Journal of Choice Modelling, 2(2), pp 200-215
www.jocm.org.uk

Using an Ex-ante Entreaty to Reduce Protest Zero Bias in Stated Preference Surveys – A Health Economic Case

Ole Bonnichsen[1,*] Jacob Ladenburg[2,†]

[1]Environmental Economics and Rural Development Division, Institute of Food and Resource Economics, Faculty of Life Sciences, University of Copenhagen, Rolighedsvej 25, 1958 Frederiksberg C, Denmark
[2]Danish Institute of Governmental Research, Nyropsgade 37, 1602 Copenhagen V, Denmark

Received 12 December 2008, revised version received 13 May 2009, accepted 4 November 2009

Abstract

When prompting respondents to act on the hypothetical markets that are presented to them in stated preference surveys, it is often observed that a proportion of the sample state a zero demand for the good in question even though their genuine demand is positive. Though the literature comes up with some means to calibrate the estimated preferences ex-post, this is far from being unproblematic and uncontroversial. In this paper we test whether a short ex-ante entreaty, presented to respondents in a health economic stated preference survey, can effectively reduce the number of protest zero bids. We first of all find that the protest reduction entreaty significantly reduces the number of protest zero bids in the survey. Furthermore, the results indicate that marginal demand for the individual attributes as well as general preferences do not seem to be influenced by the protest reduction entreaty. Compared to the traditional ex-post calibration of protest zero bids in stated preference surveys, these results suggest that an ex-ante reduction in protest zero bids by using a simple and short entreaty is possible without introducing other types of biases in the survey.

Keywords: Choice experiment, health economics, ex-ante protest reduction entreaty, zero bids, bias

* Corresponding author, T: +45-35336815, F: +45 35336802, ole@foi.dk
† T: +45-43333461, F: +45 43333401, jal@akf.dk

Licensed under a Creative Commons Attribution-Non-Commercial 2.0 UK: England & Wales License
http://creativecommons.org/licenses/by-nc/2.0/uk/

ISSN 1755-5345

1 Introduction

In an impact assessment of a new public programme that will affect the supply and demand for a good, monetary figures are fundamental. Figures representing the cost side are generally straightforward to assess. However, public programmes in general, and explicitly in the health sector, often have an impact on non-market goods and services, for which it is typically not possible to derive complementary market good and health good relations. Accordingly, the relationship between people's actual behaviour in a market and the price/qualities of the good in question would not be sufficient for inferring the economic value of the benefits of the non-market goods and policy. One solution to this problem is the application of Stated Preference Methods (SPM), such as the Contingent Valuation Method (CVM) and Choice Experiments (CE). In the application of these methods people are asked to make choices on a well defined hypothetical market for health care targeting the good in question, thus stating their preferences in terms of hypothetical Willingness To Pay (WTP) for the benefits. However, prompting respondents to act on a hypothetical market can be problematic.

SPM are prone to a range of biases, which drive a wedge between the stated behaviour and the real market behaviour. It is often observed in SPM studies that a relatively large proportion of the sample state a zero demand for the good in question even though their genuine demand is positive, also known as protest zero[1] (*PZ*) bidders. Potentially, this is a large problem in health economic SPM surveys. Many developed countries have a collectively funded health care system, which supplies health services at zero or relatively low costs. Accordingly the inclusion of an individual payment via the costs attribute of the health related good could result in respondents considering some scenarios to be unrealistic and immoral, resulting in a high number of *PZ* bids. This is observed in Bryan *et al.* (1998), where it was found that the inclusion of a cost attribute was met with resistance.

In the literature, *PZ* bids have been dealt with using ex-post approaches, such as excluding the respondents from the analysis or treating the stated zero bids as genuine. These are far from being unproblematic approaches. "Trimming" the sample by excluding *PZ* bidders could bias the estimates of demand, if the removed *PZ* bidders have significantly different preferences compared to the respondents in the effective sample. Results may also be biased if *PZ* bidders, who may hold non-zero preferences, are included and treated as genuine zero bids. However, based on the experience of using different types of ex-ante script and entreaties to reduce hypothetical biases (Cummings and Taylor 1999 and Aadland and Caplan 2003), the present paper applies a novel approach to reduce the level of *PZ* bidding. More specifically we test a short entreaty called a Protest Reduction Entreaty (*PRE*). To the authors' knowledge, such an entreaty has previously been tested only by Morse-Jones *et al*. (2007) and Mourato *et al*. (2008) in two different contexts. The former studies the preferences for tropical biodiversity protection among people living in London with a CVM study. Their entreaty focused on the potential disbelief in the validity of the scenario and was found to effectively reduce *PZ* bids. In the latter study a positive effect of an entreaty is found in the policy context of future reductions in the risks of water use restrictions.

The present paper extends the work in several dimensions. First of all, to keep the scenario description as short as possible and minimise cognitive burden, we apply and test a shorter entreaty compared to the entreaties applied in Morse-Jones *et al*. (2007)

[1] *PZ* falls in the category of protest biases (Mitchell and Carson, 1989).

and Mourato *et al*. (2008)[2]. Secondly we test the entreaty in a different setting, using a private health good (improvements of ostomy pouches) as opposed to the purely public good (existence value associated with tropical biodiversity) in Morse-Jones *et al*. (2007) and the private environmental goods (scarcity of water) in Mourato *et al*. (2008). Thirdly, we use CE rather than CVM to elicit the preferences for improvements of the ostomy pouch attributes. Additionally we apply a purely ex-ante approach by presenting all respondents to the *PRE* via the scenario description and not only to the respondents who initially state a zero WTP, as in Mourato *et al*. (2008)[3]. Finally, we focus the entreaty on the potential conflict relating to the present payment vehicle (complete or partial reimbursement) and the scenario payment vehicle (personal payment). With this specific setup, we find that the *PRE* effectively reduces the number of *PZ* bids in the survey. Furthermore, the results indicate that the entreaty has not influenced general preferences.

The paper is structured as follows: The next section presents the study design, which is followed by the procedure and hypothesis, results, discussion and a conclusion giving a summary of the main findings.

2 Study Design

The overall frame of the present study was the elicitation of ostomates' preferences for improvements in ostomy pouches. In ostomy surgery a part of the intestine is brought through the abdominal wall creating an opening where stool is then passed. An ostomy surgery is life saving and the modern stoma management appliances give ostomates the possibility to live close to full lives (Cataldo 1999). One such appliance is an ostomy pouch. Pouches are made of soft plastic, clear or skin-coloured, and they lie flat against the skin. Pouches vary according to a number of attributes, the most important of which relate to the adhesive, filter and flexibility of the system.

Collection of data was carried out through a mailed survey consisting of 1,200 questionnaires. Respondents were sampled from a nationwide group of Swedish ostomates consisting of approximately 20,000 people. In the questionnaire, respondents were presented with a scenario description which described different types of improvements to an ostomy pouch. In accordance with Lancaster's *attribute theory of value* (Lancaster 1998), the scenario introduced improvements to the current ostomy pouch with regard to three different attributes of the ostomy pouch: The flexibility of the system as a whole, the number of small starting leakages under the base plate per month and the filter lifetime (Bonnichsen 2009). An additional monthly expense was used as the payment vehicle[4]. The attributes were presented to the respondents with the descriptions shown in Table 1.

[2] Our entreaty is of 93 words, compared to 192 and 271 words in Morse-Jones *et al*. (2007) and Mourato *et al*. (2008) respectively.

[3] In Mourato *et al*. (2008) the entreaty used to reduce *PZ* bids was only presented to the respondents who initially stated a zero WTP and subsequently were classified *PZ* bidders. After the presentation of the entreaty, the respondents were asked to reconsider their stated zero preferences in a second WTP question.

[4] Ostomates had a maximum out of pocket expense of 1,800 SEK/year when the survey was conducted.

Table 1 Attributes and Attribute Levels

Attribute	Attribute Level
Flexibility of the system as a whole	Same as current Small improvement Large improvement
Number of small starting leakages under the base plate per month	3 leakages 1 leakage No leakages
Filter lifetime	7 hours 12 hours 24 hours
Additional expense per month	(0 SEK) 125 SEK 200 SEK 375 SEK 500 SEK 750 SEK 1000 SEK

A zero-priced status quo alternative was used as a benchmark. Following Banzhaf *et al*. (2001), the status quo alternative was defined by the current system of the individual[5]. This information was then used to define the status quo in estimation.

As a full factorial design comprised 162 alternatives, a D-optimal fractional factorial design consisting of 18 alternatives was identified (Louviere *et al*. 2000)[6]. These alternatives were then arranged into nine choice sets and assigned into two blocks[7], with the respondents evaluating five and four choice sets per block. Each treatment group had identical blocking structures so that the same choice sets appeared in both treatments. As such, a choice set consisted of three alternatives: The zero-priced status quo alternative and two policy generated improvement alternatives with an associated additional monthly expense. Figure 1 shows an example of a choice set used in the questionnaire.

The questionnaire used in the survey underwent numerous revisions following on from focus groups and a pilot study. Reminders and possible replacement questionnaires were mailed to respondents approximately ten days after the initial send out.

[5] In order to find the respondents' own status quo values, the questionnaire included a question asking the respondents to state the attribute levels of their current system with regard to flexibility, number of small leakages and filter lifetime.

[6] To minimise the number of dominating and non causal alternatives, the initially identified efficient design was subjected to the manual swapping procedure suggested by Huber and Zwerina (1996).

[7] In SAS, the %mktblock macro was used to assign the choice sets into blocks (Kuhfeld 2005).

	Alternative 1	**Alternative 2**	**My current system (i.e. no change)**
Flexibility of the system as a whole (base plate and coupling)	Large improvement	Same as current	-
Number of small starting leakages under the base plate per month	3 leakages	No leakages	-
Filter lifetime	24 hours	12 hours	-
Additional expense per month	750 SEK	200 SEK	0 SEK
I prefer	❑	❑	❑
	(mark one box only)		

Figure 1 Choice Set Example

2.1 Identifying Protest Zero Bidders

There exists no well established or generally agreed upon method for identifying protest bidders and it is typically done using an ad-hoc approach (Jorgensen *et al*. 1999; Boyle and Bergstrom 1999). In our survey, the *PZ* bidders were identified using a number of debriefing questions presented to the respondents, who in the choice sets always chose their status quo and thereby expressed a zero WTP for improved ostomy pouches. The debriefing questions are presented in Table 2.

Table 2 Definition of Protest and Genuine Zero Bids

Stated reason for choosing the status quo alternative in all choice questions	Protest/ Genuine
"I cannot afford to pay the extra expense"	G
"The changes were too small to be of importance"	G
"None of the alternatives were better than my current system"	G
"I am happy with no change"	G
"I am not interested in this matter"	P
"I object to paying for ostomy pouches"	P
"Coloplast should pay for this change"	P
"The government should pay for this change"	P
"I need more information/time to answer the question"	P

The type of response to the debriefing questions is categorised as being either a genuine zero WTP (G) or a *PZ* WTP (P) [8]. This is the typical method used to identify *PZ* bidders (Meyerhoff and Liebe 2006).

2.2 Procedure

In order to test for the effect of the *PRE*, a two-split sample design was utilised. In both treatments respondents were introduced to a hypothetical market entailing choices between alternative ostomy pouches. Included in this section was a description of the attributes, reasons for the variations in the attribute levels, "cheap talk" focusing on the issue of hypothetical bias and budget reminders prior to the choice session. The questionnaires for both treatments were kept identical with the exception that respondents in treatment B were given a *PRE*, while respondents in treatment A were not. The *PRE* is presented below.

"*The hypothetical expense presented to you in the next questions is not an expense that you yourself will be responsible for paying, but it has been included merely to increase the realism of your choices and for you to convey how much you believe an improvement is worth. While this hypothetical expense will have no effect on the reimbursements received in Sweden and will not result in any extra cost for ostomates, we kindly ask you to carefully consider the hypothetical expense, as if you were to pay it when making your choices.*"

In the literature, the behaviour of *P*Z bidders appears to be triggered by different protest elements, such as dissension over the use of specific payment vehicles, the policy context, ethical objections, misunderstandings or lack of information, the institutional setting of the survey and the type of good (Boyle 2003; Strazzera *et al*. 2003; Jorgensen *et al*. 1999; Meyerhoff and Liebe 2006). Accordingly, there seems to be a link between the propensity to state a *PZ* demand and the protest elements. The *PZ* behaviour thereby becomes an increasing function of the protest elements in the survey, i.e. the more elements that can trigger a protest response, the higher the propensity to protest[9].

Many of the above mentioned protest elements can be accredited the formulation and the setup of the survey. We therefore argue that the *PRE* can redirect the attention of the respondents from potential "triggering" formulations in the scenario description, to the individual assessment of the actual values that the respondents associate with the hypothetical changes in the attributes of the presented ostomy pouches. As an intended consequence, *PZ* behaviour should be reduced by using the *PRE*.

Based on these arguments the following hypotheses are put forward.

[8] Stating "I'm happy with no change" and "The changes were too small to be of importance" are perceived as economically rational reasons for zero WTP behaviour and therefore considered genuine answers. Seen from an ostomate's point of view, changing ostomy pouch system might be associated with uncertainty related to the quality and reliability of the new system. Stating "I am not interested in this matter" could be perceived not as a protest response, but merely a way of indicating satisfaction with the status quo. As it cannot be determined whether the true motivation behind this response is of protest or genuine nature, these respondents are removed. It should also be noted that only four respondents from treatment A and three from treatment B gave this response to the debriefing question.

[9] This is in equivalence with the literature. In a study analysing the relation between type of good in question and *PZ* behaviour, Jorgensen *et al.* (2001) find that the propensity to exhibit *PZ* behaviour is both a function of the type of good in question and the scope of provision of the good.

H1: Propensity to state a PZ preference is independent of the treatment (PRE) and thus identical in treatment A and B.

If the hypothesis is rejected, an effect of the *PRE* is established. This makes two further hypotheses worth exploring. The next hypothesis relates to the direction of the propensity to state a *PZ* preference.

H2: The PRE reduces the propensity to state a PZ preference.

The acceptance of H2 would establish the intended effect of the *PRE*, i.e. that *PRE* reduces the number of *PZ* bids in the survey.

In contrast to the previously applied entreaties focusing on reducing WTP and thereby hypothetical bias (Carlsson *et al*. 2005; Cummings and Taylor 1999; List 2001), the *PRE* is not intended to influence preferences as such, but merely the propensity to exhibit *PZ* behaviour. We therefore present the final hypothesis.

H3: The PRE does not influence preferences, i.e. $Preferences_A = Preferences_B$.

3 Results

Of the 1,200 questionnaires mailed to the respondents, an initial sample (including *PZ* bids) of 465 responses was obtained[10]. This was subsequently trimmed to an effective sample (excluding *PZ* bids) of 254 usable responses. The distribution of effective responses between the two treatments was found to be similar with treatment A containing 116 responses and treatment B containing 138 responses, while the distribution of responses across blocks was also found to be even. An analysis of a range of demographic background characteristics of the initial and the effective samples (Table 3) shows that the two treatments only differ significantly with respect to gender in both the initial and the effective samples[11]. In the effective sample, women account for 40 percent of the total in treatment A whereas this share is 51 percent in treatment B.

[10] Before the initial sample of 465 was established, 145 responses were removed as they stated that they use pouches without a filter. These respondents were removed as they would not be able to relate to all of the attributes presented to them in the choice sets and would therefore not be able to make the required trade-offs. With the removal of these particular respondents, the mail out sample of 1,200 is actually equivalent to approximately 900, thereby resulting in a higher response rate.

[11] To test if the *PRE* has an effect on the distributions of the various demographic background characteristics of the initial and effective samples (i.e. sampling bias), an analysis of whether the characteristics differ significantly between the initial and effective samples was carried out. The χ^2-tests reveal that there was no significant difference between the characteristics of the initial and effective samples in both treatments. Results are available from the authors upon request.

Table 3 Respondent Demographics Compared Across Treatments

	Initial sample			Effective sample		
	Treat-ment A %	Treat-ment B %	Significance in χ^2-test[a]	Treat-ment A %	Treat-ment B %	Significance in χ^2-test
Gender						
Male	39	47	*	40	51	*
Female	61	53		60	49	
Household gross income (SEK)						
<150,000	12	12	NS	11	9	NS
150,000-299,999	32	28		33	26	
300,000-499,999	32	33		29	31	
>500,000	24	28		27	34	
Age						
18-34	6	6	NS	9	7	NS
35-54	33	27		29	32	
55-74	61	67		62	62	
Education						
Primary and vocational	75	72	NS	72	68	NS
Short-middle academic	16	17		16	19	
Long-term academic	9	10		12	13	
Type of stoma[b]						
Colostomy	34	34	NS	31	33	NS
Ileostomy	66	66		69	67	
Length of time of stoma being in place						
<1 yr	1	0.5	NS	1	1	NS
1-5 yrs	31	31		31	35	
5-10 yrs	27	26		23	27	
>10 yrs	42	41		45	38	
Blocking						
Block 1	-	-		58	70	NS
Block 2	-	-		58	75	

*NS indicates no significant difference at 95% level, * indicates a significant difference at 95% level.*
[a] *The χ^2-test is employed on the basis of the actual numbers behind the percentages.*
[b] *There are three types of stoma: Colostomy, ileostomy, urostomy. The sample does not contain ostomates with urostomy.*

Due to the different distributions of gender in the two treatments, analyses were carried out on an overall level as well as on a gender specific level. This approach was taken in order to ascertain whether potential differences with regard to both the number of *PZ* bidders and the preferences of the respondents in the two treatments are caused merely by an overall impact of the *PRE* or if there is a gender specific effect.

3.1 Number of Protest Zero Bids

The aim of the *PRE* was to reduce the number of *PZ* bids. Table 4 compares the number of *PZ* bids in the two treatments.

Table 4 Distribution of *PZ* Bids Across Treatments

	Number of *PZ* bids (percentage of particular sample)			
Sample	Treatment A	Treatment B	Percentage difference[a]	χ^2-test[b]
Full	72 (32%)	48 (20%)	-38.6%	17.16^{***}
Male	29 (34%)	21 (19%)	-40.3%	11.23^{***}
Female	43 (32%)	27 (20%)	-32.2%	6.05^{*}

[a] *Weighted according to the number of respondents in each sample.*
[b] *Taking into consideration the number of respondents in each sample.*

The above analysis shows that, while there are still a high number of *PZ* bids in both treatments, there is a significant difference between the two treatments. Treatment B seems to have a considerably smaller number of *PZ* bids when compared to treatment A. More specifically, the *PRE* seems to have reduced the number of respondents stating *PZ* behaviour by almost 40 percent. Chi-tests indicate that the differences observed are significant, indicating that *H1* is rejected and *H2* is not rejected, thereby suggesting that the *PRE* has had the desired effect.

When looking at the number of *PZ* bids on a gender specific level, the *PZ* numbers could suggest that female respondents are more prone to state a *PZ* preference compared to males. However, taking into account the differences in the distribution of males and females in treatment A, Pearson Chi-tests cannot reject the hypothesis of an equal distribution of *PZ* bidders between male and female respondents, in either of the two treatments[12]. This suggests that male and female respondents are equally sensitive to the *PRE* applied.

3.2 The Parametric Analysis

Moving on to the parametric analysis and the test of *H3*, we apply a Mixed Logit model. The model does not exhibit the restrictive Independence of Irrelevant Alternatives (IIA) property[13], it allows for random taste variation and for correlation in unobserved utility over alternatives and time (Train 2003). The model is specified with the price coefficient being fixed and all other coefficients being randomly normally distributed[14]. The model describing the elicited preferences for treatments A and B is presented in Table 5.

[12] Test values are available from the authors upon request.
[13] IIA was found to be violated in early stages of analysis. To test for possible violations of IIA, the test developed by Hausman and McFadden (1984) was used.
[14] It could be argued that a normal distribution might not be entirely appropriate as it allows for negative and positive values in the sample. However, several of the respondents have expressed that they presently have a well functioning ostomy pouch in the sense that they have very few leakages and a filter lifetime of more than 12 hours. Consequently, some part of the respondents might associate the attribute levels with negative utility.

Table 5 Results of Mixed Logit Model.
Comparison of estimated marginal WTP in SEK per month

Parameter	Treatment A			Treatment B			ΔWTP
	Mean estimates	Standard deviation	WTP [95% CI][a]	Mean estimates	Standard deviation	WTP [95% CI]	(*t*-test sig.)[b]
Leaks_1							
Full	1.4582^{**}	0.0430^{NS}	265 [67-463]	1.3066^{**}	0.1792^{NS}	191 [76-306]	$74^{(-)}$
Male	1.4602^{*}	0.0172^{NS}	361 [-56-778]	1.9011^{NS}	0.1716^{NS}	179 [24-333]	$182^{(-)}$
Female	1.5488^{*}	0.0514^{NS}	231 [-1-462]	1.8234^{**}	0.1451^{NS}	229 [69-389]	$2^{(-)}$
Leaks_0							
Full	3.0125^{***}	0.0982^{NS}	548 [354-742]	2.4825^{***}	1.5637^{**}	363 [249-477]	$185^{(-)}$
Male	2.9910^{***}	0.2106^{NS}	739 [231-1247]	3.6971^{*}	4.0091^{NS}	347 [188-507]	$392^{(-)}$
Female	3.2943^{***}	0.0045^{NS}	491 [287-695]	3.1526^{***}	0.8292^{NS}	395 [251-540]	$96^{(-)}$
Filtlife_12							
Full	0.6557^{NS}	0.1945^{NS}	119 [-12-251]	0.5668^{NS}	0.0861^{NS}	83 [-14-180]	$36^{(-)}$
Male	0.3029^{NS}	0.0456^{NS}	75 [-181-330]	1.1398^{NS}	1.0495^{NS}	107 [-21-236]	$-32^{(-)}$
Female	1.0564^{NS}	0.2567^{NS}	158 [-2-317]	0.3708^{NS}	0.1526^{NS}	47 [-76-169]	$111^{(-)}$
Filtlife_24							
Full	0.6230^{NS}	0.0151^{NS}	113 [-32-259]	1.0902^{**}	1.1358^{NS}	159 [55-264]	$-46^{(-)}$
Male	-0.9516^{NS}	0.3652^{NS}	-24 [-316-269]	1.9564^{NS}	2.7429^{NS}	184 [43-325]	$-207^{(-)}$
Female	1.2742^{NS}	0.0137^{NS}	190 [19-360]	1.1375^{NS}	0.9489^{NS}	143 [-10-295]	$47^{(-)}$
Flex_small							
Full	1.5493^{**}	0.0434^{NS}	282 [124-440]	0.5815^{NS}	0.2252^{NS}	85 [-29-199]	$197^{(+)}$
Male	1.3187^{NS}	0.3578^{NS}	326 [-3-655]	1.7515^{NS}	0.5772^{NS}	165 [10-320]	$161^{(-)}$
Female	1.7839^{*}	0.0865^{NS}	266 [81-451]	-0.1995^{NS}	0.7780^{NS}	-25 [-200-150]	$291^{(+)}$
Flex_large							
Full	1.4423^{**}	1.4423^{**}	262 [106-419]	0.6820^{NS}	1.8690^{**}	100 [-28-227]	$163^{(-)}$
Male	1.1235^{NS}	0.0858^{NS}	278 [-33-588]	1.8326^{NS}	3.7744^{NS}	172 [8-336]	$106^{(-)}$
Female	1.6979^{*}	0.1151^{NS}	253 [67-440]	0.2862^{NS}	1.2824^{NS}	36 [-144-216]	$217^{(-)}$
ASC							
Full	-3.0099^{***}	2.5592^{***}	-548 [-823-(-272)]	-1.3487^{**}	3.3558^{***}	-197 [-357-(-37)]	$-351^{(+)}$
Male	-3.2599^{***}	2.1107^{***}	-805 [-1494-(-116)]	-0.6770^{NS}	6.7947^{*}	-64 [-326-198]	$-741^{(+)}$
Female	-2.9968^{***}	2.9224^{***}	-447 [-736-(-157)]	-1.2502^{NS}	2.2972^{***}	-157 [-346-32]	$-290^{(-)}$
Price							
Full	-0.0055^{***}			-0.0068^{***}			
Male	-0.0040^{**}			-0.0106^{*}			
Female	-0.0067^{***}			-0.0080^{***}			
N	505, 208, 293			621, 314, 290			
Simulations	500			500			
LL(0)	-554.8, -228.5, -321.9			-682.2, -345.0, -318.6			
LL(b)	-256.3, -108.0, -142.4			-381.8, -208.6, -153.5			
Pseudo-R^2	0.54, 0.53, 0.56			0.44, 0.40, 0.52			

*NS indicates no significance, * indicates significance at 95%, ** at 99% level and *** at 99.9% level.*

[a] *95% confidence intervals are estimated using the Delta Method in accordance with Greene (2003) and Hanemann and Kanninen (1999).*

[b] *An asymptotic t-test of the significance of the differences in WTP. (-) indicates no significant difference in WTP. (+) indicates a significant difference at the 95% level.*

Parameter estimates denote the sample marginal utility associated with a change from the average sample status quo. Looking at the model based on the full dataset (both genders), Table 5 shows that the parameter estimates follow expectations of quality improvements positively contributing to marginal utility. The parameter estimates for filter lifetime of 12 hours in both treatments, filter lifetime of 24 hours in treatment A and the variables representing increased flexibility of the system in treatment B have the expected sign, but they are found to be insignificant. In addition, it would appear that the parameter estimates from treatment B have higher internal validity than those from treatment A, in the sense that better attribute levels are always numerically associated with higher utility compared to worse attributes. In other words treatment B has a preferable preference structure.

The parameter estimates of the means indicate that respondents derive the highest marginal utility from the leakages attribute followed by flexibility and finally by filter lifetime. The coefficient for the alternative specific constant (ASC) represents the marginal utility associated with either one of the two hypothetical alternatives opposed to the status quo alternative[15]. The significance of the ASC indicates that respondents are not indifferent between staying with their current system and changing to one of the two hypothetical alternatives. The coefficient has a negative sign, which indicates that respondents have a preference for staying with their current system (i.e. disutility of change), which suggests that either the respondents have strong confidence and feel secure with their current system or that a status quo effect might be evident (Meyerhoff and Liebe 2009). Interestingly, there does not seem to be much systematic variance around the means in either of the two treatments except for Leaks_0 (treatment B only), Flex_large and the ASC.

Dividing parameter estimates by the negative price parameter results in estimates of WTP to secure the level specified improvement in the ostomy pouch attributes. Given potentially different scale parameters in the two models, the parameter estimates cannot be directly compared across models (Louviere *et al*. 2000; Swait and Louviere 1993). However, a direct comparison can be made with regard to the WTP estimates, as the scale parameter cancels out in this calculation (Train 2003).

3.2.1 Differences in WTP

The numerical differences in WTP estimates are reported in the far right column of Table 5. WTP estimates in treatment B are generally lower than in treatment A. These numerical differences could indicate that the *PRE* causes some respondents to express lower WTP and thus might have caused respondents in treatment B to be less prone to hypothetical biasness. To ascertain whether the differences are significant, a *t*-test is carried out for each of the WTP differences, testing the null of equal WTP in the two treatments.

The results of the tests for identical WTP between the two treatments show that only the *t*-tests for Flex_small and the ASC reject the null. This indicates that the *PRE* has not resulted in overall significant differences in preferences between the two treatments. The WTP for Flex_small is shown to be significantly lower for treatment B with a stated WTP of 85 SEK compared to 282 SEK for treatment A. However, accounting for the fact that there are more respondents in employment in treatment A compared to treatment B, we find that when running the models again there are no significant differences in preferences when respondents who are not in employment are excluded from the analysis. This suggests that the significance of this difference is due to a difference in respondents and not a difference in preferences.

The significant WTP difference between treatments for the ASC indicates that respondents in treatment B are more willing to choose one of the two policy-generated alternatives (less disutility of change). This could also be an indication of individuals

[15] Care should be taken with this interpretation of the ASC as it only holds under the assumption that the respondent has made the required trade-offs when making their choice between the status quo and the two hypothetical alternatives. That is to say that the respondent has to have considered all of the attributes and attribute levels of the two hypothetical alternatives before deciding to choose the status quo and not some rule of thumb, such as lexicographical preferences (Rekola 2003).

in treatment B being more accurate in their responses or merely more willing to participate in the survey. We will return to this in the discussion.

3.2.2 Differences in Preferences

An alternative way to examine the effects of the *PRE* is to consider effects on overall preferences. To formally test the hypothesis of identical preferences in the two treatments, a likelihood ratio (LR) test for nested models is conducted. The test statistic $LR = -2(LL_{Pooled\ model} - (LL_{Treatment\ A} + LL_{Treatment\ B}))$ is asymptotically χ^2-distributed with (K+1) degrees of freedom, where K is the number of variables in the models (Swait and Louviere 1993). Table 6 shows the LR tests for identical preferences, where μ is the estimated relative scale parameter between treatment A (normalised to 1) and treatment B.

Table 6 LR Tests for Equality of Model Parameters

	$LL_{TreatmentA}$ + $LL_{TreatmentB}$	$LL_{Pooled\ model}$	LR test, DF = 16	P-value
Full	-256.3-381.8 = -638.1	-649.5 ($\mu = 1.10^{NS}$)	22.8	0.120
Male	-108.0-208.6 = -316.6	-330.6 ($\mu = 1.30^{NS}$)	28.0	0.032
Female	-142.4-153.5 = -295.9	-300.7 ($\mu = 0.90^{NS}$)	9.6	0.886

*NS indicates no significance, * indicates significance at 95%, ** at 99% level and *** at 99.9% level.*
[a] *μ is the estimated relative scale parameter between treatment A (normalised to 1) and treatment B.*

The results show that on the full sample level *H3* cannot be rejected on conventional levels of significance, i.e. the respondents in the two treatments have stated identical preferences. In other words, we cannot reject that the presence of the *PRE* has not significantly affected overall preferences. However, on the gender specific level it would appear that males have significantly different preferences. We will return to this in the discussion.

4 Discussion

4.1 Non-heterogeneous Effects of the Entreaty

Recently a number of stated preference studies have focused on potential heterogeneous biases across respondents. More specifically, the susceptibility of biased behaviour appears to vary systematically with the characteristics of the respondents (Aprahamian *et al*. 2007; Flachaire and Hollard 2007). In this line Ladenburg and Olsen (2008a; 2008b) find and argue that male and female respondents might differ in susceptibility to biases and differ in how they respond to attempts to reduce the bias. Ladenburg and Olsen (2008b) thus present several studies supporting their hypothesis[16]. However, it cannot be rejected that the *PRE* has an equal effect on male and female respondents in the present study. Referring to Meyers-Levy and Maheswaran (1991) and Meyers-Levy and Sternthal (1991), our results suggest that

[16] See for example Ladenburg *et al*. (2007); Andersson and Svensson (2008); Brown and Taylor (2000); Mitani and Flores (2007) and Lundhede *et al*. (2008).

the male and female respondents have found the information in the *PRE* equally motivating and have processed the *PRE* similarly. In this light, these results are validating for the use of the *PRE* to reduce *PZ* bids in CE studies.

However, as indicated by both the gender specific differences in WTP and LR-test for equality of model parameters shown in Tables 5 and 6, there is a significant gender specific difference in preferences between treatments. More specifically, WTP of female respondents in treatment B have significantly weaker preferences for a small change in the flexibility of the system compared to the female respondents in treatment A. As such, the observed difference in WTP in the full sample model seems to be mainly due to differences in WTP between the female respondents in the two treatments. Similarly, the estimated full sample mean difference in preferences for the status quo alternative appears to be mainly due to differences between male respondents in the two treatments. The observations are supported by the LR-test, in which overall preferences of male respondents are significantly different on a 95 percent confidence level.

4.2 Length of the Entreaty

The aim of the present paper is to test a relatively short *PRE*, when compared to Mores-Jones *et al*. (2007) and Mourato *et al*. (2008). As mentioned, the motivation for this is to minimise the cognitive burden of the respondents and to keep the length of the scenario description at a minimum. Using entreaties of 192 and 271 words respectively, both Morse-Jones *et al*. (2007) and Mourato *et al*. (2008) find a significant effect of their entreaties and are able to reduce the number of *PZ* bids with 30-40 percent. As presented, we apply a script of 93 words and the results suggest that even this short entreaty is equally effective in reducing the number of *PZ* bids. More specifically, the *PRE* has reduced the number of *PZ* bids with 38.6 percent. Accordingly, in the light of the increasing demands for setting up the hypothetical market in stated preference surveys, and the fact that the length of the hypothetical market description is not an unlimited resource, our results are promising in the sense that a relatively short entreaty can be used to setup a more efficient hypothetical market.

4.3 Reducing Status Quo Effects

The aim of the present paper is to test an entreaty to reduce *PZ* behaviour in a CE application of a stated preference survey. However, as indicated by the change in the preferences for the status quo alternative, the entreaty also seems to have reduced status quo effects. Though this is not an issue that will be discussed in detail, these results are noteworthy, as they validate the use of a *PRE* in stated preference surveys. In this line, the apparent reduction of status quo effects have been achieved without inducing the respondents to increase the demand for the hypothetical good and thereby creating hypothetical biases. More specifically, even though the preferences for the status quo are significantly reduced, the marginal preferences for the attributes have not become stronger. On the contrary, the marginal demand for the attributes appears to have been reduced.

5 Conclusion

We find that our short and simple *PRE* presented to respondents prior to the actual choice sets has effectively reduced the number of *PZ* bids by almost 40 percent. The respondents' marginal demand for improvements of the individual attributes, as well as general preferences, do not seem to have been significantly influenced by the *PRE*. This implies that the ex-ante reduction in *PZ* bids conducted here is an easy-to-implement and effective method to reduce this type of bias. Furthermore, as an additional effect of the *PRE*, the entreaty seems to have had an influence on the level of status quo bias, especially for male respondents, suggesting a gender specific bias mechanism at work.

6 Acknowledgements

We gratefully acknowledge the helpful and valuable comments of Søren Bøye Olsen and Sinne Smed at the University of Copenhagen and two anonymous reviewers. We also thank Coloplast A/S for providing financial support.

7 References

Aadland, D. and Caplan, A. (2003) Willingness to pay for curbside recycling with detection and mitigation of hypothetical bias, *American Journal of Agricultural Economics*, 85(2), 492-502.

Andersson, H. and Svensson, M. (2008) Cognitive ability and scale bias in the contingent valuation method, *Environ Resource Economics*, 39, 481-495.

Aprahamian, F., Chanel, O. and Luchini, S. (2007) Modeling starting point bias and unobserved heterogeneity in contingent valuation surveys: An application to air pollution, *American Journal of Agricultural Economics*, 89, 533-547.

Banzhaf, M.R., Reed, F.R. and Mathews, K.E. (2001) Opt-out alternatives and anglers' stated preference, In: Bennett, J. and Blamey, R. (eds.): The choice modeling approach to environmental valuation, Northampton, MA: Edward Elgar Publishing.

Bonnichsen, O. (2009) Elicitation of preferences for improvements in ostomy pouches – A choice experiment. Working paper.

Boyle, K.J. (2003) Contingent valuation in practice. In: Champ, P.A., Boyle, K.J., Brown, T.C. (eds.): A primer on nonmarket valuation. Kluwer, Dordrecht, Nederlands, 111-169.

Boyle, K.J. and Bergstrom, J.C. (1999) Doubt, doubt, and doubters: the genesis of a new research agenda? In: Bateman, I.J., Willis, K.G. (eds.): Valuing environmental preferences, Theory and practice of the contingent valuation method in the US, EU and developing countries, Oxford University Press, Oxford, 183-206.

Brown, K.M. and Taylor, L.O. (2000) Do as you say, say as you do: evidence on gender differences in actual and stated contributions to public goods, *Journal of Economic Behavior and Organization*, 43(1), 127-139.

Bryan, S., Buxton, M., Sheldon, R., and Grant, A. (1998) Magnetic resonance imaging for the investigation of knee injuries: An investigation of preferences, *Health Economics*, 7(7), 595-603.

Carlsson, F., Frykblom, P. and Lagerkvist, C.J. (2005) Using cheap talk as a test of validity in choice experiments, *Economics Letters*, 89(2), 147-152.

Cataldo, P.A. (1999) Intestinal stomas: 200 years of digging, *Diseases of the Colon and Rectum*, 42(2), 137-142.

Cummings, R.G. and Taylor, L.O. (1999) Unbiased value estimates for environmental goods: A cheap talk design for the contingent valuation method, *American Economic Review*, 89(3), 649-665.

Flachaire, E. and Hollard, G. (2007) Starting point bias and respondent uncertainty in dichotomous choice contingent valuation surveys, *Resource Energy Economics*, 29(3), 183-194.

Greene, W.H. (2003) Econometric Analysis, Fifth Edition, Prentice-Hall, Inc., Upper Saddle River, New Jersey.

Hanemann, M.W. and Kanninen, B. (1999) The statistical analysis of discrete-response CV data, In: Bateman, I.J. and Willis, K.G. (eds.) Valuing Environmental Preferences – Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries, Oxford University Press, Oxford.

Hausman, J.A. and McFadden, D. (1984) Specification tests for the multinomial logit model, *Econometrica*, 52(5), 1219-1240.

Huber, J. and Zwerina, K. (1996) The importance of utility balance in efficient choice designs, *Journal of Marketing Research*, 33, 307-317.

Jorgensen, B.S., Syme, G.J., Bishop, B.J., Nancarrow, B.E. (1999) Protest responses in contingent valuation, *Environ Resource Economics*, 14(1), 131-150.

Jorgensen, B.S., Wilson, M.A. and Heberlein, T.A. (2001) Fairness in the contingent valuation of environmental goods: attitude toward paying for environmental improvements at two levels of scope, *Ecological Economics*, 36, 133–148.

Kuhfeld, W. (2005) Marketing Research Methods in SAS. Experimental Design, Choice, Conjoint and Graphical Techniques, SAS Institute Inc.

Ladenburg, J. and Olsen, S.B. (2008a) Gender-specific starting point bias in choice experiments: Evidence from an empirical study, *Journal of Environmental Economics and Management*, 56, 275-285.

Ladenburg, J. and Olsen, S.B. (2008b) Gender anomalies in choice experiments – are biases really gender dependent? Working paper.

Ladenburg, J., Olsen, S.B. and Nielsen, R.C.F. (2007) Reducing hypothetical bias in choice experiments - Testing an opt-out reminder, manuscript presented at the *European Association of Environmental and Resource Economists' 15th annual conference*, 27-30 June 2007, Thessaloniki, Greece.

Lancaster, K.J. (1998) A new Approach to Consumer Theory. In: Consumer Theory, Edward Elgar Publishing Limited, Cheltenham.

List, J.A. (2001) Do explicit warnings eliminate the hypothetical bias in elicitation procedures? Evidence from field auctions for sportscards, *The American Economic Review*, 91(3), 1498-1507.

Louviere, J.J., Hensher, D.A. and Swait, J.D. (2000) Stated choice methods, analysis and application, Cambridge University Press.

Lundhede, T.H., Olsen, S.B. and Jacobsen, J.B. (2008) Choice experiments and certainty in choice: A test of the influence of utility difference on self-reported certainty levels and evaluations of two recoding approaches to handle uncertain responses, manuscript presented at the *European Association of Environmental and Resource Economists' 16th annual conference*, 25-28 June 2008, Gothenburg, Sweden.

Mitani, Y. and Flores, N. (2007) Does gender matter for demand revelation in threshold public goods experiments?, *Economic Bulletin*, 3(27), 1-7.

Mitchell, R.C. and Carson, R.T. (1989) Using surveys to value public goods: The contingent valuation method, Resources for the Future, Washington DC, USA.

Meyerhoff, J. and Liebe, U. (2006) Protest beliefs in contingent valuation: Explaining their motivation, *Ecological Economics*, 57(4), 583-594.

Meyerhoff J. and Liebe U. (2009) Status quo effects in choice experiments: Empirical evidence on attitudes and choice task complexity, *Land Economics*, 85(3), 515-528.

Meyers-Levy, J. and Maheswaran, D. (1991) Exploring differences in males' and females' processing strategies, *Journal of Consumer Research*, 18, 63-70.

Meyers-Levy, J. and Sternthal, B. (1991) Gender differences in the use of message cues and judgements, *J Marketing Res*, 28, 84-96.

Morse-Jones, S., Mourato, S. and Atkinson, G. (2007) Exploring public preferences for tropical biodiversity: Can entreaties help to reduce protest zeros?, manuscript presented at the *European Association of Environmental and Resource Economists' 15th annual conference*, 27-30 June 2007, Thessaloniki, Greece.

Mourato, S., Atkinson, G., Morse-Jones, S., Lanz, B. and Provins, A. (2008) When to take “no” for an answer? Using entreaties to reduce protests in stated preference studies, paper selected for presentation at the *European Association of Environmental and Resource Economists' 16th annual conference*, 25-28 June 2008, Gothenburg, Sweden.

Rekola, M. (2003) Lexicographic preferences in contingent valuation: a theoretical framework with illustrations, *Land Economics*, 79(2), 277-291.

Strazzera, E., Genius, M., Scarpa, R. and Hutchinson, G. (2003) The effect of protest votes on the estimates of WTP for use values of recreational sites, *Environmental and Resource Economics*, 25(4), 461-476.

Swait, J. and Louviere, J. (1993) The role of the scale parameter in the estimation and comparison of multinomial logit models, *Journal of Marketing Research*, 30(3), 305-314.

Train, K. (2003) Discrete choice methods with simulation, Cambridge University Press.