task has resulted from a concern about the cognitive burden associated with methods (Maddala et al. 2003, Louviere et al. 2008, Bridges et al. 2011). While there is some consensus on the number of choice tasks a respondent can answer (8-16) (Coast et al. 2006, Bridges et al. 2011), the number attributes that a task should have (5-8) (Hensher 2006a), and the number of levels attribute should have (2-4) (Pinnel and Englert 1997), there currently is little written on the number of choice profiles that should be included in each task.

The use of the paired-profile, while common in health, is not consistent with recommendations used more broadly in marketing (Sandor and Wedel 2002, Green and Srinivasan 1978). DeShazo and Fermo (2002) took on a broad study to examine multiple factors such as the number of alternatives and the number attributes for each alternative in the design of conjoint experiments. Their primary aim was to investigate the cognitive burden of these experiments. However, in the paper, they determined that 3.25 (rounded to three) alternatives in each card were the most efficient. More recently, Burgess and Street (2006) concluded that presenting respondents with more than two choices increased the accuracy of the parameter estimates.

## **4. Methods**

### *a. Survey design*



The first recommended step in any conjoint analysis is to derive the characteristics that will be used to define each choice. This is usually done through a literature search, consultation with experts and qualitative interviews of patients. This particular study is part of a larger examination into hearing aid utilization. To motivate the process, a series of qualitative interviews were conducted with patients referred to study staff by their clinicians in the Johns Hopkins Hospital, Department of Otolaryngology-Head and Neck Surgery. Data were derived via open ended, in-depth interviews.

Eighteen individuals with hearing loss were invited to participate in this study. Trained fieldworkers conducted the semi-structured interviews, and encouraged participants to discuss their feelings and experiences with hearing loss and/or hearing aids. Researchers analyzed each written interview for patient-reported features of hearing aids and hearing loss. The interviews were transcribed and searched for main themes. The major themes can be grouped under perceived performance, features, costs and the impact of the aid on the user. From these groups, seven attributes were selected: performance in noisy settings, performance in quiet settings, battery life, feedback, cost, comfort and whether the aid was water and sweat resistant. The definition of the attributes and their levels is detailed in Table 1.

[Please insert Table 1]

After the selection of the 7 attributes, each with two levels, the designs of the two arms of the study were created. For both designs, the cards were randomized so

that each had an equal probability of being asked first. The paired design is a main-effects orthogonal design, where profiles were paired with their mirror image (as there were only two levels per attribute). This involved eight cards in total, and with each respondent responding to all cards. Sawtooth software generated a D-efficient design for the triplet and respondents responded to all twelve cards. For the triplet design, respondents were also asked to select their second-best option from the remaining two.

*b. Study population*

We used the Knowledge Networks online panel. The panel is probability sampled to be nationally representative. Households are recruited through address and random digit dialing and once accepted, are provided with technology to access the internet, if needed. The sampling frame includes households with and without phones or cell-phone only houses. The organization also oversamples African American and Hispanic areas based on Census tract information. Respondents take surveys online. Knowledge Networks sends reminder emails to the non-responders, followed up with a reminder telephone call to complete the survey.

500 US participants were assessed for eligibility from the panel (Figure 1). 177 individuals were subsequently excluded: 127 did not have hearing loss, 28 had hearing loss so profound that they would not benefit from hearing aids, and 22 declined to participate. This left 323 participants to be randomized to either the paired or the triplet designs. 146 individuals were randomized to the paired and 177 were randomized to the triplet. Conjoint analysis uses data from multiple

comparisons from each participant. From the paired design, we had 1168 comparisons after two comparisons were missing from two different respondents. For the triplet design, 2112 comparisons were analyzed after subtracting the seven missing from one respondent and a respondent who only chose card C.

*c. Statistical analysis*

The data of the paired and triplet experiment were analyzed with the conditional logit model/Mc Fadden's choice model (McFadden 1974, Maddala 1983) with the selection of hearing aid option A/B/C as the dependent variable. This is coded as 1 for the chosen alternative and 0 otherwise. We used dummy coding for the attributes. All attributes were coded such that positive estimates could be expected except for the cost attribute. No intercept was included in any of the models. The parameters and standard errors of the model were estimated by maximum likelihood using clogit in STATA.

For the fully ranked data we used the rank-ordered logistic regression model, which is also known as the Plackett-Luce model (Punj and Staelin 1978) and as the exploded logit model (Ben-Akiva et al. 1985). Maximum likelihood is used to estimate the model (Marden 1995). Standard errors are estimated with the observed information matrix. Parameters were tested using Wald tests based on the maximum likelihood estimates and standard errors (calculated with the observed information matrix). Significance was set at  $p < 0.05$ .

One issue with the dataset is that there is no crossover in the paired and triplet samples. No one in the paired sample answered any of the questions in the

triplet, and vice versa. So there is some concern that our parameters may not be directly comparable. Swait and Louviere (2003) developed a procedure to test whether the two samples have different underlying parameters. In the conditional logit model the estimated parameters are confounded with the scale parameter, the inverse of the variance (Swait and Louviere 2003). Swait and Louviere set out three scenarios that could happen with comparisons of two data sets. The first is that the relative utilities are the same in both sample, but that any differences observed is due to chance. The second is where the underlying parameters are the same but the scale parameter is different. The third is that both the coefficients and the scale parameters are different. To address the issue we first tested whether the coefficients  $\beta_{\text{quiet},i}$ ,  $\beta_{\text{comfort},i}$ ,  $\beta_{\text{feedback},i}$ ,  $\beta_{\text{battery},i}$ ,  $\beta_{\text{cost},i}$ ,  $\beta_{\text{noisy},i}$ ,  $\beta_{\text{water},i}$ ,  $i=1,2$  (where  $i=1$  for paired and  $i=2$  for triplet experiment) are equal while permitting the scale factors to differ between the paired and triplet experiment:

$$H_{0,A}: \beta_{\text{quiet},1} = \beta_{\text{quiet},2}, \beta_{\text{comfort},1} = \beta_{\text{comfort},2}, \beta_{\text{feedback},1} = \beta_{\text{feedback},2}, \beta_{\text{battery},1} = \beta_{\text{battery},2}, \\ \beta_{\text{cost},1} = \beta_{\text{cost},2}, \beta_{\text{noisy},1} = \beta_{\text{noisy},2}, \beta_{\text{water},1} = \beta_{\text{water},2}.$$

If this test was not rejected, we tested whether the scale parameters  $\tau_{\text{paired}}$  and  $\tau_{\text{triplet}}$  differ between the two experiments:

$$H_{0,B}: \tau_{\text{paired}} = \tau_{\text{triplet}}.$$

Both hypotheses were tested using standard likelihood ratio statistics according to Swait and Louviere. If  $H_{0,A}$  is not rejected and  $H_{0,B}$  is rejected we conclude that the underlying parameters are the same but the scale factors are different. Moreover with testing hypothesis  $H_{0,A}$  we obtained an estimate of the relative scale factor  $\tau_{\text{triplet}}/\tau_{\text{paired}}$  between the two samples and this can be interpreted as a measure of the homogeneity of the error variances of the two samples.

Ordinary least squares was used to estimate individual-level preferences with the depended variable being the choice of card. No intercept was estimated and the attributes dummies are the independent variables. If a respondent's parameter estimate was one for a certain attribute, this indicates that the respondent was not trading any of the other attributes.

Willingness to pay for each attribute relative to baseline level was calculated by dividing the estimated attribute coefficient by the negative cost parameter estimate, multiplied by \$2000. Bootstrap (bias-corrected) confidence intervals were calculated based on 1000 replications (Hole 2007). Differences in descriptive statistics between the samples were tested with t-test for normal data, chi-squared test and Wilcoxon test was used for non-normal or heavily skewed data. Data were analyzed using STATA version 11.1 (Stata Corp, College Station, TX, USA) and R (R Development Core Team, 2010) for the individual OLS estimates.

## 5. Results

The average age of participants was 64.1 years (SD=12.7) and 32.8% were female. In Table 2, descriptive characteristics are given for the paired and triplet design separately. The two samples differ significantly in race (Fisher exact test,  $p=0.022$ ). The duration of completing the task differs significantly between the two experiments (Wilcoxon test,  $z=8.536$ ,  $p<0.0001$ ). The triplet task takes almost twice as long with a median 21.0 versus 11.5 minutes for the paired.

For the paired design, functionality in noisy settings contributes is the relatively most important characteristic influence choice of hearing aids. A one-unit increase in the aid's functioning in noisy environments means a respondent is 2.82 times as likely to purchase a hearing aid with improved functioning. Comfort and functionality in quiet environments were the next most important attributes with odds ratios of 1.75 and 1.74, respectively. All coefficients are significant at  $p<0.001$  except for battery life ( $p=0.230$ ), which is not a significant contributor to individuals' utility functions.

The triplet design produced generally more extreme results. The odds ratio for functioning in noisy environments jumps to 4.67 from 2.82 in the paired design. Quiet settings and comfort maintain their closeness, nearly tying for second place (OR=2.30 quiet, 2.32 comfort). Waterproof is the next most important attribute (1.61) and now all of the attributes including battery life are significant at  $p<0.000$ . The median duration to complete the triplet experiment was 21 minutes. This is statistically different from the median time, 11 minutes, to complete the paired experiment (Wilcoxon rank test,  $z=-8.536$ ,  $p<0.0001$ )

The fully ranked design provided the most extreme valuations. Individuals under this design are over 5 times as likely to purchase a hearing aid that works better in noisy settings. The distance between the cost and the functionality in quiet settings expands, so that comfort is now clearly more important (OR=2.48 quiet, 2.62 comfort). All coefficients are significant at  $p < 0.000$ . Additionally, the standard errors are decreasing from the paired to the triplet, and are smallest for the fully ranked design. This indicated the efficiency of the designs is improving as we ask respondents to give us more information. In Figure 2 the estimated odds ratios are shown with their confidence limits.

In Figure 5 the parameter estimates of the triplet and fully ranked experiment are plotted against the parameter estimates of the paired experiment. Moreover, with OLS we plotted a line (without intercept) through these pairs of estimates. Both for the triplet and for the fully ranked experiment the slope of the line is larger than one, 1.46 and 1.55 respectively. This straight line suggests that the paired, triplet and fully ranked coefficients differ only by a multiplicative scalar, namely the ratio between the scale parameters, and that the scale parameter of the paired experiment is smaller than the triplet or fully ranked experiment.

The estimated relative scale between the paired and triplet experiment equals 1.49. Hypothesis  $H_{0,A}$  is not rejected ( $\chi^2=3.23$ ,  $df=8$ ,  $p=0.919$ ). Hypothesis  $H_{0,B}$  is rejected ( $\chi^2=890.2$ ,  $df=1$ ,  $p < 0.0001$ ), implying that the parameters differ between the two experiments, but only up to a scaling factor. Moreover, the estimated relative scale factor 1.49 can be interpreted as a measure of heterogeneity of the error variances between the two experiments since it equals the ratio of the error

variances of the paired sample and the triplet sample. It implies that the relative efficiency of the triplet experiment over the paired experiment.

In examining decision heuristics, 32.2 % of the respondents made choices that were dominated by only one attribute in the paired design. 17.1% of the respondents made choices that were purely dominated by the functionality in noisy environment attribute. Figure 3 shows an example of how the distribution of one of the attributes changes from the paired to the full-ranked triplet design. Figure 4 shows the distribution of the individual parameters for all attributes and the two experiments.

An additional sensitivity check was performed to weight our sample with national sampling weights. After weighting, the fully ranked experiment still provided the most extreme valuations (available from the authors upon request). There is no statistically difference between the weights of the two experiments (t-value = 0.2171. p-value = 0.4141).

Willingness-to-pay (WTP) for a given attribute is an approximation of how one's welfare is improved with increases in certain attributes. This is the key way the results of conjoint experiments are translated into useful information for policymakers. Our results imply that lexicographic decision makers bias the willingness-to-pay estimates. Respondents are willing to pay \$5,392 for a hearing aid that functions better in noisy settings according to the fully ranked data, but only willing to pay \$3,618 under the paired design. Table 4 summarizes the various differences in WTP across the three designs. Researchers attempting to use the



WTP values in economic evaluations, or policy makers trying to use them to adjust fees on programs could end up with less accurate estimations in a paired design.

## **6. Discussion**

Designs incorporating three cards are better able to distinguish relative preference orderings. The paired format is the only design to have a relative ordering of the attributes different than the other two. In the paired design, comfort is rated higher than quiet settings, but not by much. The odds ratios are 1.75 (comfort) and 1.74 (quiet settings). In all of the triplet card designs, functionality in quiet settings is ranked higher than cost. We hypothesize that this is likely due to the fact that the triplet design forces respondents to trade on more than one attribute.

Figure 2 illustrates this process. This figure is created from the individual regression models. Conjoint analysis is able to run individual-level models as each individual has multiple data points—one for each choice. In Figure 2, one can see the spike in valuations at zero, and then the smaller spikes around other values for the paired design. For the triplet design, the distribution looks more normal. We think this is because the respondents in the triplet design are forced to trade on other options. With only two levels per attribute, triplet cards would have one level of each attribute appearing twice on each choice set. For those who were focused solely on one attribute, if the level of their favored attribute is the same, they have to move to another attribute in order to select a hypothetical hearing aid. Our results

support the conclusions of Lancsar and Louviere (2006) that the designs of conjoint experiments are important in encouraging seemingly irrational preferences.

Giving respondents an opt-out choice has also been widely discussed as a way to improve the accuracy for lexicographic decision makers (Haaijer et al. 2001, Dhar 1997, Brazell et al. 2006) with Brazell et al. (2006) providing some of the most recent research on how to improve those designs. In health, Ryan et al. (2004) performed a conjoint with an opt-out choice for cervical cancer screenings and concluded that it did improve the accuracy of the responses. However, we feel that our triplet design provides better utility estimation with a third option instead of a third option consisting of nothing. No relative preference information is captured when a respondent selects the opt-out option. Additionally, providing opt-out options may not provide relevant information for policy makers looking to use the conclusions of conjoint analysis to adjust benefits or design screening programs, especially if the screening program is designed for everyone within certain age ranges.

The main limitation of this study is that we were not able to do a crossover or fold-over design so that those randomized to the paired design did not answer any of the triplet conjoint profiles. This would allow us to better analyze how decision processes change when moving from the paired to the triplet design. Further research should use a fold-over design to test whether our conclusions are supported. An additional item for further research, which we are testing, is whether a paired design with levels of some attributes appearing twice on the profiles would yield results as efficient as the paired and fully ranked designs.

This experiment is the first of its kind in health care decision-making. We find that the triplet designs, and more specifically the fully-ranked triplet design provide better assessments of preferences, with more efficiency. Conjoint has become increasingly used to evaluate health care technologies and policies and knowing what designs elicit better preference estimations are important. Our study shows the dramatic changes in the willingness to pay estimates that can happen with more refined relative utility values.

## **References**

- Ben-Akiva ME, Lerman SR. Discrete Choice Analysis: Theory and application to travel demand. Cambridge, MA. MIT press. 1985.
- Bertoli S, Staehelin K, Zemp E, Schindler C, Bodmer D, Probst R. Survey on hearing aid use and satisfaction in Switzerland and their determinants. *Int J Audiol*. 2009 Apr;48(4):183-95.
- Bishai D, Brice R, Girod I, Saleh A, Ehreth J. Conjoint analysis of French and German parents' willingness to pay for meningococcal vaccine. *Pharmacoeconomics*. 2007;25(2):143-154.
- Brazell JD, Diener CG, Karniouchina E, Moore WL, Séverin V, Uldry PF. The no-choice option and dual response choice designs. *Market Lett*. 2006; 17: 255-268.
- Bridges, J. (2003) 'Stated preference methods in health care evaluation: an emerging methodological paradigm in health economics', *Applied Health Economics and Health Policy*, 2(4): 213-224.
- Bridges J, Hauber AB, Marshall, D, Lloyd, A, Prosser, L, Regier, DA, Johnson FR, Mauskopf, J. A Checklist for Conjoint Analysis Applications in Health: Report of the ISPOR Conjoint Analysis Good Research Practices Taskforce. *Value in Health*, 2011, In Press.
- Bridges J, Kinter E, Kidane L et al. Things are looking up since we started listening to patients: Recent trends in the application of conjoint analysis in health 1970-2007. *Patient*. 2008; 1(4): 273-282.
- Bridges J, Kinter E, Schmeding, A, Rudolph I and Mühlbacher I. Can patients with schizophrenia complete choice-based conjoint analysis tasks? *The Patient*, 2011, 4(4): In Press.
- Bridges, J, Cohen, JP, Grist PG, Mühlbacher AC. International experience with comparative effectiveness research: case studies from England/Wales and Germany. *Advances in Health Economics and Health Services Research*. 2010;22: 29-50.
- Burgess, Leonie and Street, Deborah J. The optimal size of choice sets in choice experiments. *Statistics*. 2006; 40: 6, 507-515.
- Caussade S, Ortúzar J, Rizzi L, Hensher DA. Assessing the Influence of Design Dimensions on Stated Choice Experiment Estimates. *Transportation Research*. 2005; 39 (7), 621-640.
- Coast J, Flynn TN, Salisbury C, Louviere J, Peters TJ. Maximising responses to discrete choice experiments: a randomised trial. *Applied Health Economics and Health Policy*. 2006; 5(4): 249-260.
- DeShazo JR, Fermo G. Designing Choice Sets for Stated Preference Methods: The effects of Complexity on Choice Consistency. *Journal of Environmental Economics and Management*. 2002; 44:123-143.

- Dhar R. Consumer Preference for a No-Choice Option. *The Journal of Consumer Research*. 1997 Sep;24(2):215-231.
- Facey K, Boivin A, Gracia J, Hansen HP, Lo Scalzo A, Mossman J, Single A. Patients' perspectives in health technology assessment: a route to robust evidence and fair deliberation. *Int J Technol Assess Health Care*. 2010; 26(3):334-40.
- Fitzpatrick EM, Leblanc S. Exploring the factors influencing discontinued hearing aid use in patients with unilateral cochlear implants. *Trends Amplif*. 2010 Dec;14(4):199-210.
- Fraenkel L, Chodkowski D, Lim J, Garcia-Tsao G. (2010) Patients' preferences for treatment of hepatitis C. *Med Decis Making*. 2010 Jan-Feb;30(1):45-57.
- Fraenkel L. Feasibility of Using Modified Adaptive Conjoint Analysis Importance Questions. *Patient*. 2010;3(4):209-215.
- Franks JR, Beckmann NJ. Rejection of hearing aids: attitudes of a geriatric sample. *Ear Hear*. 1985 Jun;6(3):161-6.
- Gilbride TJ, Allenby GM. "A choice model with conjunctive, disjunctive and compensatory screening rules." *Marketing Science*. 2004 23(3):391-406.
- Green PE, Srinivasan V. Conjoint Analysis in Consumer Research: Issues and Outlook. *Journal of Consumer Research* 1978; 5(2): 103-23.
- Griffith JM, Lewis CL, Hawley S, Sheridan SL, Pignone MP. Randomized trial of presenting absolute v. relative risk reduction in the elicitation of patient values for heart disease prevention with conjoint analysis, *Med Decis Making*. 2009;29:167:174.
- Haaier R, Kamakura W, Wedel M. The 'no-choice' alternative in conjoint experiments. *International Journal of Market Research*. 2001; 43(1): 93-106.
- Hensher D, Jou RC, Rose JM, Li Z, Huang GL. A comparative investigation of the effects of the design dimensions of choice experiments on car commuters' route choice behaviour and valuation of time in Taiwan and Australia, *International Journal of Transport Economics*, 2011; XXXVIII (2), 147-172.
- Hensher DA. How do Respondents Process Stated Choice Experiments? – Attribute consideration under varying information load, *Journal of Applied Econometrics*. 2006a; 21 , 861-878
- Hensher DA. Revealing differences in behavioural response due to the dimensionality of stated choice designs: an initial assessment, *Environmental and Resource Economics*. 2006b; 34 (1): 7-44.
- Hole AR. A comparison of approaches to estimating confidence intervals for willingness to pay measures. *Health Economics*. 2007; 16: 827-840.
- IQWiG. Methodik fuer die Bewertung von Verhaeltnissen zwischen Nutzen und Kosten im System der deutschen gesetzlichen Krankenversicherung. Version 1.1 vom 09.10.2008. Koeln: Institut fuer Qualitaet und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), 2008.

Johnson FR. Comment on revealing differences in willingness to pay due to the dimensionality of stated choice designs: An initial assessment. *Environ Resource Econ.* 2006;34:45-50.

Kinter E, Bridges J. (2011) A comparison of two methods for experimental design to value patient relevant outcomes using conjoint analysis, Draft.

Lancaster K. A new approach to Consumer Theory. *Journal of Political Economy.* 1966; 74(2): 132-157.

Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision-making: a user's guide. *Pharmacoeconomics* 2008; 26(8): 661-677.

Lancsar E, Louviere J. Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? *Health Econ.* 2006 Aug;15(8):797-811

Lancsar E, Savage E. Deriving welfare measures from discrete choice experiments: inconsistency between current methods and random utility and welfare theory. *Health Economics.* 2004 Sep 1;13(9):901-907.

Louviere J, Islam T, Wasi N, Street D, Burgess, L (2008) Designing Discrete Choice Experiments: Do Optimal Designs Come at a Price? *Journal of Consumer Research*, 35: 360-375

Louviere JJ, Flynn TN, Carson RT. Discrete choice experiments are not conjoint analysis. *Journal of Choice Modeling.* 2010;3(3):57-72.

Luce RD, Tukey JW. Simultaneous Conjoint Measurement: A new Type of Fundamental Measurement. *Journal of Mathematical Psychology.* 1964; I, 1-27.

Maddala GS. Limited dependent and qualitative variables in econometrics, J. Grandmont, Editor. 1983, Cambridge University Press.

Maddala, T, Phillips CA. Johnson FR. An experiment on simplifying conjoint analysis designs for measuring preferences. *Health Economics.* 2003; 12: 1035 – 1047

Marden JI. *Analyzing and Modeling Rank Data.* London: Chapman & Hall, 1995.

Marshall D, Bridges J, Hauber AB et al. Conjoint Analysis Applications in Health - How are studies being designed and reported? An update on current practice in the published literature between 2005 and 2008. *Patient* 2010; 3(4): 249-256.

McFadden D. Conditional logit analysis of qualitative choice behavior, in *Frontiers in Econometrics*, P. Zarembka, Editor. Academic Press: New York. p. 105-142; 1974.

Medicare.gov. Your Medicare Coverage. Online: <http://www.medicare.gov/coverage/Search/Results.asp?State=MD|Maryland&Coverage=34|Hearing+Exams+and+Hearing+Aids&submitState=View+Results+%3E>. Accessed: 16 September 2011.

Miguel FS, Ryan M, Amaya-Amaya M. 'Irrational' stated preferences: a quantitative and qualitative investigation. *Health Econ.* 2005 Mar;14(3):307-322.

- NICHD 2011 Quick Statistics. National Institute on Deafness and other Communication Disorders Online: <http://www.nidcd.nih.gov/health/statistics/quick.htm>. Accessed 16 September 2011.
- Orme B. Getting Started with Conjoint Analysis: Strategies for Product Design and Product Research (2nd Edition), Research Publishers: Madison, WI; 2010.
- Payne JW, Bettman JR, Johnson EJ. The adaptive decision maker. Cambridge University Press; 1993.
- Pereira C, Mulligan M, Bridges J, Bishai D. Determinants of influenza vaccine purchasing decision in the US: a conjoint analysis. *Vaccine*. 2011; 29(7): 1443-7.
- Phillips KA, Maddala T, Johnson FR. Measuring preferences for health care interventions using conjoint analysis: an application to HIV testing. *Health Services Research*, 2002. 37:1681-705.
- Pinnel J, Englert S. The number of choice alternatives in discrete choice modeling. *Sawtooth Software Conference Proceedings: Sequim, WA*, 1997.
- Punj GN, Staelin R. The choice process for graduate business schools. *Journal of Marketing Research*. 1978; 15: 588-598
- R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>; 2010.
- Rose J, Hensher DA, Caussade S, Ortuzar, DeDios J, Rong-Chang J. Identifying differences in preferences due to dimensionality in stated choice experiments: a cross cultural analysis, *Journal of Transport Geography*. 2009; 17 (1), 21-29
- Ryan M, Farrar S. Using Conjoint analysis to elicit preferences for health care. *BMJ*, 2000; 320: 1530-1533.
- Ryan M, Gerard K. Using discrete choice experiments to value health care programmes: current practice and future research reflections. *Appl Health Econ Health Policy* 2003; 2: 55-64.
- Ryan M, Skåtun D. Modelling non-demanders in choice experiments. *Health Econ*. 2004 Apr; 13(4):397-402.
- Ryan M, Watson V, Entwistle V. Rationalising the 'irrational': a think aloud study of discrete choice experiment responses. *Health Econ*. 2009 Mar;18(3):321-336.
- Ryan M. Using conjoint analysis to take account of patient preferences and go beyond health outcomes: an application to in vitro fertilisation. *Social Science & Medicine*. 1999 Feb;48(4):535-546.
- Ryan M.(1999) A role for conjoint analysis in technology assessment in health care? *Int J Technol Assess Health Care*. 1999 Summer;15(3):443-57.
- Sandor Z, Wedel M. Profile construction in experimental choice designs for mixed logit models. *Marketing science*. 2002 Fall; 21(4):455-475.

- Shumway M. Preference weights for cost-outcome analyses of schizophrenia treatments: comparison of four stakeholder groups. *Schizophrenia Bulletin*. 2003; 29(2): 257-66.
- StataCorp. *Stata Statistical Software: Release 11*. College Station, TX: StataCorp LP; 2009.
- Swait J, Louviere J. The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of Marketing Research*. 2003;30:305-314.
- Thurstone LL. Law of comparative judgement. *Psychol. Rev.* 1927; **34**: 273–286.
- Trowman R, Chung H, Longson C, Littlejohns P, Clark P. The National Institute for Health and Clinical Excellence and Its Role in Assessing the Value of New Cancer Treatments in England and Wales. *Clinical Cancer Research*. 2011;17(15):4930–5.
- Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. *Science*. 1974, 185: 1124-1131.
- Van Houtven G, Johnson FR, Kilambi V, Hauber AB. Eliciting Benefit–Risk Preferences and Probability-Weighted Utility Using Choice-Format Conjoint Analysis. *Medical Decision Making*. 2011 May;31(3):469 -480.
- van Til JA, Stiggelbout AM, Ijzerman MJ. The effect of information on preferences stated in a choice-based conjoint analysis. *Patient Educ Couns*. 2009 Feb;74(2):264-71. Epub 2008 Oct 26.
- Viney R., Lancsar E., Louviere J. Discrete choice experiments to measure consumer preferences for health and healthcare. *Expert Rev Pharmacoeconomics Outcomes Res* 2002;2:89-96.
- Viscusi WK, Magat WA, Huber J. Pricing Environmental Health Risks: Survey Assessments of Risk-Risk and Risk-Dollar Trade-offs for Chronic Bronchitis. *Journal of Environmental Economics and Management*, 1991; 21:32–51.
- Vroomen JM, Zweifel P. Preferences for health insurance and health status: does it matter whether you are Dutch or German? *Eur J Health Econ*. 2011 Feb;12(1):87–95.



**TABLE 1: ATTRIBUTES AND THEIR LEVELS.**

<b>Attribute</b>	<b>Definition</b>	<b>Levels</b>	
Battery Changes	How often the aid's batteries need to be changed.	2 times a month	4 times per month
Water and Sweat Resistance	The hearing aid's capacity to withstand moisture from the ear and/or from the environment.	Somewhat water/sweat resistance	Not so water/sweat resistance
Quiet Settings	Situations where there is only one source of sound, such as in one-on-one conversations	More effective for quiet settings	Somewhat effective for quite settings
Feedback Occurrence	The high-pitched squealing noise that a hearing aid can make	Feedback occurs 2 times a month	Feedback occurs 4 times a month
Cost	The amount of money patient spends when buying the hearing aid	\$3,000	\$5,000
Noisy	Situations where there are multiple sounds coming from multiple sources	More effective for noisy settings	Somewhat effective for noisy settings
Physical Comfort	How the hearing aid feels in the ear.	Rarely uncomfortable	Occasionally uncomfortable

**TABLE 2: DEMOGRAPHIC CHARACTERISTICS**

		<b>Paired Experiment (n=146)</b>	<b>Triplet Experiment (n=177)</b>	<b>P-values</b>
Age, years (mean, SD)		62.9(13.3)	65.1 (12.2)	0.13 (t=-1.52)
Sex	Female	96	121	0.62
	Male	50	56	( $\chi^2=0.25$ )
Education	Less than high school	6	21	0.07
	High school	56	56	( $\chi^2=7.1$ )
	Some college	39	50	
	Bachelor's degree or higher	45	50	
Income	0-\$24999	31	32	0.52
	\$25000-\$49999	34	54	( $\chi^2=2.3$ )
	\$50000-\$99999	58	67	
	\$100000 or more	23	24	
Region	Northeast	16	34	0.24
	Midwest	44	48	( $\chi^2=4.2$ )
	South	51	57	
	West	35	38	
Race	White, non-Hispanic	113	160	0.024
	Black, non-Hispanic	6	3	( $\chi^2=11.2$ )
	Other, non-Hispanic	5	1	
	Hispanic	15	9	
	2+ races, non-Hispanic	7	4	
Duration (median, range)	Minutes	11.5 (0-9958)	21 (7-5185)	p<0.000 (z=- 8.536)

**TABLE 3: REGRESSION RESULTS**

<b>Profile Design</b>						
<b>Attribute</b>	<b>Paired</b>		<b>Triplet</b>		<b>Fully Ranked</b>	
	Coefficients	Odds Ratio	Coefficients	Odds Ratio	Coefficient	Odds Ratio
Quiet Settings	0.55*** (0.08)	1.74	0.83*** (0.06)	2.30	0.91*** (0.05)	2.48
Comfort	0.56*** (0.08)	1.75	0.84*** (0.06)	2.32	0.96*** (0.05)	2.62
Feedback	0.29*** (0.07)	1.33	0.29*** (0.07)	1.34	0.44*** (0.05)	1.55
Battery Life	0.09 (0.07)	1.09	0.20** (0.06)	1.22	0.22*** (0.05)	1.24
Cost	-0.57*** (0.08)	0.56	-0.73*** (0.07)	0.48	-0.61*** (0.05)	0.54
Waterproof	0.23** (0.07)	1.25	0.47*** (0.06)	1.61	0.45*** (0.04)	1.57
Noisy Settings	1.04*** (0.08)	2.82	1.54*** (0.07)	4.67	1.64*** (0.05)	5.19

Notes: Coefficients are significant at \*p<0.05, \*\*p<0.01, \*\*\*p<0.001.

**TABLE 4: WILLINGNESS TO PAY**

	Paired	Triplet	Fully Ranked
Quiet Settings	\$1,924.90 (\$1,223.3 – \$3,181.4)	\$2,285.30 (\$1,818.5 – \$3,134.7)	\$2,978.30 (\$2,483.1 – \$3,964.6)
Comfort	\$1,946.60 (\$1,254.2 – \$3,339.2)	\$2,305.20 (\$1,845.4 – \$3,102.7)	\$3,155.00 (\$2,723.3 – \$4,166.6)
Feedback	\$996.20 (\$213.5 – \$1,915.1)	\$799.00 (\$459.3 – \$1,591.8)	\$1,426.80 (\$1,161.5 – \$1,973.4)
Battery Life	\$309.10 (-\$493.0 – \$1,004.1)	\$543.60 (\$155.2 – \$1,138.9)	\$709.70 (\$365.4 – \$1,324.0)
Waterproof	\$788.90 (\$46.1 – \$1,627.7)	\$1,301.10 (\$978.0 – \$1,785.3)	\$1,471.50 (\$1,157.6 – \$2,021.3)
Noisy Settings	\$3,618.20 (\$2,527.4 – \$5,278.0)	\$4,227.50 (\$3,391.2 – \$5,616.1)	\$5,391.50 (\$4,589.2 – \$7,177.3)

Note: The confidence intervals estimated using a bias-corrected bootstrap method (see A. Hole, 2007).

**Figure 1: Example Card**









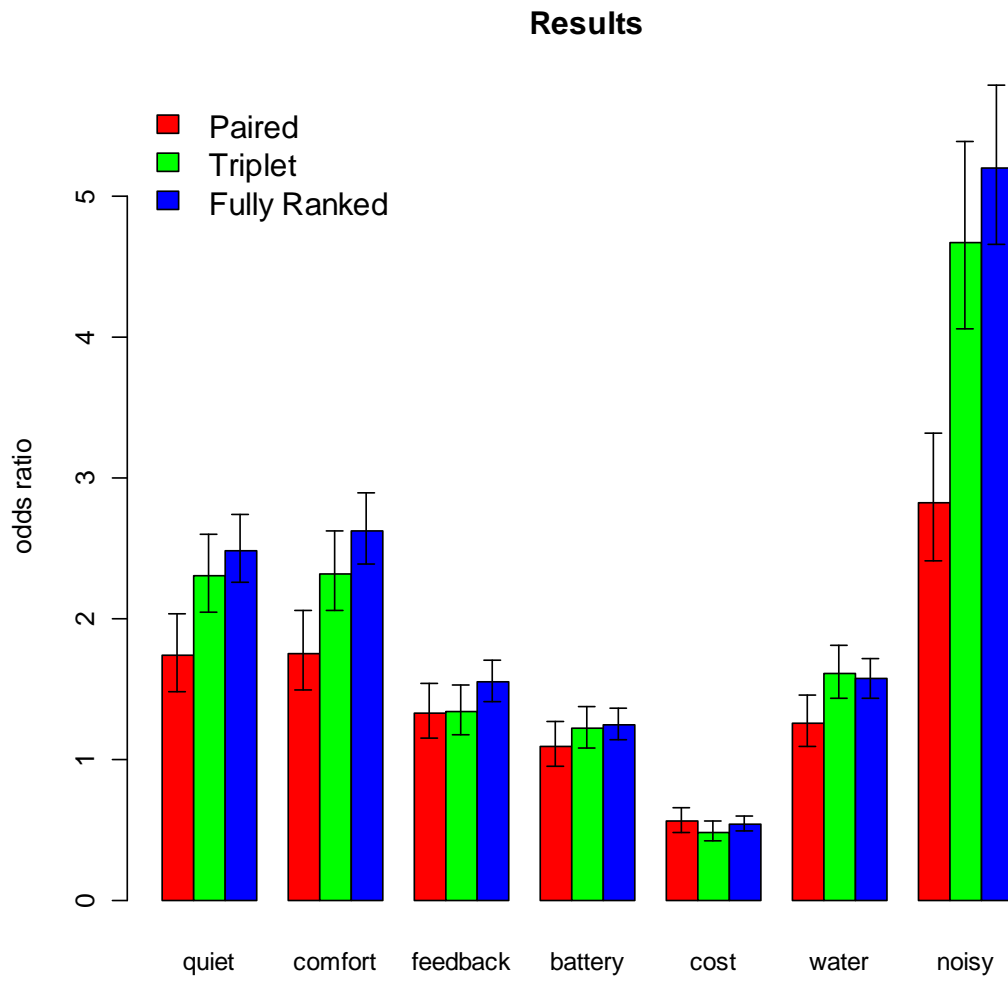
Hearing Aid Feature	Hearing Aid A	Hearing Aid B
<b>Battery Changes</b>	Four (4) times a month	Two (2) times a month
<b>Water and sweat resistance</b>	 <p>somewhat water / sweat resistant</p>	 <p>not so water / sweat resistant</p>
<b>Quiet settings</b>	 <p>somewhat effective for quiet settings</p>	 <p>more effective for quiet settings</p>
<b>Feedback Occurrence</b>	Feedback occurs four (4) times a month	Feedback occurs two (2) times a month
<b>Purchase Cost</b>	\$ 5,000	\$ 3,000
<b>Noisy settings</b>	 <p>somewhat effective for noisy settings</p>	 <p>more effective for noisy settings</p>
<b>Physical comfort</b>	 <p>occasionally uncomfortable</p>	 <p>rarely uncomfortable</p>

Figure 2: Main Results



**Figure 3: Example of the Change in Distribution of One Attribute**

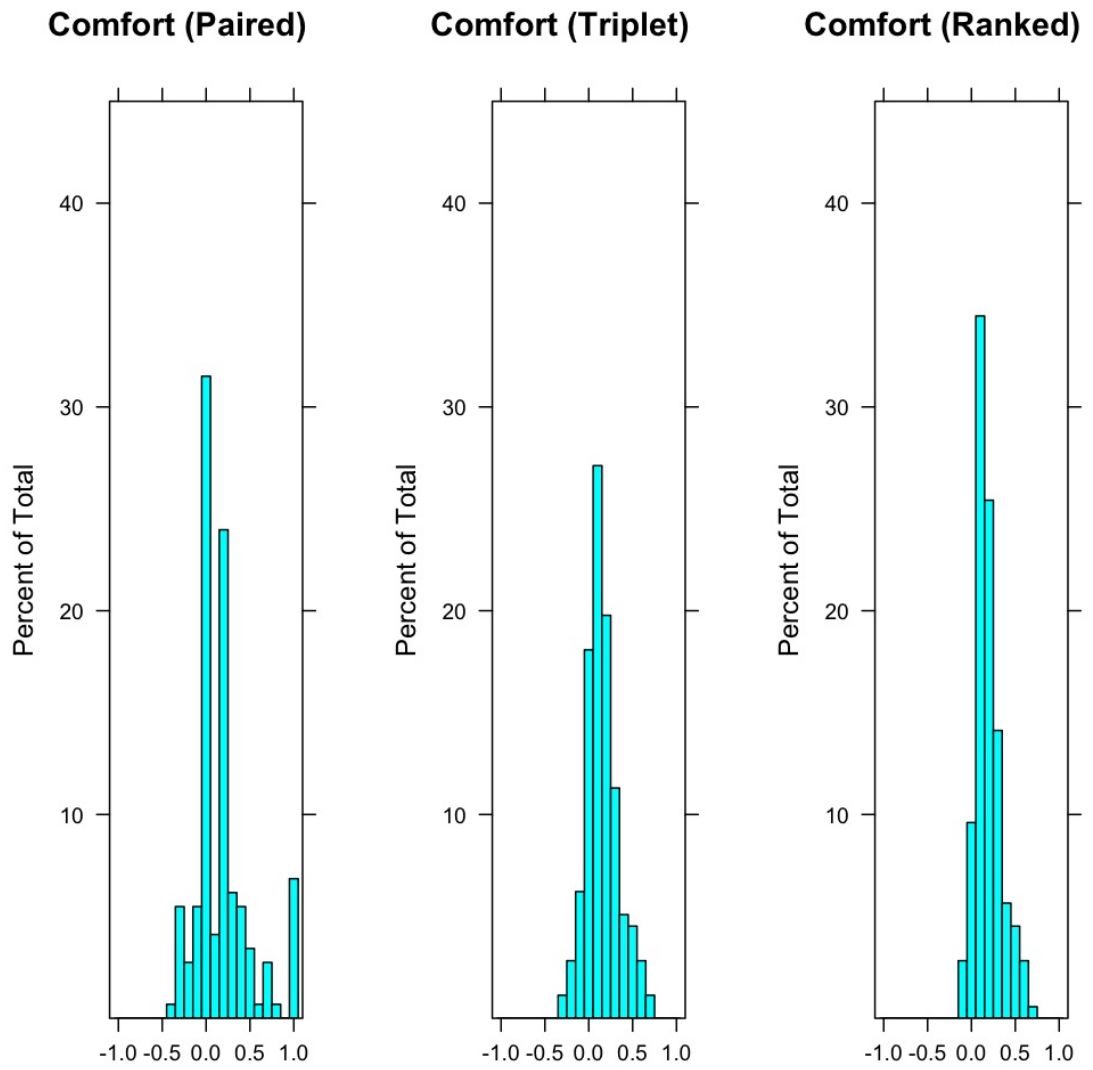


Figure 4: Comparison of Distribution, All Parameters

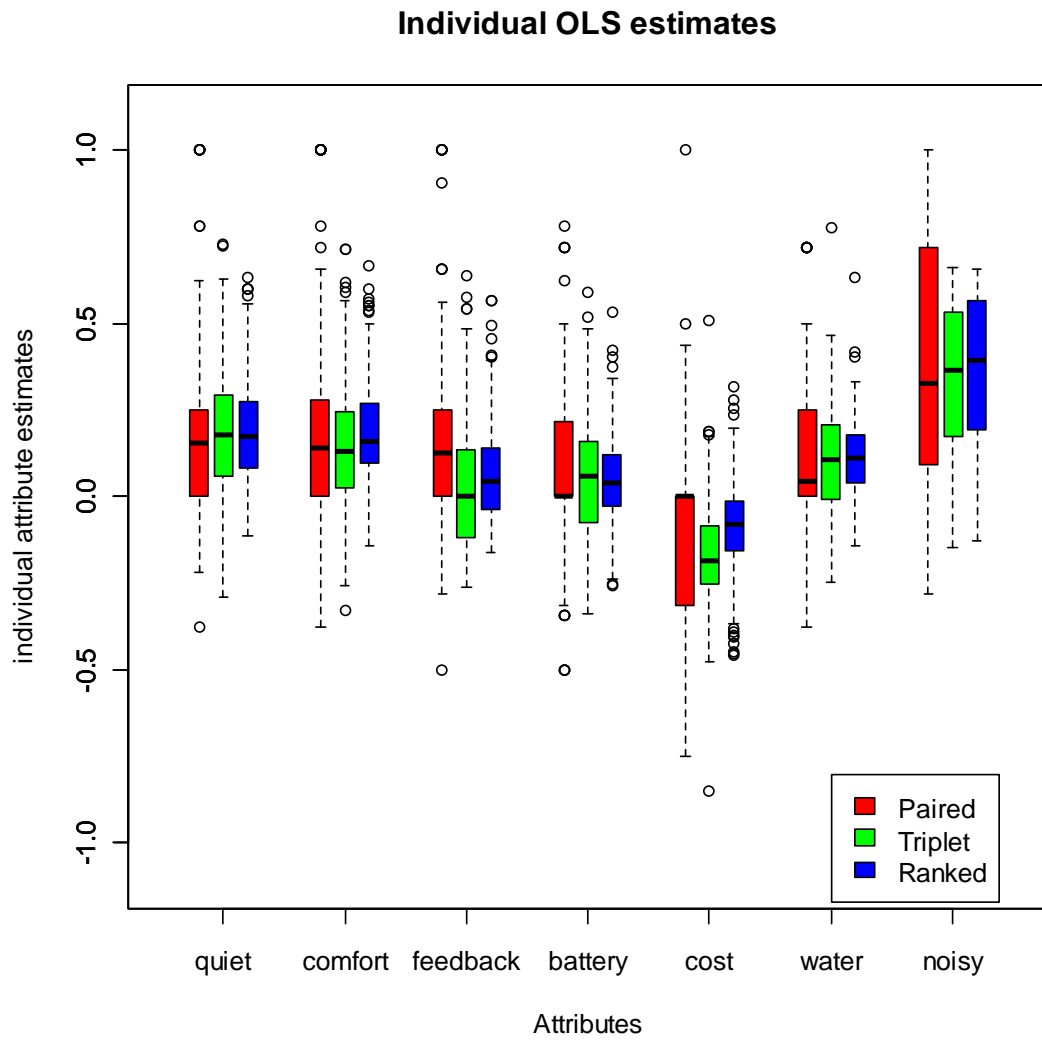




Figure 5: Comparison of parameter estimates.

