Stability of Willingness to Pay: does time and treatment allocation in a Randomised Controlled Trial influence willingness to pay?

Abstract

Background

Willingness To Pay (WTP) estimates are only useful to policymakers if they are generalisable beyond the moment when they are collected. To understand the 'shelf-life' of preference estimates, preference stability needs be tested over substantial periods of time.

Methods

We test the stability of WTP for preventative dental care (scale and polish) using a payment-card Contingent Valuation question administered to 909 randomised controlled trial participants at four time points: baseline (pre-randomisation) and at annual intervals for three years. Trial participants were regular attenders at a NHS dental practices. Participants were randomly offered different frequencies (intensities) of scale polish (no scale and polish, one scale and polish per year, two scale and polishes per year). We also examine whether treatment allocation to these different treatment intensities influences the stability of WTP. Interval regression methods were used to test for changes in WTP over time whilst controlling for changes in two determinants of WTP. Individual level changes were also examined as well as the WTP function over time.

Results

We find that at the aggregate level mean WTP values are stable over time. The results were similar by trial arm. Individuals allocated to the arm with the highest scale and polish intensity (two per year) had a slight increase in WTP towards the latter part of the trial. There is considerable variation at the individual level. The WTP function is stable over time.

Conclusions

The payment card contingent valuation method can produce stable WTP values in health over time. Future research should explore the generalisability of these results in other populations, for less familiar health care services and using alternative elicitation methods.

Keywords: contingent valuation; stability; willingness to pay; dentistry

Highlights

- Stated preferences are commonly used to value health care.
- Willingness to Pay (WTP) estimates are only useful if they have a ‘shelf-life’.
- Little is known about the stability of WTP for health care.
- We test the stability of WTP for dental care over three years.
- Our results show that the contingent valuation method can produce stable WTP values.

1. INTRODUCTION

The absence of well-functioning markets in health care means that there are very limited opportunities to measure revealed preferences through observing behaviour. This is similar to other areas with non-market goods such as the environment. Health economists therefore apply survey-based, stated preference methods to value health and health care. (1, 2) Stated preference methods assume that individuals' responses to hypothetical valuation tasks are based on complete, stable and rational preferences that are consistent with the axioms of utility theory. Preference estimates are only useful to policymakers if they are generalisable beyond the moment when they are collected. Tests of preference stability over substantial periods of time are needed to understand the 'shelf-life' of preference estimates(3). However, most previous contingent valuation (CV) studies of test-retest reliability in health care have short time periods of less than 6 weeks (4-8). Two exceptions are Thompson et al (9) and Settumba et al (10) who test reliability over 12 and 10 months respectively. Two studies explore the stability of Willingness To Pay (WTP) in discrete choice experiment (DCE) tasks (11, 12). Skjoldborg et al (11) find no significant differences in marginal WTP for attributes of rheumatoid arthritis treatment elicited at three time points up to four months apart. Price et al (12) find no differences in marginal WTP for attributes describing the mortality and morbidity reduction from improved tap water quality elicited from different samples at two time points eight years apart. Preference stability tests of WTP conducted over longer time periods are needed to understand the 'shelf-life' of preference estimates.

WTP is expected to change under certain circumstances (13). WTP elicited in response to the same stated preference tasks across time points should be unchanged if the determinants of WTP (such as income, price of complements and substitutes, inflation etc) are unchanged. However, WTP is expected to change if the determinants of preferences change. For example,

a large reduction in income should reduce an individual's WTP. WTP is also expected to change over time if the relationship between determinants and WTP change (3). While a short duration between valuation surveys reduces the likelihood that determinants or the relationship between determinants and WTP change between waves, such changes are still possible (7, 14). Any unexpected changes in WTP values in health care may suggest that the method itself is not able to elicit robust stable values in health care. However, unexpected changes in WTP values may also occur if individuals are unfamiliar with the good and have incomplete preferences. Individuals may become more familiar with the good over time and this can influence the stability of their WTP values. To better understand whether the elicitation method itself can produce stable estimates in health care, we examine the stability of WTP for a familiar health care good (scale and polish). We do this in a sample of regular attenders at UK NHS dental practices who have experience with the good that is being valued.

We use a unique dataset in which WTP was elicited at four time points over a relatively long period (baseline and at annual intervals for three years). We compare the average WTP as well as the WTP function over time. The data were collected as part of a Randomised Controlled Trial (RCT) which also allows us to examine whether context matters. Participants were randomly offered different frequencies (intensities) of scale polish (no scale and polish, one scale and polish per year, two scale and polishes per year). Being allocated to different treatment intensities should not affect WTP if individuals' responses to hypothetical valuation tasks are based on complete, stable and rational preferences that are consistent with the axioms of utility theory. However, it can be hypothesised that WTP may be affected in at least two ways. First, being allocated to the no scale and polish arm may lead to disappointment and what has been termed resentful demoralisation (16, 17). Trial participants who do not receive their preferred treatment allocation may be less motivated and may not report accurately during

follow-up. This may lead to instability in WTP values. Their reported WTP before allocation to a treatment arm may therefore be different from their reported WTP after allocation. Secondly, the differences in intensity across arms can lead to differences in experience and this may have an impact on stated WTP. Utility theory assumes that individuals make decisions with full information. Unlike choices about daily essentials such as groceries, individuals seldom make decisions about healthcare goods and services. In this case, individuals may not have complete preferences for these unfamiliar goods and services (18). As individuals gain experience of the good or service they may learn about their preferences (19-22). Information and (familial) experience of the health condition has been shown to influence WTP (23-26) However, as a scale and polish is a familiar good and given that the study was part of a pragmatic trial (scale and polish was not withheld from patients requesting it and patients could also obtain additional private scale and polish treatments) we hypothesise that any differences in WTP are more likely to be caused by the allocation itself rather than differences in the frequency of service experienced.

The aim of this paper is to test the stability of WTP values over time in a familiar healthcare good over a long time period and whether treatment allocation to different treatment intensities (0, 1 or 2 treatments per year) influences WTP.

2. METHODS

2.1. The iQuaD trial

The data are from the Improving the Quality of Dentistry (IQuaD) multicentre pragmatic split-plot randomised open trial with a cluster factorial design (27). Sixty-three NHS dental practices across Scotland and North-East England were randomised to provide routine or personalised oral hygiene advice. Within these dental practices, participants were randomised to three

groups which were offered different frequencies of National Health Service (NHS) provided scale and polish (none, one per year, or two per year for three years). A scale and polish is the thorough cleaning of teeth and gums by a dentist or dental hygienist. Scaling removes hard tartar from teeth and polishing helps to clean stains off tooth surfaces. It is one of the most frequently provided dental procedures in the UK. In England, in 2019/2020, 45% of all adult courses of treatment delivered in primary care included a scale and polish as part of the treatment course (15). In line with usual practice, participants were required to contribute to the cost of their NHS dental care, unless they were exempt from paying charges. The treatments were provided by NHS dentists and hygienists.

2.2. Sample

The participants were dentate adults who were regular NHS attenders (attended for a dental check-up in the previous two years) and who did not have severe gum disease. In total, 1877 trial participants were recruited. Dental practices sent out invitation letters, a patient information sheet and baseline questionnaire (including the CV task) to potentially eligible participants. The study team obtained consent from potentially eligible participants and then collected the baseline clinical measurements and questionnaires. The baseline measurements took place between February 2012 and July 3013. All trial participants received a scale and polish at baseline, after completing the baseline questionnaire (including CV task) and before trial allocation was known. A letter was sent to all participants to inform them of their scale and polish allocation.

Participants completed a questionnaire at baseline (pre-randomisation) and at annual intervals for three years of follow-up. All questionnaires were self-completed postal questionnaires. Of

the 1873[1] trial participants, 1119 (59.7%) returned the self-complete questionnaires at all time points. The majority of these (81.2%) completed the CV question in each year (N=909). This means that complete CV data are available for 48.5% of trial participants.

2.3. The contingent valuation (CV) task

A payment card CV task was used to elicit each participant's WTP for scale and polish.. The payment card method is commonly used to elicit WTP for health care 23.

The good or service first needs to be described to respondents. A scale and polish is a familiar service for regular dental attenders. The following information was provided in the information sheet: "It is well known that dental plaque is the main cause of gum disease. Effective oral hygiene (tooth brushing and inter-dental aids) for plaque control and the removal of calculus (tartar) by your dentist or hygienist with a scale and polish are considered necessary to prevent and treat gum disease."

The CV question presented to respondents is shown in Figure 1. The same question was used at all time points. The bid levels in the payment card were chosen as follows. A lower bound of £0 was included to allow that respondents may not value the service. The upper bound of £75 was selected based on the maximum private price for scale and polish treatment across UK providers on an internet price comparison website (28). The remaining bids were selected to cover the range using an exponential scale (29). The bids were then rounded to the nearest whole pound multiple of £5. The bids £10.50 and £17.50 were added as these were the average

[1] There were four post-randomisation exclusions where randomisation took place in error.

patient co-charge in Scotland and England at the time of the study design (2012), rounded to the nearest 50p, respectively.

2.2 Analysis

We test for stability of WTP over time and the WTP function over time using regression analysis.

2.2.1 Regression method

The payment card response data provide an interval-censored signal about the WTP for a scale and polish of individual *i* in time period *t* (30). We assume that respondents' WTP (WTP_{it}) falls in the interval $b_{it,l} \leq WTP_{it} < b_{it,u}$ where $b_{it,l}$ is the highest bid amount for which respondent *i* places an X in the box and $b_{it,u}$ is the next highest bid level. The exact WTP value is not known but the interval within which the value lies is observed. Interval regression fits a linear model to interval-censored data such as the WTP data collected in this study. The coefficients from interval regression can be interpreted the same as in Ordinary Least Squares (OLS). INTREG in Stata 15.0 is used to estimate the models.

2.2.2 Comparing WTP over time

To examine stability of WTP over time relative to the baseline, we included three dummy variables (Year 1, Year 2, Year 3) in the interval regression model. Statistically significant coefficients on one of more of the dummy variables indicate instability of WTP values. We also conduct a Wald test for joint significance of all three dummy variables. Dummy variables may be jointly significant even if they are not individually statistically significant.

We would expect individuals' WTP to change over time if they experience a change in characteristics that impact on WTP. It is therefore important to include time-varying characteristics that are hypothesised to influence WTP. We included two time-varying characteristics. First, we include whether a respondent is exempt from co charges[2]. We define this as a dummy variable that takes the value of 1 if a respondent is exempt and zero otherwise. Exemption is associated with respondents' socio-economic status.[3] Information on income was not available as the data are from participant trial questionnaires which generally do not collect this type of information. However, a change in exemption status is likely to represent a large income shock which would be expected to change an individual's WTP. It is hypothesised that those who are exempt will have a lower WTP. People who are exempt may also be less familiar with paying for dental care which may have an impact on their WTP. Secondly, we include whether the individual uses an electric toothbrush (dummy variable: Electric toothbrush) or not. It is hypothesised that individuals who are willing to buy an electric toothbrush care more about their dental health than those who are not and therefore they would have higher WTP.

We used fixed effects interval regression as this allows us to test the stability of WTP while controlling for respondents' characteristics that are constant over time:

$$WTP_i = f(X_{it}, t) + i\alpha_i + \varepsilon_{it}$$

Where (X_{it}) are the time varying characteristics, individual fixed effect ($i\alpha_i$) is the individual fixed effect and ε_{it} is an error term. We repeat the analysis by RCT arm to test whether stability over time varies across the RCT arms. The analysis presented uses a balanced panel of respondents who completed the payment card task at all four time points.

[2] Exemption status is taken from routine records (administrative dental claims data, linked to trial participant characteristics as part of the study).

[3] Exempt patients are those who receive income support, universal credit, tax credit exemptions, are younger than 18 years, receiving pension credit or are pregnant or have given birth in the last 12 months

2.2.3. Hypotheses

During the data collection period (2012 to 2016) inflation was low (around 1.4%) and there were no major macroeconomic or oral health information shocks. It is therefore hypothesised that average WTP should be stable over this period. We hypothesise that being allocated to the lowest treatment intensities is associated with the largest change in WTP especially at Year 1 due to resentful demoralisation.

2.2.4. Comparing WTP function over time

An individuals' WTP may not be stable if the relationship between individual characteristics and WTP changes over time. For example, new information may become available that scale and polish is particularly important for older individuals. This means that the relationship between age and WTP for scale and polish may change and as a result mean WTP may change over time. We estimate a separate WTP function for each time point. In each case, we estimate an interval regression model as in section 2.2, except without fixed-effects. We include exemption status and use of electric toothbrush as well as several baseline covariates, namely age (dummy variables: Age 35-44, Age 45-54, Age 55-64, Age>65; omitted category Age 17-34), gender (dummy variable: Male), whether practice employs dental hygienist or not (dummy variable: Dental Hygienist) and UK country (dummy variable: England). Previous evidence suggests that some individuals base their WTP responses on the estimated cost of the service (see for example Donaldson et al (31)). Country is therefore included as co-charges vary across England and Scotland. Patient co-charges are higher in England (if patient has for example a check-up and scale and polish only) compared to Scotland (where there are no co-charges for check-ups) and it is therefore hypothesised that WTP may be higher in England if responses are influenced by actual service cost to participants. A Chow test is used to test whether the

coefficients in the baseline WTP function are statistically significantly different from the coefficients at each of the three later time points (year 1, year 2 and year 3).

2.2.5. Individual level changes

The main motivation for the paper is to test the 'shelf life' of the values at the population average to inform cost-benefit analyses (CBA). However, demonstrating stability of WTP over time at the mean level does not exclude the possibility of changes in WTP at the individual level which are cancelled out at the mean level. To explore individual level changes, we report the number of respondents with no change in bid amount chosen compared to baseline, a one interval increase in bid amount chosen (for example from £10.50 to £15), two or more intervals increase in bid amount chosen, one interval decrease in bid amount chosen and two or more intervals decrease in bid amount chosen. It could be argued that smaller changes are more likely to be due to imprecision in preferences whilst larger changes may be more likely to indicate a change in WTP. We also report the difference between the maximum and minimum bid amount chosen across all years.

2.2.6. Robustness checks

We perform robustness checks of the main analysis. First, we exclude respondents who report zero WTP from the analysis. Those respondents who report a WTP of £0 may be protest respondent (32). Due to space constraints within the trial questionnaires, we were unable to include any follow-up questions to the CV tasks to understand whether any £0 responses were protests. We therefore rerun the analysis removing all £0 responses. Secondly, the analysis is estimated using an unbalanced panel. This can provide an indication as to whether there is a selection bias due to non-response.

3. RESULTS

Table 1 shows the baseline characteristics of the sample who completed all WTP questions. The majority of the sample (68.4%) prefer to have two or more scale and polishes per year at their stated maximum WTP at baseline. 92.4% of the sample visited their NHS dentist in the last year and 61.3% had a scale and polish at their last visit. There are more females than males in the sample and most of the sample is resident in Scotland. Appendix 1 shows the baseline characteristics of the total sample and those respondents who had one or more missing WTP responses. Note that missing values on the WTP questions are mainly due to respondents not returning the full trial questionnaire and are therefore not directly related to the WTP question (see Methods section).

Table 2 shows the frequencies of the CV responses at baseline and Appendix 2 shows the frequencies of the CV responses across all time points. There are relatively few zero responses. Only 6 respondents reported a WTP of £0 at all time points. All individuals who ticked £0 then indicated that they would like to receive a scale and polish. None of the individuals in our balanced panel reported that they were willing to pay more than £75. There is approximately a normal distribution in terms of distribution of responses by bid amount apart from the gap at £17.50. This may the result of the prominence effect where respondents are more likely to choose prominent numbers such as 1, 2, 5, 10, 20 and 50 (33).

Figure 2 shows the boxplots of mean WTP by year and randomised allocation (assuming the midpoint of the interval) and Appendix 3 shows the descriptive statistics of mean WTP by year and arm. Mean WTP seems similar across time points and across arms with confidence intervals clearly overlapping suggesting that WTP is relatively stable. However, these summary statistics do not control for changes in circumstances over time.

3.1 Comparing WTP over time

Table 3 shows the results of the fixed effects interval regression model for the balanced panel (Appendix 4 shows the full regression results). The first model includes the full sample and includes the time dummies and covariates. The time dummies are not statistically significant indicating that mean WTP is stable over time for the whole sample. The model is then estimated by RCT arm to examine whether the stability of WTP varies across trial arms. WTP is relatively stable over time in all trial arms suggesting that allocating to different treatment intensities does not have a significant impact on WTP. Only three of the nine coefficients are statistically significant at the 5% level. WTP is £1.36 lower at the end of year one compared to baseline in the one scale and polish arm. Given a mean WTP of £20.53 at baseline, this means that WTP is 6.6% lower. WTP is £0.94 higher at the end of year 2 and £1.09 higher at the end of year 3 compared to baseline in the two scale and polish arm. Given a mean WTP of £18.96 at baseline, this means that WTP is 5.0% higher at the end of year 2 and 5.6% at the end of year 3. It is interesting to note that the coefficients are negative for the lower intensity RCT arms (No S&P and One S&P) whilst they are positive for the two scale and polish arm.

3.2. Comparing WTP function over time

Table 4 shows the regression results of the determinants of WTP at each time point. Similar characteristics are associated with WTP across the different time points. Being exempt from dental charges and being registered with larger practices is associated with lower WTP at each time point. Using an electric toothbrush and being resident in England is associated with higher WTP at each time point. The association between age and WTP varies across the time points with no statistically significant association at baseline and year 3 but a significant association

at year 1 and 2. The Chow test results show that we cannot reject the null hypotheses of equal coefficients in the baseline regression model and at the three later time points.

3.3. Individual level changes

Table 5 shows the changes in bid amount chosen between years 1, 2 and 3 and baseline. A considerable proportion of respondents change their bid amount. This includes around 15% who decrease their bid amount by two or more intervals and between 12.9% and 15.2% who increase their bid amount by two or more intervals. The proportion increasing and decreasing their bids are roughly similar which explains why the mean is stable even though there are many individual changes. Table 5 also shows the difference between maximum and minimum bid amount chosen across all years. In total, 142 respondents (15.6%) choose the same bid amount throughout. Around 57% of respondents change their bid amount by one or two intervals across all years. A smaller proportion (27.3%) change their bid amount by three or more intervals across all years. Appendix 5 shows the bid amounts chosen at each time point by chosen baseline bid amount. The majority of respondents are clustered on or just beside the diagonal line (shaded in grey) which represents the same bid amount chosen at both time points. However, for several other respondents the difference between amounts chosen is considerable.

3.4 Robustness checks

The robustness checks re-estimated the fixed effects interval regression model of WTP for different samples. The results are reported in Table 6 (full regression results are reported in Appendix 6). The results are generally similar across the different specifications. WTP is £0.45 lower at the end of year one compared to baseline when using the unbalanced panel.

4. DISCUSSION

The aim of this paper was to test the stability of WTP values over time in a familiar healthcare good over a long time period. We find that both mean WTP and the WTP function were stable over a 3-year time period (measured at four time points). We also examined whether treatment allocation to different treatment intensities (0, 1 or 2 treatments per year) influenced WTP. Any effects are likely to be due to the allocation itself rather than differences in experience given that scale and polish is a familiar good and participants were part of a pragmatic trial (scale and polish was not withheld from patients requesting it). The findings suggest that preference estimates are generalisable beyond the moment when they are collected and have a reasonable 'shelf-life'. Our findings are in line with previous studies which have typically used shorter time periods to test stability (4-12). We find that random allocation in an RCT to different treatment intensities did not have a consistent impact on WTP. It is interesting to note though that WTP is lower compared to baseline in the no scale and polish arm and the one scale and polish arm which is line with the resentful demoralisation hypothesis (16.17). However, only one of these effects is statistically significant.

Whilst mean WTP was generally stable there were a substantial number of individual level changes. Some of these changes may be due to imprecise preferences. However, a proportion of respondents changed their bid amount by two or more intervals. The proportion of respondents increasing and decreasing their bid amounts was similar which explains why the mean WTP was stable despite a substantial number of individual level changes

Stable WTP values do not necessarily imply that the estimates reflect individuals' true preferences. This requires external validity tests which are beyond the scope of this paper. It could be argued that the use of heuristics may have resulted in stable WTP estimates. Whilst

it is likely that heuristics have been used by at least some respondents we think it is unlikely to be the main reason for stable WTP values. Firstly, there is unlikely to be a consistent relationship between individual characteristics and WTP if the majority of the sample used heuristics. Secondly, individual level WTP values would also be expected to be stable which was not the case in our study. We did find some possible evidence of a prominence effect and cost-based responses (WTP being higher in the region with higher user charges for dental care) suggesting that the external validity of the WTP estimates should be examined in future research.

Any unexpected changes in WTP values in health care may be due to the elicitation method itself and/or the unfamiliarity with the good. Our study tested the stability of WTP values in a familiar health care good to test whether the method itself can produce stable estimates. Individuals in our sample have experienced scale and polish (and therefore more likely to have complete preferences) and unlike other NHS services many patients must pay a co-charge and are therefore used to considering their WTP for this service. It is important to test whether stable WTP values can be estimated for unfamiliar or less familiar health care goods.

There are several limitations. First, the study was conducted using RCT participants. Individuals who consent to take part in a RCT may be atypical and may be more engaged and more likely to complete questionnaires in a consistent manner. Second, the WTP questions were asked as part of a relatively large self-complete questionnaire. The WTP question had to be short and it was not possible to identify protest responses or include techniques that have been shown to improve response validity such as a cheap talk script (34-36). However, this may be less important when the service is familiar, and most participants are used to paying. Third, the payment card CV method was used rather than the dichotomous choice method

which is the method recommended in the NOAA report (18). The payment card CV method and open-ended methods more generally are commonly used in health. In a recent review of the determinants of WTP for health services using the contingent valuation method, around 25% of papers used the payment card CV method and 37% used open-ended methods more generally (37). It is therefore important to test stability using the payment card CV method. However, results from the payment card CV method cannot necessarily be generalised to other elicitation formats as each format has its own bases and limitations. The payment card method has been shown to have a number of biases including range bias. It should be noted that in this study these biases associated are likely to be constant across arms and time. Future research should examine stability in WTP for health care using other elicitation methods. Also, stability was tested using data from a RCT. It is important to examine stability in other samples including a general population sample. Fourth, only around half of the sample had a complete set of WTP responses. The missing WTP values were mainly due to survey non-response rather than item non-response to the CV question. Fifth, information on income was not available. Income is an important determinant of WTP and the analysis should therefore ideally control for changes in income. However, the analysis did include a proxy for income (exemption from dental charges).

Conclusion

We find that WTP values for scale and polish elicited using a payment card contingent valuation question were stable over time. This suggests that WTP values are transferable and can be used in cost benefit analyses in time periods other than the one in which the WTP values were elicited. Future research should explore stability of WTP values for other, less familiar, health care services, in other populations including a general population sample and using different elicitation methods such as the dichotomous choice CV method and DCEs.

Acknowledgements

See Title Page for Acknowledgements.

References

1. Johnston RJ, Boyle KJ, Vic Adamowicz W, Bennett J, Brouwer R, Ann Cameron T, Michael Hanemann W, Hanley N, Ryan M, Scarpa R, Tourangeau R, Vossler CA. Contemporary guidance for stated preference studies. Journal of the Association of Environmental and Resource Economists. 2017;4(2):319-405. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85076196670&doi=10.1086%2f691697&partnerID=40&md5=8aac425082aecbf63c05bf6c369a2f99

2. Soekhai V, de Bekker-Grob EW, Ellis AR, Vass CM. Discrete choice experiments in health economics: Past, present and future. Pharmacoeconomics. 2019;37(2):201-26. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85056203303&doi=10.1007%2fs40273-018-0734-2&partnerID=40&md5=db89bdd2470cdd988d28f422a638c811

3. Loomis JB. Test-retest reliability of the contingent valuation method: A comparison of general population and visitor responses. American Journal of Agricultural Economics. 1989;71(1):76-84. Available from: https://doi.org/10.2307/1241776

4. Martín-Fernández J, del Cura-González MI, Rodríguez-Martínez G, Ariza-Cardiel G, Zamora J, Gómez-Gascón T, Polentinos-Castro E, Pérez-Rivas FJ, Domínguez-Bidagor J, Beamud-Lagos M, Tello-Bernabé ME, Conde-López JF, Aguado-Arroyo Ó, Sanz-Bayona M, Gil-Lacruz A. Economic valuation of health care services in public health systems: A study about willingness to pay (WTP) for nursing consultations. PLoS ONE. 2013;8(4) Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84876568666&doi=10.1371%2fjournal.pone.0062840&partnerID=40&md5=bcdcc34e8e4b3016751b60bcfacf569a

5. Smith RD. The reliability of willingness to pay for changes in health status. Applied Health Economics and Health Policy. 2004;3(1):35-8. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-22544464469&doi=10.2165%2f00148365-200403010-00007&partnerID=40&md5=6441492dd8291b93027de33427e26cd1

6. Shiell A, Hawe P. Test-retest reliability of willingness to pay. Eur J Health Econ. 2006 Sep;7(3):176-81.

7. Onwujekwe O, Fox-Rushby J, Hanson K. Inter-rater and test-retest reliability of three contingent valuation question formats in south-east nigeria. Health Econ. 2005 May;14(5):529-36.

8. O'Brien B, Viramontes JL. Willingness to pay: A valid and reliable measure of health state preference? Medical Decision Making. 1994;14(3):289-97. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-0028290332&doi=10.1177%2f0272989X9401400311&partnerID=40&md5=860b20683b9c65da36d8bce67f894bb2

9. Thompson MS, Read JL, Liang M. Feasibility of willingness-to-pay measurement in chronic arthritis. Med Decis Making. 1984;4(2):195-215.

10. Settumba SN, Shanahan M, Botha W, Ramli MZ, Chambers GM. Reliability and validity of the contingent valuation method for estimating willingness to pay: A case of in vitro fertilisation. Applied Health Economics and Health Policy. 2019;17(1):103-10. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85055338645&doi=10.1007%2fs40258-018-0433-3&partnerID=40&md5=2cec206e860477b7da81017beeacd674

11. Skjoldborg US, Lauridsen J, Junker P. Reliability of the discrete choice experiment at the input and output level in patients with rheumatoid arthritis. Value in Health. 2009;12(1):153-8. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-58849149512&doi=10.1111%2fj.1524-4733.2008.00402.x&partnerID=40&md5=8933721402082a0efa8e125452e73857

12. Price J, Dupont D, Adamowicz W. As time goes by: Examination of temporal stability across stated preference question formats. Environmental and Resource Economics. 2017;68(3):643-62. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84973621852&doi=10.1007%2fs10640-016-0039-x&partnerID=40&md5=4cf07b0abf98027eee2883dc74d2e515

13. Carson RT, Hanemann WM, Kopp RJ, Krosnick JA, Mitchell RC, Presser S, Rudd PA, Smith VK, Conaway M, Martin K. Temporal reliability of estimates from contingent valuation. Land Econ [Internet]. 1997;73(2):151-63. Available from: http://www.jstor.org/stable/3147279

14. Dong H, Kouyate B, Cairns J, Mugisha F, Sauerborn R. Willingness-to-pay for community-based insurance in burkina faso. Health Econ. 2003 Oct;12(10):849-62.

15. National Health Service. NHS dental statistics for England - 2019-20 annual report. https://digital.nhs.uk/data-and-information/publications/statistical/nhs-dental-statistics/2019-20-annual-report: 2020

16. Cook, Thomas D.,, Campbell,Donald T.,. Quasi-experimentation : design & analysis issues for field settings. Boston: Houghton Mifflin; 1979.

17. Torgerson D, Sibbald B. Understanding controlled trials: What is a patient preference trial? BMJ. 1998;316(7128):360. Available from: http://www.bmj.com/content/316/7128/360.abstract

18. Arrow K, Solow R, Portney PR, Leamer EE, Radner R, Schuman H. Report of the NOAA panel on contingent valuation. Fed Regist. 1993;58(10):4601-14.

19. Braga J, Starmer C. Preference anomalies, preference elicitation and the discovered preference hypothesis. Environmental and Resource Economics. 2005;32(1):55-89. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-24044536650&doi=10.1007%2fs10640-005-6028-0&partnerID=40&md5=5b9d6b1db897a4c25880bd822fa627a8

20. Cameron TA, Englin J. Respondent experience and contingent valuation of environmental goods. J Environ Econ Manage. 1997;33(3):296-313. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-0031193460&doi=10.1006%2fjeem.1997.0995&partnerID=40&md5=7eb261186c601393942057f70bac2ccf

21. Matthews Y, Scarpa R, Marsh D. Stability of willingness-to-pay for coastal management: A choice experiment across three time periods. Ecol Econ. 2017;138:64-73. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85016309493&doi=10.1016%2fj.ecolecon.2017.03.031&partnerID=40&md5=6f19053282655c7789477ac00f0d8694

22. Ryan M, San Miguel F. Revisiting the axiom of completeness in health care. Health Econ. 2003;12(4):295-307. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-0037387822&doi=10.1002%2fhec.730&partnerID=40&md5=9a8e4a695bbf3c3ab75cefbb95f2d85b

23. Hirth RA, Bloom BS, Chernew ME, Fendrick AM. Willingness to pay for diagnostic certainty: Comparing patients, physicians, and managed care executives. J Gen Intern Med [Internet]. 1999 Mar;14(3):193-5.

24. Wagner TH, Hu T, Dueñas GV, Kaplan CP, Nguyen BH, Pasick RJ. Does willingness to pay vary by race/ethnicity? an analysis using mammography among low-income women. Health Policy [Internet]. 2001;58(3):275-88. Available from: https://www.sciencedirect.com/science/article/pii/S0168851001001774

25. Neumann PJ, Johannesson M. The willingness to pay for in vitro fertilization: A pilot study using contingent valuation. Med Care [Internet]. 1994;32(7) Available from: https://journals.lww.com/lww-medicalcare/fulltext/1994/07000/the_willingness_to_pay_for_in_vitro_fertilization_.3.aspx

26. Wagner TH, Hu T, Dueñas GV, Kaplan CP, Nguyen BH, Pasick RJ. Does willingness to pay vary by race/ethnicity? an analysis using mammography among low-income women. Health Policy [Internet]. 2001;58(3):275-88. Available from: https://www.sciencedirect.com/science/article/pii/S0168851001001774

27. Ramsay CR, Clarkson JE, Duncan A, Lamont TJ, Heasman PA, Boyers D, Goulao B, Bonetti D, Bruce R, Gouick J, Heasman L, Lovelock-Hempleman LA, Macpherson LE, McCracken G, I., McDonald AM, McLaren-Neil F, Mitchell FE, Norrie JDT, van der Pol M, Sim K, Steele JG, Sharp A, Watt G, Worthington H, V., Young L. Improving the quality of dentistry (IQuaD): A cluster factorial randomised controlled trial comparing the effectiveness

and cost-benefit of oral hygiene advice and/or periodontal instrumentation with routine care for the prevention and management of periodontal disease in dentate adults attending dental primary care. Health Technol Assess. 2018 JUL;22(38):1.

28. What price£?. Tooth Scale and Polish [Internet]. cited 03 2021]. Available from: http://www.whatprice.co.uk/prices/dentist/tooth-scale-and-polish.html#ixzz1xTI3LIkY

29. Rowe RD, Schulze WD, Breffle WS. A test for payment card biases. J Environ Econ Manage. 1996;31(2):178-85. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-0030236459&doi=10.1006%2fjeem.1996.0039&partnerID=40&md5=a5bd0e7124638fae9e1d14a8647f4c25

30. Cameron TA, Huppert DD. OLS versus ML estimation of non-market resource values with payment card interval data. J Environ Econ Manage [Internet]. 1989;17(3):230-46. Available from: https://www.sciencedirect.com/science/article/pii/0095069689900181

31. Donaldson C, Shackley P, Abdalla M. Using willingness to pay to value close substitutes: Carrier screening for cystic fibrosis revisited. Health Econ [Internet]. 1997;6(2):145-59. Available from: https://doi.org/10.1002/(SICI)1099-1050(199703)6:23.0.CO 2-N

32. Pennington M, Gomes M, Donaldson C. Handling protest responses in contingent valuation surveys. Medical Decision Making. 2017;37(6):623-34. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85022019044&doi=10.1177%2f0272989X17691771&partnerID=40&md5=2633d62256b43a9c4a2b1835da336e8f

33. Whynes DK, Philips Z, Frew E. Think of a number… any number? Health Econ [Internet]. 2005;14(11):1191-5. Available from: https://doi.org/10.1002/hec.1001

34. Murphy JJ, Allen PG, Stevens TH, Weatherhead D. A meta-analysis of hypothetical bias in stated preference valuation. Environmental and Resource Economics. 2005;30(3):313-25. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-17644395703&doi=10.1007%2fs10640-004-3332-z&partnerID=40&md5=a8bd76766d44d8ad79b869cbac80226f

35. Cummings RG, Taylor LO. Unbiased value estimates for environmental goods: A cheap talk design for the contingent valuation method. Am Econ Rev. 1999;89(3):649-65. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-0000301020&doi=10.1257%2faer.89.3.649&partnerID=40&md5=b84cb46d2aed7cec896e63fdc7a2b09c

36. Özdemir S, Johnson FR, Hauber AB. Hypothetical bias, cheap talk, and stated willingness to pay for health care. J Health Econ. 2009;28(4):894-901. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-70349257530&doi=10.1016%2fj.jhealeco.2009.04.004&partnerID=40&md5=f513f433f58cdf9fab545905accd1432

37. Steigenberger C, Flatscher-Thoeni M, Siebert U, Leiter AM. Determinants of willingness to pay for health services: A systematic review of contingent valuation studies. The European

Journal of Health Economics. 2022;23(9):1455-82. Available from: https://doi.org/10.1007/s10198-022-01437-x

TABLE 1. Characteristics of participants (complete cases – N=909)

	N	%
Age		
<35	113	12.4
35-44	133	14.6
45-54	217	23.9
55-64	239	26.3
>=65	207	22.8
Gender		
Female	586	64.5
Male	323	35.5
Exempt from dental charges		
Non-Exempt	774	85.1
Exempt	135	14.9
Uses electric brush		
No	584	64.2
Yes	311	34.2
Missing	14	1.5
Practice employs a hygienist		
No	214	23.5
Yes	695	76.5
Country		
Scotland	642	70.6
England	267	29.4
Date of last visit to dentist		
<1 year ago	840	92.4
1-2 years ago	60	6.6
>2 years ago	5	0.6
Missing	4	0.4
How often prefer to have scale and polish		
>more than 2 a year	192	21.1
2 a year	430	47.3
Once a year	208	22.9
Once every 2 years	32	3.5
never	19	2.1
Missing	28	3.1
Scale and polish at last visit		
Yes	557	61.3
No	339	36.3
Missing	22	2.4

TABLE 2. Contingent valuation task responses at baseline (N=909)

Bid amount (£)	Total N	%
0	21	2.3
1	1	0.1
5	71	7.8
10.5	181	19.9
15	202	22.2
17.5	55	6.1
20	255	28.1
30	102	11.2
50	19	2.1
75	2	0.2

TABLE 3. Fixed effects interval regression of willingness to pay

	Full sample		No S&P		One S&P		Two S&P	
	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value
Year 1	-0.334	(0.20)	-0.197	(0.66)	-1.355***	(<0.01)	0.508	(0.26)
Year 2	0.0938	(0.72)	-0.0419	(0.93)	-0.651	(0.15)	0.935**	(0.04)
Year 3	0.218	(0.40)	-0.308	(0.50)	-0.0985	(0.83)	1.087**	(0.02)
Constant	19.745***	(<0.01)	19.765***	(<0.01)	20.277***	(<0.01)	19.100***	(<0.01)
Observations	3445		1196		1080		1172	
Individuals	862		299		270		293	
McFadden R^2	0.0009		0.0002		0.0043		0.0034	

* p<0.10; ** p<0.05; *** p<0.01. Wald test for joint significance of Year 1, Year 2, Year 3 – Chi^2(p-value): 4.93 (0.18); 0.60(0.90); 11.33(0.01); 7.14 (0.07).

TABLE 4. Determinants of willingness to pay at each timepoint

	Baseline		**Year 1**		**Year 2**		**Year 3**	
	Coefficient	t-statistic	Coefficient	t-statistic	Coefficient	t-statistic	Coefficient	t-statistic
One S&P	0.749	(0.87)	-0.530	(-0.62)	0.100	(0.11)	0.971	(1.11)
Two S&P	-0.566	(-0.67)	0.257	(0.31)	0.545	(0.60)	1.145	(1.34)
Personalised advice[1]	-0.0943	(-0.13)	0.0126	(0.02)	1.117	(1.48)	0.583	(0.81)
Aged between 35 and 44	-1.193	(-0.89)	0.990	(0.75)	1.593	(1.11)	1.702	(1.24)
Aged between 45 and 54	0.520	(0.42)	2.738**	(2.28)	2.281*	(1.74)	1.471	(1.18)
Aged between 55 and 64	0.742	(0.60)	2.490**	(2.07)	2.483*	(1.89)	1.939	(1.55)
Aged 65 and over	-0.936	(-0.74)	0.886	(0.72)	2.436*	(1.81)	1.728	(1.36)
Male	0.0961	(0.13)	0.482	(0.66)	0.296	(0.37)	0.608	(0.80)
Exempt from dental charges	-3.497***	(-3.65)	-3.933***	(-4.04)	-4.199***	(-3.94)	-4.736***	(-4.62)
Uses electric brush	2.498***	(3.40)	2.344***	(3.28)	1.679**	(2.18)	2.651***	(3.65)
Practice employs a hygienist	0.512	(0.59)	0.657	(0.76)	-0.101	(-0.11)	0.161	(0.18)
England	2.416***	(2.89)	1.569*	(1.90)	2.919***	(3.24)	1.865**	(2.19)
Constant	18.75***	(10.88)	16.63***	(9.86)	15.45***	(8.37)	15.92***	(9.09)
Observations	862		862		862		862	
McFadden R^2	0.014		0.013		0.012		0.015	
Chow test (vs baseline) Chi^2 (p-value)			6.54 (0.92)		8.78 (0.79)		8.55 (0.81)	

* p<0.10; ** p<0.05; *** p<0.01; [1] this variable controls for the randomisation of practices to providing routine or personalised oral hygiene advic

TABLE 5. Individual level changes

	Year 1 vs baseline		Year 2 vs baseline		Year 3 vs baseline	
Change in bid amount chosen	N	%	N	%	N	%
No change	387	42.6	363	39.9	348	38.3
1 decrease	134	14.7	122	13.4	112	12.3
2 or more decreases	133	14.6	139	15.3	135	14.9
1 increase	138	15.2	147	16.2	183	20.1
2 or more increases	117	12.9	138	15.2	131	14.4

Interval difference between highest and lowest bid amount chosen across all years		
	N	%
0	142	15.6
1	297	32.7
2	222	24.4
3	158	17.4
4	67	7.4
5	9	1.0
6	12	1.3
7	2	0.2

TABLE 6. Robustness checks

	Excluding zeros		**Unbalanced panel**	
	Coefficient	p-value	Coefficient	p-value
Year 1	-0.229	(0.36)	-0.454**	(0.03)
Year 2	0.230	(0.36)	-0.0133	(0.95)
Year 3	0.258	(0.30)	0.217	(0.31)
Constant	19.963***	(<0.01)	20.321***	(<0.01)
Observations	3365		5255	
Individuals	856		1743	
McFadden R^2	0.0010		0.04382	

* p<0.10; ** p<0.05; *** p<0.01. . Wald test for joint significance of Year 1, Year 2, Year 3 – Chi^2(p-value): 5.00 (0.17); 10.60(0.01).

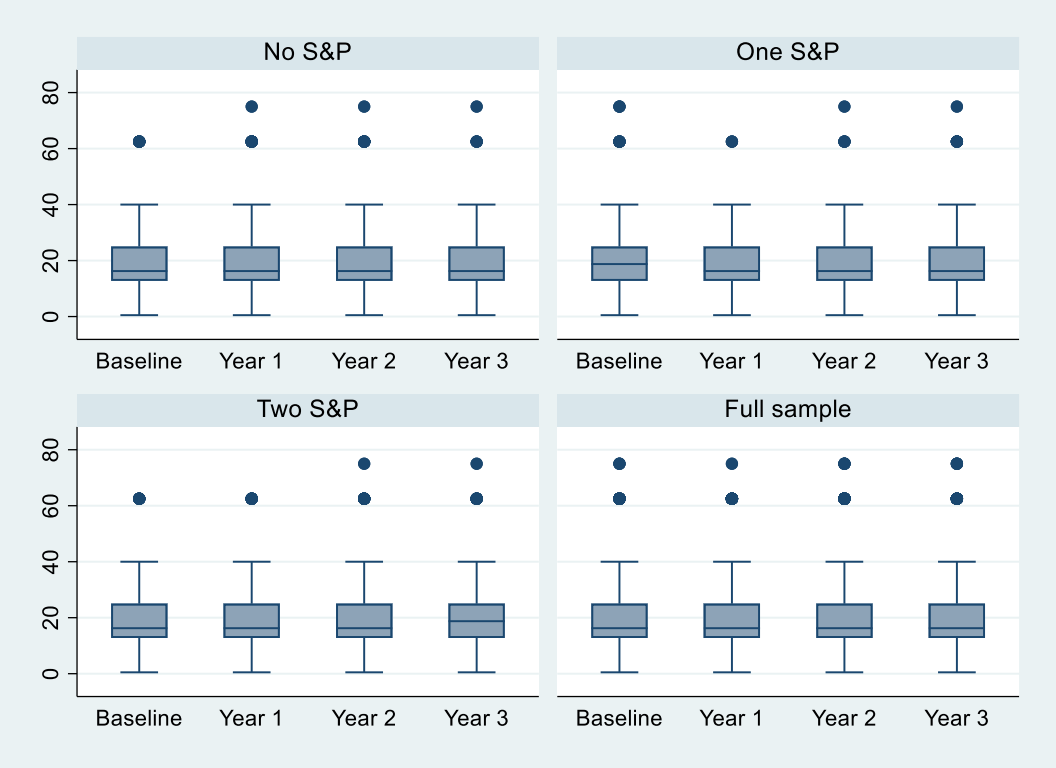
FIGURE 1: Payment card CV question included in all waves

We would like to find out how much you value scale and polish. What is the maximum amount of money that you would be willing to pay out of pocket for a scale and polish? *Please place an 'X' in the appropriate box.*

£0		£10.50		£20		£75	
£1		£15		£30			
£5		£17.50		£50			

If more than £75, please specify how much you would be willing to pay

FIGURE 2. Boxplots of WTP (midpoint) by year and arm



APPENDIX 1. Characteristics of participants and complete cases

	(Some) missing		**Complete WTP**		
	N	%	N	%	χ^2 (p-value)*
Age					220.87 (0.00)
<35	345	35.8	113	12.4	
35-44	205	21.3	133	14.6	
45-54	206	21.4	217	23.9	
55-64	99	10.3	239	26.3	
>=65	108	11.2	207	22.8	
Gender					0.21 (0.65)
Female	611	63.4	586	64.5	
Male	352	36.6	323	35.5	
Exempt from dental charges					72.79 (0.00)
Non-exempt	659	68.4	774	85.1	
Exempt	304	31.6	135	14.9	
Uses electric brush					21.01 (0.00)
No	656	68.0	584	64.2	
Yes	265	27.5	311	34.2	
Missing	43	4.5	14	1.5	
Practice employs a hygienist					1.69 (0.19)
No	252	26.1	214	23.5	
Yes	712	73.9	695	76.5	
Country					1.22 (0.27)
Scotland	703	72.9	642	70.6	
England	261	27.1	267	29.4	

* Pearson Chi -square test for differences between some missing and complete WTP samples.

APPENDIX 2. Frequencies of willingness to pay responses across all time points and by RCT arm

Bid amount		**No S&P**		**One S&P**		**Two S&P**		**Total**
(£)	N	%	N	%	N	%	N	%
Baseline								
0	8	2.8	6	2.1	7	2.3	21	2.3
1	1	0.4	0	0.0	0	0.0	1	0.1
5	20	7.0	22	7.7	29	9.6	71	7.8
10.5	64	22.5	45	15.8	72	23.8	181	19.9
15	79	27.8	60	21.1	63	20.8	202	22.2
17.5	20	7.0	22	7.7	13	4.3	55	6.1
20	86	30.3	94	33.1	75	24.8	255	28.1
30	36	12.7	28	9.9	38	12.5	102	11.2
50	8	2.8	5	1.8	6	2.0	19	2.1
75	0	0.0	2	0.7	0	0.0	2	0.2
Year 1								
0	9	3.2	6	2.1	7	2.3	22	2.4
5	19	6.7	19	6.7	20	6.6	58	6.4
10.5	81	28.5	63	22.2	81	26.7	225	24.8
15	60	21.1	76	26.8	51	16.8	187	20.6
17.5	24	8.5	7	2.5	12	4.0	43	4.7
20	89	31.3	80	28.2	89	29.4	258	28.4
30	30	10.6	31	10.9	34	11.2	95	10.5
50	9	3.2	2	0.7	9	3.0	20	2.2
75	1	0.4	0	0.0	0	0.0	1	0.1
Year 2								
0	6	2.1	8	2.8	10	3.3	24	2.6
1	0	0.0	1	0.4	0	0.0	1	0.1
5	27	9.5	12	4.2	17	5.6	56	6.2
10.5	71	25.0	55	19.4	69	22.8	195	21.5
15	70	24.6	75	26.4	59	19.5	204	22.4
17.5	17	6.0	12	4.2	15	5.0	44	4.8
20	87	30.6	79	27.8	82	27.1	248	27.3
30	34	12.0	38	13.4	40	13.2	112	12.3
50	9	3.2	3	1.1	10	3.3	22	2.4
75	1	0.4	1	0.4	1	0.3	3	0.3
Year 3								
0	7	2.5	3	1.1	10	3.3	20	2.2
1	0	0.0	1	0.4	0	0.0	1	0.1
5	24	8.5	14	4.9	15	5.0	53	5.8
10.5	70	24.6	54	19.0	60	19.8	184	20.2
15	75	26.4	75	26.4	66	21.8	216	23.8
17.5	23	8.1	17	6.0	20	6.6	60	6.6
20	74	26.1	79	27.8	79	26.1	232	25.5

30	45	15.8	35	12.3	42	13.9	122	13.4
50	3	1.1	5	1.8	9	3.0	17	1.9
75	1	0.4	1	0.4	1	0.3	3	0.3
100	0	0.0	0	0.0	1	0.3	1	0.1

APPENDIX 3. Summary statistics of the willingness to pay values by year and arm (using the midpoint of the bid interval)

	Baseline	**Year1**	**Year2**	**Year3**
Mean	20.9	20.6	21.3	21.3
Median	16.3	16.3	16.3	16.3
Sd	11.4	11.3	12.0	11.6
Minimum	0	0	0	0
Maximum	75	75	75	75
N	909	909	909	909
No scale and Polish				
Mean	20.9	20.7	20.9	20.6
Median	16.3	16.3	16.3	16.3
Sd	11.4	11.8	12.0	11.1
Minimum	0	0	0	0
Maximum	62.5	75	75	75
N	322	322	322	322
1 Scale and Polish				
Mean	21.4	20.0	21.1	21.5
Median	18.8	16.3	16.3	16.3
Sd	11.4	9.9	11.0	11.1
Minimum	0	0	0	0
Maximum	75	62.5	75	75
N	284	284	284	284
2 Scale and Polish				
Mean	20.4	21.1	21.8	21.9
Median	16.3	16.3	16.3	18.8
Sd	11.5	11.9	12.8	12.5
Minimum	0	0	0	0
Maximum	62.5	62.5	75	75
N	303	303	303	303

APPENDIX 4. Full interval regression results

	Full sample		No S&P		One S&P		Two S&P	
	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value
Year 1	-0.334	(0.20)	-0.197	(0.66)	-1.355***	(<0.01)	0.508	(0.26)
Year 2	0.0938	(0.72)	-0.0419	(0.93)	-0.651	(0.15)	0.935**	(0.04)
Year 3	0.218	(0.40)	-0.308	(0.50)	-0.0985	(0.83)	1.087**	(0.02)
Exempt from dental charges	-0.160	(0.83)	-0.0524	(0.97)	0.204	(0.88)	-0.300	(0.80)
Uses electric brush	0.771*	(0.08)	0.358	(0.63)	0.936	(0.24)	1.202	(0.12)
Constant	19.745***	(<0.01)	19.765***	(<0.01)	20.277***	(<0.01)	19.100***	(<0.01)
Observations	3445		1196		1080		1172	
Individuals	862		299		270		293	
McFadden R^2	0.0009		0.0002		0.0043		0.0034	

* p<0.10; ** p<0.05; *** p<0.01

APPENDIX 5. Chosen bid amount at year 1, 2 and 3 by baseline bid amount (N)

Bid amount	**Year 1**									
(in £)	0	1	5	10.5	15	17.5	20	30	50	75
Baseline										
0	10	0	2	5	2	0	2	0	0	0
1	0	0	1	0	0	0	0	0	0	0
5	1	0	24	29	10	2	5	0	0	0
10.5	5	0	15	97	33	6	21	3	1	0
15	3	0	10	56	72	6	51	3	1	0
17.5	1	0	2	3	12	14	21	1	1	0
20	2	0	4	29	49	13	119	39	0	0
30	0	0	0	5	9	2	32	44	9	1
50	0	0	0	1	0	0	6	5	7	0
75	0	0	0	0	0	0	1	0	1	0
	Year 2									
	0	1	5	10.5	15	17.5	20	30	50	75
Baseline										
0	9	0	1	5	4	0	1	1	0	0
1	0	0	1	0	0	0	0	0	0	0
5	4	1	22	27	8	0	6	2	1	0
10.5	6	0	16	83	44	5	20	7	0	0
15	2	0	9	46	72	8	54	10	1	0
17.5	0	0	1	5	12	17	14	6	0	0
20	3	0	5	24	53	10	114	42	4	0
30	0	0	0	4	10	2	33	41	10	2
50	0	0	1	1	1	2	5	3	5	1
75	0	0	0	0	0	0	1	0	1	0
	Year 3									
(in £)	0	1	5	10.5	15	17.5	20	30	50	75
Baseline										
0	6	0	5	4	4	0	2	0	0	0
1	0	0	1	0	0	0	0	0	0	0
5	3	1	20	31	12	0	4	0	0	0
10.5	5	0	12	72	55	10	22	5	0	0
15	2	0	8	40	75	19	43	13	2	0
17.5	0	0	1	9	9	15	20	1	0	0
20	3	0	3	23	53	15	107	48	3	0
30	1	0	2	4	8	1	28	49	8	1
50	0	0	1	1	0	0	7	6	3	1
75	0	0	0	0	0	0	0	0	1	1

APPENDIX 6. Robustness checks

	Excluding zero		Unbalanced panel	
	Coefficient	p-value	Coefficient	p-value
Year 1	-0.229	(0.36)	-0.454**	(0.03)
Year 2	0.230	(0.36)	-0.0133	(0.95)
Year 3	0.258	(0.30)	0.217	(0.31)
Exempt from dental charges	-0.104	(0.88)	-0.312	(0.56)
Uses electric brush	0.706*	(0.09)	0.676*	(0.06)
Constant	19.963***	(<0.01)	20.321***	(<0.01)
Observations	3365		5255	
Individuals	856		1743	
McFadden R^2	0.0010		0.0438	

* p<0.10; ** p<0.05; *** p<0.01;