

Reducing Protest Zero Bias in Stated Preference Surveys – Testing a Protest Reduction Entreaty in a Health Economic Choice Experiment

Ole Bonnichsen^a, Jacob Ladenburg^b

^a *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*

^b *Danish Institute of Governmental Research, Nyropsgade 37, 1602 Copenhagen V, Denmark*

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

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. As such, 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 zero¹ (PZ) bidders. Potentially, this is a large problem in health economic SPM surveys. The reason is that 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 hypothetical cost 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 the effect of including a cost attribute in a split sample survey was investigated. 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 PZ bidders who are removed 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 & Taylor 1999, Aadland & 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), designed to reduce the respondents’ propensity to state a PZ bid. To the authors’ knowledge, such an entreaty has previously only been tested by the papers of Morse-Jones *et al.* (2007) and Mourato *et al.* (2008). Morse-Jones *et al.* (2007) and Mourato *et al.* (2008) test the effect of an entreaty in two different setups. In Morse-Jones *et al.* (2007), they use the case of preferences for tropical biodiversity protection among people living in London and find in a CVM study, that an entreaty focusing on the potential disbelief in the validity of the scenario, effectively reduced PZ bids. In Mourato *et al.* (2008), the policy context is reductions in the risks of water use restrictions in the future. Here they also find a positive effect of an entreaty.

The present paper extends the work in several dimensions. First of all, in order to keep the scenario description as short as possible and keep the cognitive burden of the respondents at a minimum we apply and test a shorter entreaty compared to the entreaties applied in Morse-Jones *et al.* (2007) 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 is done ex-post in Mourato *et al.* (2008)³. Finally, we focus the entreaty on the issue of a 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

¹ PZ falls in the category of protest biases (Mitchell & Carson, 1989).

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.

survey. Furthermore, the results indicate that the entreaty does not have an influence on the general preference structure.

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. An additional monthly expense was used as the payment vehicle⁴. The attributes were presented to the respondents with the descriptions shown in Table 1 (Bonnichsen, 2009).

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	

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

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 individual⁵.

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 9 choice sets and assigned into two blocks⁷, with the respondents evaluating five and four choice sets, respectively. 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.

	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	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>(mark one box only)</i>			

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.

2.1 Identifying protest zero bidders

There exists no well established or generally agreed upon method for identifying protest bidders and this is typically done using an ad-hoc approach (Jorgensen *et al.*, 1999; Boyle & Bergstrom, 1999). In our survey, the PZ bidders were identified using a number of debriefing questions. These questions were presented to the respondents, who in the four or five 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.

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 & Zwerina (1996).

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

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 & Liebe, 2006).

2.2 Procedure

In order to test for the effect of the PRE, a two-split sample design was utilised. In both splits respondents were introduced to a hypothetical market entailing choices between alternative ostomy pouches. Included in this section was a description of the attributes, “cheap talk” focusing on the issue of hypothetical bias and budget reminders prior to the choice session. The questionnaires for both samples were kept identical with the exception that respondents in split B were given a PRE, while respondents in split 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.”

To model the effect of the PRE, a theoretical choice framework is setup by elaborating on the work by Ladenburg & Olsen (2008a)⁹. Assume that a utility maximising individual i is faced with a choice a related to a single action, $a \in (0,1)$. In the present setup this means that respondents choose one alternative ostomy pouch from a choice set consisting of three alternative ostomy pouches. In the model, the choice of action influences the individual’s utility (U_i) through two separable components; a wealth component W_i and a moral component M_i .

$$U_i(a, n, s, q, z) = M_i(a, n, s, z) + W_i(a, q, c) \quad (1)$$

⁸Stating “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” may 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 4 respondents from split A and 3 from split B have this response to the debriefing question.

⁹Ladenburg & Olsen (2008) test the influence of a price vector in an instructional choice set used in a CE concerning protection of specific types of nature when planning motorways. In their paper, they attune a more theoretical model initially suggested by Levitt & List (2007).

The wealth component W_i depends on the perceived utility q of choosing action a and the implicit cost of that choice c . Specifically, $\partial W_i / \partial q > 0$ and $\partial W_i / \partial c < 0$ meaning that W_i increases as the perceived value of the ostomy pouch improvements increases and cost decreases.

With regard to M_i ¹⁰, n represents the costs of social norms, also referred to as social and moral costs in the literature. In the initial model by Levitt & List (2007), n is included to control for factors such as social norms and legal rules, which prescribe specific behaviour in a particular society. As such, the size and the sign of the n element might be influenced by several factors. Following the argumentation in Ladenburg & Olsen (2008a), it can be argued that n might entail a moral responsibility to choose the ostomy pouch with the best levels of the attributes, even though the relative cost of that action exceeds the benefits that would accrue to respondent i . Such behaviour could be motivated by individual beliefs concerning that harmful things might occur due to inaction (Schwartz, 1970; 1976). On the other hand, the individual might also derive utility specifically by accepting high-cost ostomy pouch improvements, because they see themselves as honest and socially responsible individuals (Andreoni & Vesterlund, 2001; Nyborg, 2000; Akerlof & Dickens, 1982). Independently of motivation for the cost of social norms (negatively motivated by not doing or positively motivated by doing), the cost of social norms positively influences M_i so that $\partial M_i / \partial n > 0$.

In the present CE setup, this would mean that individuals might have a tendency to choose a policy generated alternative (with better attribute levels compared to the individuals' present ostomy pouch) too frequently, when compared to the choices made solely based on the W_i component.

The second component in M_i is s , which refers to the fact that the choice of ostomy pouch might also be governed by moral utility if the respondent feels scrutinized by filling in the questionnaire. The final variable is z_i , which is added to the moral utility component to capture the choice behaviour in which individual i chooses the status quo ostomy pouch, even though individual i would gain a net benefit by choosing an improved ostomy pouch with higher costs (i.e. the individual is a PZ bidder). In the literature, the behaviour of PZ bidders appears to be triggered by different protest elements, p_k in the valuation experiment/scenario description, 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 good¹¹ (Boyle, 2003; Strazzera *et al.*, 2003; Jorgensen *et al.*, 1999; Meyerhoff & Liebe, 2006). Let p_k represent the k^{th} protest element in the scenario description and under the assumption that p_k has a positive influence on the propensity to exhibit PZ behaviour, it is argued that $\partial z_i / \partial (p_k) > 0$. Accordingly, the PZ behaviour becomes an increasing function of the protest elements in the survey, i.e. the higher number of

¹⁰ In Levitt & List (2007) M_i also included a parameter v representing the externality an action can have on other individuals. More specifically, v represents the loss to other people as a consequence of the gain individual i achieved by action a in for example a lab experiment involving payoffs. In a more policy relevant setting, Ladenburg & Olsen (2008a) argue that v can represent the negative externality with regard to the level of provision of an environmental good. However, in the present study, it is argued that a link between the choice of action and the impact on other individuals is not directly established. First of all, the good in question is a private good and the scenario description does not facilitate a dependency between individual i and other ostomates via the action a . If the text in the questionnaire had a research and development orientated profile, such a dependency (and thereby the relevance of including v in the utility function) would probably have been more relevant.

¹¹ It could be argued that z is a part of n , however in the present application, n is interpreted as general social norms in society or within groups of individuals, whereas z is assumed to be an individual's assessment of the valuation process with regard to evoking a PZ response.

elements which can trigger a protest response, the higher the propensity to protest¹². However, in the stated preference data we do not observe z_i on a continuous scale, but only in a dichotomous dimension in the sense that either the individual gives a PZ bid or not.

In order to link the nature of the stated behaviour with the model, we introduce an individual PZ threshold λ_i , which when exceeded ($z_i(p) > \lambda_i$), the respondents exhibit PZ behaviour. Via $z_i(p)$, the propensity to exhibit PZ choices thereby becomes a function of the individual's perceptions of the valuation setup and the contents of the scenario description. This is a suitable framework for testing the influence of the PRE embedded in the scenario description.

2.3 Hypothesis

PZ behaviour is, as put forward, both a function of the characteristics of respondent i and the different protest elements, p_k , in the survey setup and scenario description. By including the PRE, we try to add another element to the survey, p_{PRE} , which via the entreaty formulation directly aims to reduce the impact of the other p_k elements on the propensity to state PZ behaviour. Consequently the following main hypothesis is put forward:

H1: Propensity to state a PZ preference is independent of the treatment (PRE) and thus identical in split A and B, i.e. $\partial z_i / \partial (p_{PRE}) = 0$

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, i.e. $\partial z_i / \partial (p_{PRE}) < 0$

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 & Taylor, 1999; List, 2001), the PRE is not intended to influence the preference structure 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. $\text{Preferences}_{\text{Split A}} = \text{Preferences}_{\text{Split B}}$.

3. Results

Of the 1,200 questionnaires mailed to the respondents, an initial sample (including PZ bids) of 465 responses was obtained¹³. This was subsequently trimmed to an effective sample (excluding PZ bids) of 254 usable responses. The distribution of effective responses between the two splits was found to be similar and not

¹² 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.

¹³ 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.

significantly different, with split A containing 116 responses and split B containing 138 responses. An analysis of a range of demographic background characteristics of the initial and the effective samples (see Table 3) shows that the two splits only differ significantly with respect to gender in both the initial and the effective samples¹⁴. With regard to the effective sample, women account for 40% of the total in split A whereas this share is 51% in split B.

TABLE 3
Respondent demographics compared across splits

	Initial sample			Effective sample		
	Split A %	Split B %	Significance in χ^2 -test ^a	Split A %	Split 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		11	9	
150,000-299,999	32	28	NS	33	26	NS
300,000-499,999	32	33		29	31	
>500,000	24	28		27	34	
Age						
18-34	6	6		9	7	
35-54	33	27	NS	29	32	NS
55-74	61	67		62	62	
Education						
Primary and vocational	75	72		72	68	
Short-middle academic (college)	16	17	NS	16	19	NS
Long-term academic (university)	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 in place						
<1 yr	1	0.5		1	1	
1-5 yrs	31	31	NS	31	35	NS
5-10 yrs	27	26		23	27	
>10 yrs	42	41		45	38	

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 splits, analyses are carried out on an overall level as well as on a gender specific level. This approach is 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 splits are caused merely by an overall impact of PRE or if there is a gender specific effect.

¹⁴ 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 splits. Results are available from the authors upon request.

3.1 Number of protest zero bids

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

TABLE 4
Distribution of PZ bids across splits

Sample	Number of PZ bids (% of particular sample)		% difference ^a	χ^2 -test ^b
	Split A	Split 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 splits, there is a significant difference between the two splits. Split B seems to have a considerably smaller number of PZ bids when compared to split A. More specifically, the PRE seems to have reduced the number of respondents stating PZ behaviour with almost 40% on a full sample level. Referring to section 2.2 this can be interpreted as $\partial z_i / \partial (p_{PRE})$. 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 split 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 splits¹⁵. This also 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 Multinomial Probit Model¹⁶ (MNP). The model is specified by allowing for identically distributed error components of the three alternatives and allows for correlation between the two policy-generated alternatives (Bunch, 1989; Weeks, 1997; Train, 2003). The MNP describing the elicited preferences for splits A and B is presented in Table 5. The following parametric models are based on the full sample as well as on gender specific samples.

¹⁵ Test values are available from the authors upon request.

¹⁶ This type of model was chosen to avoid the restrictive Independence of Irrelevant Alternatives (IIA) property (Train, 2003), which was found to be violated in early stages of analysis. To test for possible violations of IIA, the test developed by Hausman & McFadden (1984) was used.

TABLE 5

Results of Multinomial Probit Model. Comparison of estimated marginal WTP in SEK per month

Parameter	Split A		Split B		Δ WTP (T-test sig.) ^b
	Estimates	WTP [95% CI] ^a	Estimates	WTP [95% CI]	
Leaks_1					
Full	0.4681***	213 [89-337]	0.4800***	232 [102-362]	-19 ⁽⁻⁾
Male	0.3838 ^{NS}	186 [4-368]	0.3400*	259 [23-495]	-73 ⁽⁻⁾
Female	0.4831*	197 [20-373]	0.8032**	240 [90-390]	-43 ⁽⁻⁾
Leaks_0					
Full	0.8245***	375 [214-536]	0.8305***	402 [253-551]	-27 ⁽⁻⁾
Male	0.8040*	389 [105-673]	0.5892**	448 [142-755]	-59 ⁽⁻⁾
Female	0.8919***	363 [175-551]	1.3301***	398 [241-556]	-35 ⁽⁻⁾
Filtlife_12					
Full	0.0917 ^{NS}	42 [-69-153]	0.1790 ^{NS}	87 [-18-191]	-45 ⁽⁻⁾
Male	0.2409 ^{NS}	117 [-64-297]	0.2172 ^{NS}	165 [-40-371]	-48 ⁽⁻⁾
Female	0.1082 ^{NS}	44 [-107-195]	0.0617 ^{NS}	18 [-90-46]	26 ⁽⁻⁾
Filtlife_24					
Full	0.0830 ^{NS}	38 [-58-133]	0.4009**	194 [89-299]	-156 ⁽⁺⁾
Male	-0.0909 ^{NS}	-44 [-192-104]	0.2756*	210 [19-401]	-254 ⁽⁺⁾
Female	0.3276 ^{NS}	133 [-16-283]	0.5453**	163 [46-281]	-30 ⁽⁻⁾
Flex_small					
Full	0.5214**	237 [98-376]	0.1912 ^{NS}	92 [-34-219]	145 ⁽⁻⁾
Male	0.5054 ^{NS}	245 [32-457]	0.2474 ^{NS}	184 [-24-391]	61 ⁽⁻⁾
Female	0.6279*	256 [70-441]	-0.0743 ^{NS}	-22 [-169-125]	278 ⁽⁺⁾
Flex_large					
Full	0.4886**	222 [92-353]	0.4061**	196 [81-312]	26 ⁽⁻⁾
Male	0.3308 ^{NS}	160 [-32-353]	0.3641*	277 [84-471]	-117 ⁽⁻⁾
Female	0.6843*	279 [96-462]	0.2945 ^{NS}	88 [-46-222]	191 ⁽⁻⁾
ASC					
Full	-0.9533***	-433 [-652-(-215)]	-0.6738***	-326 [-507-(-145)]	-107 ⁽⁻⁾
Male	-1.0022***	-485 [-841-(-130)]	-0.5279**	-402 [-773-(-30)]	-83 ⁽⁻⁾
Female	-1.0016***	-408 [-670-(-146)]	-0.6325*	-189 [-356-(-23)]	-219 ⁽⁻⁾
Price					
Full	-0.0022***		-0.0021***		
Male	-0.0021***		-0.0013***		
Female	-0.0025***		-0.0033***		
Rho_12					
Full	0.9088***		0.5618**		
Male	0.9439***		0.7142***		
Female	0.8698***		0.5985 ^{NS}		
N	505, 208, 293		621, 314, 290		
Simulations	250		250		
LL(0)	-554.8, -228.5, -321.9		-682.2, -345.0, -318.6		
LL(b)	-291.3, -119.0, -166.3		-461.5, -266.1, -172.1		
Pseudo-R ²	0.475, 0.479, 0.483		0.324, 0.229, 0.460		

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 & 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 marginal utility associated with a change from the person specific status quo attribute levels. Generally the parameter estimates are significant and follow expectations of quality improvements positively contributing to marginal utility. The significance of the parameter estimates is generally uniform across splits with the exceptions of Filtlife_24 and Flex_small. These attributes show

varying levels of significance across splits. The parameter estimates for Filtlife_12 are found to not be significantly different from zero in any of the models, but this is to be expected since the respondents own current filter lifetime is reported as more than 9 hours on average.

The parameter estimates 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 alternative¹⁷. 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).

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 & Louviere, 1993). But 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. In the model based on the full sample, WTP estimates in split B are generally higher than in split A. The same tendency is apparent in the models based on the gender specific samples. These numerical differences could indicate that the PRE causes respondents in split B to express higher WTP and thus might have caused some level of hypothetical biasness. On the other hand, the observed differences in the average WTP might also be caused by the inclusion of respondents who without the PRE would have stated a PZ bid. 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 splits.

In the model based on the full sample, the results of the tests for identical WTP between the two splits show that none of the t-tests reject the null, with the exception of Filtlife_24. This indicates that the PRE has not resulted in significant differences in preferences between the two splits. The WTP for Filtlife_24 is shown to be significantly higher for split B than for split A. Respondents in split A state a WTP of 38 SEK compared to 194 SEK in split B. An analysis of the distribution of respondents' self reported own current values for filter lifetime, shows that respondents in split B report a shorter filter lifetime at the lower end of the distribution. More specifically, split B has more than twice as many respondents reporting a current filter lifetime of less than 5 hours than split A¹⁸. Potentially, this could explain the observed differences in WTP.

¹⁷ 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).

¹⁸ Looking at the entire distribution, a χ^2 -test reveals that there is a significant difference between the reported own current filter lifetimes of the two splits. On average respondents in split B report filter lifetimes of 9 hours while those in split A report 9.6 hours.

On a gender specific level, Flex_small is found to have significantly different WTP values for females. More specifically, females in split A have stated a WTP of 256 SEK compared to -22 SEK in split B. When accounting for the fact that there are more women in employment in split A compared to split B, we find when running the models again there are no significant differences in preferences when women who are not in employment are excluded from the analysis. Filtlife_24 is also found to have significantly different WTP between splits for male respondents. The same argumentation used above to explain the Filtlife_24 differences in the model based on the full sample, can also be used to explain this gender specific difference in WTP.

The above reasoning for the observed significant differences in WTP suggests that the significance of these differences is due to a difference in respondents and not a difference in preferences.

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 splits, a likelihood ratio (LR) test for nested models is conducted. The test statistic $LR = -2(LL_{\text{pooled model}} - (LL_{\text{split A}} + LL_{\text{split B}}))$ is asymptotically χ^2 -distributed with $(K+1)$ degrees of freedom, where K is the number of variables in the models (Swait & Louviere, 1993). The LR tests for identical preferences are carried out on models based on the full sample and on gender specific samples in Table 6 where μ is the estimated relative scale parameter between split A (normalised to 1) and split B.

TABLE 6
LR tests for equality of model parameters

	$LL_{\text{splitA}} + LL_{\text{splitB}}$	$LL_{\text{pooled model}}$	LR test, DF = 9	P-value
Full	-291.3-461.5 = -752.8	-759.0 ($\mu^a=0.66^{***}$)	12.4	0.1917
Male	-119.0-266.1 = -385.1	-389.9 ($\mu =0.4^{***}$)	9.7	0.3753
Female	-166.3-172.1 = -338.4	-343.0 ($\mu=1.07^{NS}$)	9.2	0.4190

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

The results show that H3 cannot be rejected, i.e. the respondents in the two splits have stated identical preferences in the full model and in the gender specific models. In other words, we cannot reject that the presence of the PRE has not significantly affected overall preferences. It is interesting to note that male respondents have a significant relative scale parameter. This suggests that males in split B tend to have more variance in their choices. 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 & Hollard, 2007). In this line Ladenburg & 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 & Olsen

(2008b) thus present several studies supporting their hypothesis¹⁹. However, as put forward in the results of the present study, it cannot be rejected that the PRE has an equal effect on male and female respondents. Referring to Meyers-Levy & Maheswaran (1991) and Meyers-Levy & Sternthal (1991), our results suggest that 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.

4.2 Length of the entreaty

The aim of the present paper is to test a relatively short PRE, when compared to Morse-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 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%. 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% in the full sample case. 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 Difference in estimation precision

On an average level, the PRE does not seem to have an effect on the preference structure. However, as indicated by the gender specific Pseudo-R² in Table 5 and the significant estimated ratio between scale parameters in Table 6, the PRE seems to have had an influence on the precision of the estimated models based on the male sample. More specifically, the precision in terms of a lower relative scale parameter in split B of 0.4 when compared to split A, suggests that the variance of the estimated model has increased when the male respondents are given the PRE. Referring also to the drop in the Pseudo-R² in the models based on the male sample of 0.479 in split A to 0.229 in split B, it would seem that male respondents in split B find it more difficult to transform the attributes of the alternatives into consistent choices (Adamowicz *et al.*, 1998). This could be due to the model specification which fits the other samples well, not being as good a fit for the sample based only on males from split B.

5. Conclusion

We find that our short and simple protest reduction entreaty (PRE) presented to respondents prior to the actual choice sets has effectively reduced the number of protest zero (PZ) bids by almost 40%. 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.

6. Acknowledgements

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

¹⁹ See for example Ladenburg *et al.* (2007), Andersson & Svensson (2007); Brown & Taylor (2000), Mitani & Flores (2007) and Lundhede *et al.* (2008).

7. References

- Aadland, D. & 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.
- Adamowicz, W.L., Louviere, J. & Swait, J. (1998): Introduction to attribute-based stated choice methods. National Oceanic and Atmospheric Administration, Washington, USA.
- Akerlof, G.A. & Dickens, W.T. (1982): The economic consequences of cognitive dissonance, *Amer. Econ. Rev.* 72 (3) (1982) 307-319.
- Andersson, H., Svensson, M. (2007): Cognitive ability and scale bias in the contingent valuation method. *Environ Resource Econ* (in press).
- Andreoni, J., & Vesterlund, L. (2001): Which is the fairer sex? Gender differences in altruism, *Quart. J. Econ.* 116 (1) 293-312.
- Aprahamian, F., Chanel, O. & 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. & Mathews, K.E. (2001): Opt-Out Alternatives and Anglers' Stated Preference. In *The Choice Modeling Approach to Environmental Valuation*, eds. J. Bennett and R. Blamey, Northampton, MA: Edward Elgar Publishing.
- Bonnichsen, O. (2009): Elicitation of Preferences for Improvements in Ostomy Pouches – A Discrete 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, Netherlands, pp 111-169.
- Boyle, K.J. & 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, pp 183-206.
- Brown K.M. & Taylor, L.O. (2000): Do as you say, say as you do: evidence on gender differences in actual and stated contributions to public goods. *J Econ Behav Organ* 43(1):127-139.
- Bryan, S., Buxton, M., Sheldon, R., & Grant, A. (1998): Magnetic resonance imaging for the investigation of knee injuries: An investigation of preferences, *Health Economics*, vol. 7, no. 7, pp. 595-603.
- Bunch, D.S. (1989): Estimability in the Multinomial Probit Model, *TronSpnR.e s.-B*, Vol. 25B, No. I, pp. 1-12.
- Carlsson, F., Frykblom, P. & 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, *Dis. Colon Rectum* 42:137-142.
- Cummings, R.G. & 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. & Hollard G. (2007): Starting point bias and respondent uncertainty in dichotomous choice contingent valuation surveys. *Resource Energ Econ* 29(3):183-194.

- Greene, W.H. (2003): *Econometric Analysis*, Fifth Edition, Prentice-Hall, Inc., Upper Saddle River, New Jersey.
- Hanemann, M.W. & Kanninen, B. (1999): The statistical analysis of discrete-response CV data, 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.
- Hausman, J.A. & McFadden, D. (1984): Specification Tests for the Multinomial Logit Model, *Econometrica*, vol. 52, no. 5, pp. 1219-1240.
- Huber, J. & Zwerina, K. (1996): The Importance of Utility Balance in Efficient Choice Designs, *Journal of Marketing Research*, Vol. 33, pp. 307-317.
- Jorgensen, B.S., Syme, G.J., Bishop, B.J., Nancarrow, B.E. (1999): Protest Responses in Contingent Valuation, *Environ Resource Econ* 14(1): 131-150.
- Jorgensen, B.S., Wilson, M.A. & 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, pp. 133–148.
- Ladenburg, J. & Olsen, S.B. (2008a): Gender-specific starting point bias in choice experiments: Evidence from an empirical study. *Journal of Environmental Economics and Management*. Doi: 10.1016/j.jeem.2008.01.004.
- Ladenburg, J., Olsen, S.B. (2008b): Gender anomalies in choice experiments – are biases really gender dependent? Working paper.
- Ladenburg, J., Olsen, S.B. & Nielsen, R.C.F. (2007): Reducing hypothetical bias in Choice Experiments - Testing an Opt-Out Reminder. Paper 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.
- Levitt, S.D. & List, J.A. (2007): What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World? *J. Econ. Perspect.* 21 (2) 153-174.
- List, J.A. (2001): Do explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Field Auctions for Sports cards. *The American Economic Review* 91(3):1498-1507.
- Louviere, J.J.; Hensher, D.A. & Swait, J.D. (2000): *Stated Choice Methods, Analysis and Application*, Cambridge University Press.
- Lundhede, T.H., Olsen, S.B. & 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 submitted to the European Association of Environmental and Resource Economists' 16th annual conference 25-28 June 2008, Gothenburg, Sweden.
- Mitani, Y. & Flores, N. (2007): Does gender matter for demand revelation in threshold public goods experiments? *Econ Bull* 3(27): 1-7.
- Mitchell, R.C. & Carson, R.T. (1989): *Using Surveys to Value Public Goods: The Contingent Valuation Method*, Resources for the Future, Washington DC, USA.

- Meyerhoff, J. & Liebe, U. (2006): Protest beliefs in contingent valuation: Explaining their motivation. *Ecol Econ* 57(4): 583-594.
- Meyers-Levy, J. & Maheswaran, D. (1991): Exploring differences in males' and females' processing strategies, *J. Cons. Res.* 18, 63-70.
- Meyers-Levy, J. & Sternthal, B. (1991): Gender differences in the use of message cues and judgements, *J. Marketing Res.* 28, 84-96.
- Morse-Jones, S., Mourato, S. & Atkinson, G. (2007): Exploring Public Preferences for Tropical Biodiversity: Can Entreaties Help to Reduce Protest Zeros? *Environmental valuation – Contingent valuation (II) EAERE 2007*.
- Mourato, S., Atkinson, G., Morse-Jones, S., Lanz, B. & 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 16th Annual Conference of the European Association of Environmental and Resource Economists, Gothenburg (Sweden).
- Nyborg, K. (2000): Homo economicus and Homo Politicus: interpretation and aggregation of environmental values, *J. Econ. Behav. Org.* 42 305-322.
- Rekola, M. (2003): Lexicographic preferences in contingent valuation: a theoretical framework with illustrations. *Land Economics* 79 (2), 277-291.
- Schwartz, S.H. (1970): Elicitation of moral obligation and self-sacrificing behavior: an experimental study of volunteering to be a bone marrow donor. *J Pers Soc Psychol.* Aug; 15(4): 283-93.
- Schwartz, S.H. & Gottlieb, A. (1976): Bystander reactions to a violent theft: crime in Jerusalem. *J Pers Soc Psychol.* Dec; 34(6): 1188-99.
- Strazzera, E., Genius, M., Scarpa, R. & 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. & Louviere, J. (1993): The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models, *Journal of Marketing Research*, vol. 30, no. 3, pp. 305-314.
- Train, K. (2003): *Discrete Choice Methods with Simulation*. Cambridge University Press.
- Weeks, M. (1997): The Multinomial Probit Model Revisited: A Discussion of Parameter Estimability, Identification and Specification Testing. *Journal of Economic Surveys*, Volume 11, Number 3, September 1997, pp. 297-320(24).