

**Work in progress. Please do not quote or cite**

## **Testing the External Validity of Discrete Choice Experiments: Evidence from a Chlamydia Screening study**

Verity Watson\*, Mandy Ryan and Emma Watson

A major criticism of stated preference valuation methods is that the results are based on respondents making hypothetical decisions. Opponents of the methods state that these are not what respondents would do if faced with the same decision in real life. This study tested the external validity of the choice experiment stated preference approach. Patients at a family planning clinic were asked to complete a discrete choice questionnaire, concerned with Chlamydia screening, before their consultation. The questionnaire included 16 screening options and, for each option, subjects were asked if they would be prepared to go through with screening. Possible responses were 'yes' or 'no'. Included amongst these options was the 'actual choice', which they were offered on completion of the questionnaire. There was no significant difference in the proportions who said they would take the test in the hypothetical and real contexts, providing support for the external validity of the DCE approach.

The study also explored the effect of using visual aids versus written descriptions as methods of providing information in choice experiments. One challenge in designing stated preference studies is providing sufficient information to enable respondents to make informed choices, but not too much where they become fatigued. The use of pictorial representations may be particularly useful in health economics when researchers are interested in treatments or problems which respondents are not familiar with. This study tests for differences in the parameters of the model depending upon whether respondents were provided information with images or not, and also compares external validity results with and without pictorial aids. The results suggest that whilst pictures had an impact on parameter estimates, they did not influence external validity results.

\* *Correspondence to:* Verity Watson, Health Economics Research Unit, Institute of Applied Health Sciences, University Medical buildings, Foresterhill, Aberdeen, AB25 9ZD. Tel: 01224 555937, Fax: 01224 662994, <http://www.abdn.ac.uk/heru/> Email: v.watson@abdn.ac.uk

**Acknowledgements:** Financial support from the University of Aberdeen is acknowledged. HERU is funded by the Chief Scientist Office of the Scottish Executive Health Department (SEHD). The views expressed here are those of the authors and not necessarily those of SEHD.

## 1. Introduction.

Stated preference methods such as contingent valuation and discrete choice experiments, where respondents are asked to provide responses to hypothetical questions, are widely used in several areas of economics including Health, Transport and the Environment. One potential criticism of their use is hypothetical bias i.e. whether responses to hypothetical questions are not the same as respondents' behaviour when faced with actual decisions (Diamond and Hausman, 1994). This paper reports the results of an experiment designed to test the external validity of a discrete choice experiment. Section two presents a systematic review of the external validity literature for both contingent valuation methods and discrete choice experiments. This provides an overview of current evidence and a discussion of how this work can be further developed. Section three discusses the design of the hypothetical discrete choice experiment, the actual experiment and the tests of external validity. Section four presents the results and section five comments on these results and suggests future areas for research. This is very much work in progress and all comments will be gratefully received.

## 2. Previous Empirical Studies.

A systematic literature search was conducted to identify published English language studies testing the external validity of contingent valuation and discrete choice experiment studies. Searches of the following databases were undertaken: MEDLINE; EMBASE; HEALTHSTAR; Social Science Citation Index; PsychLIT; EconLIT; and Health Management Information Consortium database. To reflect the different terms used to refer to discrete choice experiments, search strategies were formulated for each individual database using the following free text terms: 'conjoint', 'conjoint analysis', 'conjoint measurement', 'conjoint studies', 'conjoint choice experiments', 'part-worth utilities', 'pairwise choices', 'discrete choice experiments', 'discrete choice modelling', 'discrete choice conjoint experiments' and 'stated preference'. For contingent valuation, contingent valuation and willingness to pay search strategies were formulated. Some databases included appropriate controlled thesaurus/vocabulary terms, for example in MEDLINE an appropriate MeSH term would be 'external validity' or 'validity'. However, sometimes a controlled vocabulary term was not available. Studies were also identified from the bibliographies of contingent valuation and discrete choice external validity studies identified in the literature search and the websites of academic institutions (the latter to identify discussion papers and published conference papers).

Identified literature suggested that external validity of stated preference methods has been tested in three ways. Firstly, under *laboratory conditions*, where respondents provide hypothetical valuations of goods, which are compared to actual valuation across or between samples. Secondly, *field experiments*, where hypothetical willingness to pay is obtained using a stated preference method and actual willingness to pay is obtained by offering respondents the opportunity to purchase the good. Thirdly, *stated preferences are compared to revealed preferences*. Strictly speaking this may be referred to as a test of the convergent validity of the valuation methods. However in the literature it is

often referred to as a test of external validity. The papers identified are listed in Table 1 (*this review is not complete, and we would welcome information on literature you are aware of*).

### 2.1 Laboratory Based Experiments

Tests of external validity conducted in laboratory conditions have tested whether hypothetical valuations differ from those implied by real purchase decisions. These experiments have mainly valued marketable private goods, using homogenous subject groups, usually university students or staff, and they took place in artificial settings not associated with the purchase of the good on offer, such as university lecture theatres.

Cummings et al (1995), tested the external validity of the dichotomous choice willingness to pay approach using three private goods - an electric juicer, a box of chocolates, and a calculator. They justified using private goods by arguing that these were more convenient to value, and that respondents were familiar with private goods. Cummings et al reasoned if the hypothetical valuation of private goods failed an external validity test, it was unlikely that, less familiar, public goods would pass such a test. In two of the three experiments, the chocolates and the calculator, recruited subjects were undergraduate university students. In the third experiment, the electric juicer, the respondents were recruited from a church congregation. The juicer and chocolate experiments used both within subject and between subject designs. The within subject respondents were initially asked a hypothetical dichotomous choice willingness to pay question. Subsequently they were offered the opportunity to buy the good at the same price. In the between subject design a second sample were only offered the opportunity to buy the good there and then for a set price. This allowed the authors to test if an ordering effects existed in the within subject experiment. In each of the experiments a chi-squared test was used, where the null hypothesis was that the likelihood of a respondent stating yes was unrelated to whether the valuation question was real or hypothetical. Willingness to pay was unrelated to question type in these experiments, thus indicating external validity.

A second external validity test of dichotomous choice contingent valuation was undertaken by Johannesson et al (1998) for a box of Belgian chocolates with undergraduate university students as respondents. Respondents were split into two groups. The first received the hypothetical experiment followed by the opportunity to purchase the box of chocolates at the same price. The second group were only given the opportunity to purchase the box of chocolates at a defined price. The authors used a t-test to compare the mean willingness to pay for both samples. In contrast to the results of the Cummings et al (1995), Johannesson et al (1998) found hypothetical willingness to pay to be greater than actual willingness to pay.

Blumenschein, et al (1998) tested the external validity of a hypothetical dichotomous choice contingent valuation experiment for a pair of sunglasses. Respondents were undergraduate university students who received \$5 for participating. The experiments used a both a within sample and between sample design. One group of respondents were asked the hypothetical dichotomous choice question followed by the opportunity to actually purchase the sunglasses 'here and now'. A second group were only asked the actual purchase question. The hypothetical question was followed by a question asking how certain the respondents were; those who answered 'yes' were asked if they were

'definitely sure' or 'probably sure'. The authors tested for no difference in the proportion of hypothetical and actual 'yes' responses in the within sample design by using a chi-squared test, and in the between sample design using a sign test. The null hypothesis of no difference was rejected for both the within and between sample designs. When only the 'definitely sure' yes responses were counted the null hypothesis of no difference between the hypothetical and actual willingness to pay could not be rejected.

The external validity of an open-ended contingent valuation method was tested by Loomis et al (1996) where hypothetical and actual valuations of an art print are compared. The respondents were 100 university clerical and administrative staff. Each was paid \$20 for attending a valuation session on campus. The sample was split between those respondents who answered a hypothetical question and those respondents given the opportunity to buy the print. The populations were tested to ensure that they were not statistically different prior to testing for differences in respondents' willingness to pay. The willingness to pay measures were not distributed normally, thus non-parametric tests were used. These tested if the responses to the hypothetical and actual experiments were significantly different. Loomis et al found hypothetical willingness to pay amounts for the art print was greater than the actual valuation.

Loomis et al (1997) also tested the external validity of the dichotomous choice contingent valuation using the same good and method described above. The results from this dichotomous choice experiment were compared to the results of the open-ended experiment discussed above. This provided four valuation treatments; hypothetical open-ended hypothetical dichotomous choice, actual open-ended, and actual dichotomous choice. Using a Mann-Whitney test Loomis et al found that hypothetical valuations were greater than the actual valuations for both the dichotomous choice and the open-ended experiments.

Frykblom (1997) compared hypothetical willingness to pay amounts elicited using bidding game and dichotomous choice contingent valuation questions for a private good (an atlas) to the actual dichotomous choice valuation. Respondents were 169 university students recruited in lectures. The students were split between the three valuation techniques: hypothetical open-ended (HOE); hypothetical dichotomous choice (HDC); and real (actual) dichotomous choice (RDC). As there were differences in the socio economic characteristics across the samples the hypotheses that  $HOE=HDC$ ,  $HOE=RDC$  and  $HDC=RDC$  were tested using parametric tests, to test whether mean and variances of responses were the same. The results indicated that the hypothesis that hypothetical open-ended and hypothetical dichotomous choice resulted in the same valuation could not be rejected. However, the hypothesis that the hypothetical and the real question modes resulted in the same valuation could be rejected.

A test of external validity of stated preferences was conducted by Bjornstad et al. (1997) using a referendum for environmental improvement programmes. The referenda required respondents to vote on two environmental programmes: a citizen's guide for drinking water quality or funding a land acquisition programme. In this experiment if the referenda were passed, everyone would have to pay a specified amount. Respondents were recruited from undergraduate students. Each was asked more than one referendum question. This allowed a learning effect to be tested for. Respondents

were asked either: 2 hypothetical referenda one for each programme; 2 real referenda one for each programme or hypothetical referenda for the two programmes plus an actual referendum for one of the programmes. The information collected in the actual referendum was used as a benchmark against which the hypothetical information is compared. Respondents were given a participation fee; if the referendum in the real experiment was passed then this fee was reduced by the required funding amount. A Pearson chi squared test was used to test the external validity. The null hypothesis of no difference in voting behaviour between actual and hypothetical referenda was rejected. However, when the responses for the referenda with three questions were tested the null hypothesis of no difference between the real and the hypothetical referenda could not be rejected. This implied that a learning design might reduce hypothetical bias in contingent valuation experiments.

Carlsson and Martinsson (2001) conducted the only test of the external validity of discrete choice experiments identified. In this experiment university students were recruited after a lecture. Initially they were given a hypothetical discrete choice experiment about donations to three wildlife projects run by the worldwide fund for nature (WWF). The projects were the preservation of rainforests, the preservation of Mediterranean areas, and environmental work in the Baltic Sea. The attributes in this experiment were the project (Mediterranean, Baltic or Rainforest), the amount of money the respondent would receive for participation and the donation that the project would receive. Respondents were asked to complete 16 choice scenarios where the attribute levels varied. Following the hypothetical experiments respondents were asked to complete a further 16 scenarios and told that one of these choice sets would be randomly chosen to be the actual choice. Consequently, the respondent received the participation fee and the wildlife fund received the donation specified. The authors tested the hypothesis of equal parameters between the two models, and could not reject this hypothesis. Following this a t-test was used to compare marginal willingness to pay for each of the attributes, again the null hypothesis of equal marginal willingness to pay could not be rejected. Thus the authors do not reject the hypothesis that the preferences were the same over both experiments, thus concluding external validity in the experiment.

## *2.2 Field Experiments*

The external validity of stated preference methods has also been tested in real-life situations where respondents are offered a good or service hypothetically. Following this the results can be compared **between** samples where another set of respondents were offered the actual purchase decision, or **within** sample by following the hypothetical stated preference experiment with an actual experiment.

Within the health economics literature these tests of external validity **all** have been with reference to CVM. Blumenschein et al (2001) offered respondents a pharmacist led asthma management programme designed to help obtain the optimal management of asthma. All respondents had asthma and had received asthma medication from the pharmacy in the past 6 months. Ten pharmacies took part in the study, 5 offered the hypothetical programme and 5 offered the actual management programme. Those pharmacies in the real group agreed to offer a management programme to any respondent who purchased the good, where the American Pharmacy Services Corporation provided the training. The study recruited 173 asthma patients using a between

sample design 84 patients in the hypothetical group and 89 in the real group. In the real group respondents were given the opportunity to enrol in the programme, whereas in the hypothetical group respondents were asked a hypothetical dichotomous choice question. In both cases the prices varied between pharmacies. This experiment has added credibility in comparison to laboratory tests of external validity as it is undertaken in a location associated with the provision of health care, thus increasing realism. When the characteristics of the samples were compared they were found to have no statistically significant differences. The authors consider the null hypothesis that there are no differences in the proportion of yes responses for the hypothetical and the real experiment using non-parametric tests. The results reject the null hypothesis. Blumenschein et al state that it is necessary to establish a link between the hypothetical and real willingness to pay, this would be possible if the experiment had a within subject design.

Blumenschein et al. (2001) follow up the hypothetical WTP question for the asthma treatment programme with a question asking respondents how certain they are that they will pay their stated amount on a scale from 1-10 from very uncertain to very certain. When only those respondents who stated that they were very certain in the hypothetical analysis are compared to the actual responses then no significant difference between hypothetical yes responses and real yes responses is found. However the authors express some concern, stating that the results do not validate the calibration method, as it would be logical that if only 'definitely yes' responses were included that the WTP would fall. The authors state that it may only be coincidence that the proportion of hypothetical yes responses was similar to the proportion of real yes responses.

Onwujekwe (2001) tested the external validity of dichotomous choice and payment scale contingent valuation elicitation methods using insecticide treated mosquito nets in rural Nigeria. A within sample design was used. Respondents were initially visited at home and asked how much they would hypothetically pay for the mosquito nets, which were shown to them by the interviewer. Within 1-2 months randomly selected respondents were visited again and given the opportunity to buy the mosquito nets at a specified price. This experiment, which allowed the actual and the hypothetical valuations to be compared, found that there was no statistically significant difference between the proportion of respondents who would buy the nets in the hypothetical treatment and those who actual bought the nets.

Within the environmental literature MacMillan et al (1999) tested the external validity of the open-ended contingent valuation method by comparing hypothetical willingness to pay to actual donations to the Isle of Eigg trust. The trust raises money to support the isle of Eigg and implement environmental protection programmes. Respondents were members of the UK general population contacted by mail. A between sample design was used with respondents being split between those who received the hypothetical donation question and those who were asked to make an actual donation. The hypothetical questions were presented in one of two ways - a 'realistic style' which mimics other charitable fundraising styles and a 'neutral design' more similar to that used in contingent valuation studies. In contrast to other tests of external validity, the results found that the actual donations were greater than the hypothetical donations.

Table 1: Review of External Validity studies in WTP literature

	Elicitation Format	Comparison	Test	Good	Familiar Context	Sample Type and Size	Conclusion
Bjornstad et al. (1997)	DC CVM	Real Referendum	Within Sample	Donation to Non Profit organisation	No	Students - 286	Hypothetical differs from Real
Blumenschein et al (2001)	DC WTP	Real Decision	Between Sample	Asthma Treatment Programme	Yes	Asthma Sufferers - 171	Hypo > Actual WTP
Blumenschein, K., et.al (1997)	Vickery Auction	Real valuation	Within Sample	Sunglasses (Private Good)	No	Students - 61	Hypothetical WTP > Real
Blumenschein, K., et.al (1998)	DC CVM	Real Valuation	Between and Within Sample	Sunglasses (Private Good)	No	Students - 133	DC WTP > Real Value
Carlsson and Martinsson (2001)	DCE	Real Decision	Within Sample	Donation to WWF	No	University Students - 35	Hypo = Actual WTP
Clarke, P. M., (2002)	DBDC CVM	TCM	Between Sample	Mammographic Screening (Private Good)	No	Random Sample - 595	DC WTP > TCM Value
Cummings et al (1995)	DC CVM	Real Valuation	Within and Between Sample	3 Private goods (a Juicer, Chocolates and Calculator)	No	Student and Church Congregation - 232	Hypothetical > Real
Frykblom, P., (1997)	DC and OE CVM	Real valuation	Between Samples	An Atlas (Private Good)	No	Recruited Students - 169	(OE WTP=DC WTP) > Actual Value
Getzner, M., (2000)	OE WTP and DC WTP	Real valuation	Within Sample	Endangered Animal Protection Programme (Public Good)	No	Students - 35	Hypothetical WTP < Real WTP
Johannesson (1999)	DC WTP	Real Purchase	Between and Within Sample	Box of Chocolates	No	University Students - 242	Hypo > Actual WTP
Kennedy, C. A., (2002)	DC CVM	Revealed Preference Data	Between Sample	Radon Prevention (Private Good)	No	Convenience sample - 142	Hypothetical WTP = Revealed preference Value
List, J. A, and Shogren, J. F., (1998)	Vickery Auction	Real Valuation	Within Sample	Baseball Trading Cards (Private Good)	Yes	Random Sample - 265	Hypothetical WTP > Real Valuation
Loomis, J., et al (1996)	OE WTP	Real valuation	Between and Within Sample	Art print	No	University Staff - 100	Hypothetical WTP > Real Valuation
MacMillan, D. C., et al (1999)	OE WTP	Real Valuation	Between Sample	Isle of Eigg Trust	Yes	Random Sample 1700	Hypo = Real WTP
Onwujekwe, O., (2001)	DBDC CVM and Bidding game	Real Valuation	Within Sample	Insecticide Treated Mosiquito Nets (Private Good)	Yes	Random Sample - 840	Hypothetical = Actual values
Smith, V. K., and Mansfield, C., (1998)	DC WTA	Real valuation	Within Sample	Respondents Time (Private Good)	No	Random sample - 1037	Hypothetical response = Real Response

The difference between donations was reduced with the 'realistic style' question. This adds to the evidence that realism may improve external validity.

### *2.3 Stated V Revealed Preference*

Responses obtained from stated preference experiments may also be compared to results of revealed preference studies. Although these tests are referred to as tests of external validity strictly speaking these are tests of the convergent validity of the stated preference and revealed preference methods. Two such studies were identified in the health economics literature. Kennedy (2002) compared hypothetical valuations obtained from a contingent valuation study with revealed preference valuations for the prevention of radon induced lung cancer. This experiment valued a market good that provides a health improvement. The revealed preference sample was taken from households found to have a radon level above the nationally defined safe level. These households were offered the opportunity to purchase various levels of remediation at the market prices. The other half of the sample were respondents who were offered the hypothetical remediation option. The author was unable to compare the characteristics of the two samples as confidentiality prevented details of the actual sample being disclosed. Both methods resulted in similar welfare estimates, indicating that the methods may have external validity.

Clarke (2002) tested for hypothetical bias in contingent valuation estimates by testing the convergent validity of dichotomous choice willingness to pay contingent valuation question when compared with a revealed preference travel cost method for mammographic screening in rural Australia. Women were contacted to complete a telephone survey where they were asked their willingness to pay for screening using a double bounded dichotomous choice contingent valuation question. These were compared to the travel cost estimates obtained by asking respondents when they last attended a screening clinic and where they were screened. Costs are assumed to be the sum of the travel cost (\$0.50 per km) the cost of time and user fees. The contingent valuation estimates were found to be significantly greater than those from the travel cost method. This may indicate the presence of hypothetical bias. However Clarke noted that travel cost method only obtains use values for goods whereas contingent valuation will value both use and non-use values. Thus the difference may have been due to the valuation of non-use values.

## **3. Moving Forward: An Experiment to Test the External Validity of Discrete Choice Experiments.**

From the literature summarised in section 2 it can be seen that the external validity tests for stated preferences have focused on contingent valuation studies. Only one study was identified which tested the external validity of discrete choice experiments (Carlsson and Martinsson, 2001). The external validity (and convergent validity) studies can be split in to two types of experiments; laboratory experiments and field experiments. It is suggested that field experiments, although they are more difficult to design, are preferred to laboratory experiments as they increase the realism of the hypothetical setting. Much of the work carried out to date uses between sample designs where the



differences found may be due to differences in the populations. Although the characteristics of the samples are compared these difference may arise in areas for which there is no data available.

Many studies discussed above have been applied to market goods (such as chocolates and art paintings) in artificial circumstances using students, as these are easier to develop and are lower cost. However, this may exacerbate the hypothetical nature of the experiments, especially when the good is offered in a setting unfamiliar to respondents. The experiment discussed in this paper was designed to test the external validity of discrete choice experiments when applied to healthcare provision – in this case Chlamydia screening.

### *3.1 Experimental Design*

#### *3.1.1 The good*

The good being valued in this experiment is a screening test for *Chlamydia trachomatis* (commonly referred to as Chlamydia). Chlamydia is the most common sexual transmitted infection in the UK. Chlamydia is asymptomatic in at least 75% of women. However, if left untreated, it may lead to infertility. National guidelines recommend that women be regularly screened for Chlamydia. This experiment was designed to identify women's preferences for Chlamydia screening. Prior to the questionnaire respondents were provided with an information sheet explaining Chlamydia and the consequences of it not being treated (see Appendix 1).

#### *3.1.2. The Discrete Choice Experiment*

173 respondents were recruited when they visited a family planning clinic in Aberdeen. Whilst waiting for their appointment they completed a discrete choice experiment. In this experiment they were asked to state whether or not they would have screening if it took a form specified in each of 16 choices presented in table 2. Each choice differed according to five attributes; place of screening (family planning clinic, Genito-Urinary medicine (GUM) clinic, GP clinic or at home), type of screening (full pelvic examination, perineal swab and urine test), cost to the respondent of chlamydia screening (free - £20), chance of developing pelvic inflammatory disease if chlamydia is not treated (0% - 25%) and the support provided when respondents are informed of the screening results (No support or the support of a trained healthcare advisor). These levels were selected from the range of options which are currently feasible, and with the advice of the Doctor from the family planning clinic (who was also involved in the design of the study). The cost attribute was determined from the results of a pilot study completed by 45 women. This pilot also lead to changes in question wording for the risk attribute.

From the set of attributes and levels discussed there are 384 possible screening options. These were reduced to a more manageable number screening options using fractional factorial design generated from experimental design package (SPEED) to maintain desirable statistical properties for estimating the model and hypothesis testing. For each of the 16 choices respondents were asked to state whether they would be screened for chlamydia or not be screened (Table 2).

Table 2: Choices presented to individuals in experiment

	Place of Screening	Type of Screening	Cost to you of Chlamydia Screening	Chance of developing Pelvic Inflammatory Disease (PID) if you have Chlamydia and it is not treated	Type of Information and Support when you are given Screening Results	I would have Chlamydia Screening	I would <b>not</b> have Chlamydia Screening
Choice 1	Family Planning Clinic	Full Pelvic Examination	£5	10%	None		
Choice 2	Family Planning Clinic	Perineal Swab	£10	1%	None		
Choice 3	Genito Urinary Medicine (GUM) Clinic	Urine Test	£10	10%	Support of Trained Health Advisor		
Choice 4	At Home	Perineal Swab	£5	5%	Support of a Trained Health Advisor		
Choice 5	At Home	Urine Test	Free	1%	None		
Choice 6	At GP Clinic	Full Pelvic Examination	£20	1%	Support of a Trained Health Advisor		
Choice 7	Family Planning Clinic	Urine Test	£20	5%	Support of a Trained Health Advisor		
Choice 8	Genito Urinary Medicine (GUM) Clinic	Urine Test	£5	5%	Support of a Trained Health Advisor		
Choice 9	Genito Urinary Medicine (GUM) Clinic	Full Pelvic Examination	Free	5%	None		
Choice 10	At Home	Urine Test	£20	10%	None		
Choice 11	At GP Clinic	Perineal Swab	Free	10%	Support of a Trained Health Advisor		
Choice 12	Family Planning Clinic	Urine Test	Free	25%	Support of a Trained Health Advisor		
Choice 13	Genito Urinary Medicine (GUM) Clinic	Perineal Swab	£20	25%	None		
Choice 14	At Home	Full Pelvic Examination	£10	25%	Support of a Trained Health Advisor		
Choice 15	At GP Clinic	Perineal Swab	Free	10%	Support of a Trained Health Advisor		
Choice 16	At GP Clinic	Urine Test	£5	25%	None		

Most discrete choice experiments use text, with little consideration given to pictorial representations of the subject matter. Given current developments in high-powered graphics packages, and the ability to download images from the web, consideration should be given to alternative styles of presenting health goods to respondents. More generally, consideration should be given to whether the type of representation (text, pictorial, interview (with or without pictures), computer based (with or without pictures) influences the model parameters. In the marketing literature Vriens et al (1998) suggest that whilst pictorial representations improve respondent's understanding of the design attributes, verbal representation facilitate judgement. In this study part of the sample were provided with an information sheet that included pictures (see Appendix 1) (86) and part without (87).

The responses to the discrete choice experiment were analysed using a random effects probit where the dependant variable is whether or not respondents would undergo the hypothetical screening option and the independent variables the defined attributes. A constant term is included in the analysis to account for respondents' preference for screening or no screening all other things being equal. Four random effects probit models were estimated; *Model I* was the pooled data (not distinguishing between pictures and no pictures); *Model II* was the pooled data, including a dummy variable to test if pictures influenced respondents probability of being screened for Chlamydia; *Model III* included only those respondents who received pictures and *Model IV* those who did not receive pictures. without pictures. Following estimation of Models III and IV, a Log Likelihood test was used to test the null hypothesis of equal coefficients between these Models.

### 3.1.3. Testing External Validity

A within sample design was used to test the external validity of discrete choice experiments. Following the discrete choice experiment respondents had an appointment with the doctor at the clinic. During the appointment they were offered the opportunity to be screened there and then for Chlamydia. The test offered is the same as screening option 12 in the discrete choice experiment (table 2). At this point the information that was provided to the patients was identical to what had been described in the questionnaire and all patients saw the same doctor. This should minimise information or interviewer bias. To enable the external validity of the experiment to be tested following the appointment with the Doctor it was noted on the questionnaire whether the patient took the Chlamydia screening test offered that day.

The external validity of the discrete choice experiment was tested by comparing the proportion of respondents who stated they would take the screening option described in choice 12 and the proportion of respondents who took the actual test. A number of tests were conducted on this data (*and we would welcome suggestions on other tests*).

i) A McNemar test was used to test whether the proportions saying 'yes' and 'no' were significantly different across the hypothetical and actual choice (Altman, 1991).

ii) A non-parametric one tailed binomial test was used to test whether the proportion of externally valid responses ('yes-yes' or 'no-no') are significantly greater than 0.5, thus implying that the presence of external validity were not random.

In the case of each respondent it was noted whether they received an information sheet with pictures or not. This information was used to test if pictures had any effect on external validity.

From the responses it is possible to identify those responses that are consistent between the hypothetical discrete choice experiment and the actual test. From these an external validity variable may be determined. As this would be dichotomous a logit regression may be used to test for factors that influence respondents external validity, such as difficulty of the experiment and socio-economics characteristics.

#### 4. Results

From *Model I* in Table 3 it can be seen that: the constant and all the attributes except 'risk of developing PID' are significant. The 'family planning clinic' is the preferred location and 'urine test' the preferred screening method. Respondents would rather have the support of a trained healthcare advisor and would select the lowest cost screening option. The positive constant implies that respondents would rather be screened for Chlamydia than not, *ceteris paribus*.

Table 3: Results of Regression Models

	<i>Model I</i> pooled	<i>Model II</i> pooled	<i>Model III</i> with pictures	<i>Model IV</i> without pictures
GUM Clinic	-0.280	-0.282	-0.260	-0.310
(t statistic)	(-2.84)	(-2.18)	(-1.91)	(-2.09)
G.P. Clinic	-0.350	-0.338	-0.424	-0.241
	(-3.54)	(-3.36)	(-3.10)	(-1.61)
At Home	-0.533	-0.504	-0.381	-0.645
	(-5.75)	(-5.35)	(-2.97)	(-4.62)
Full Pelvic Exam	-0.572	-0.584	-0.572	-0.602
	(-7.02)	(-7.06)	(-5.03)	(-4.96)
Perineal Swab	-0.579	-0.583	-0.551	-0.623
	(-7.08)	(-6.99)	(-4.85)	(-5.05)
Cost	-0.057	-0.057	-0.059	-0.057
	(-12.50)	(-12.32)	(-9.24)	(-8.18)
Risk PID	0.002	0.002	0.001	0.004
	(0.54)	(0.58)	(0.10)	(0.76)
Support	0.465	0.464	0.409	0.532
	(6.18)	(6.08)	(3.94)	(4.67)
Constant	1.340	1.243	1.458	1.222
	(10.68)	(8.41)	(8.19)	(6.76)
Pictures	.	0.185	.	.
		(1.12)		
Log Likelihood	-1123.057	-1084.088	-580.804	-499.396
No Observations	2142	2072	1137	935
Rho	0.434	0.438	0.445	0.423

In *Model II* the pictures dummy variable is not significant at the 10% level, suggesting that the pictures had no effect on whether respondents were screened for Chlamydia, *ceteris paribus*. However, the log-likelihood test (from *Models III* and *IV*) suggested that the null hypothesis of equal parameters should be rejected<sup>1</sup> (p-value=0.000), thus implying that including pictures in the information sheet had an effect on parameters of the model.

#### 4.1 Results of the External Validity test.

Allowing for item non-response in the choice experiment there are 126 comparable choice situations. Table 4 presents a comparison of the responses to the choice experiment and the actual choice. 102 respondents answered in exactly the same way when faced with the actual choice as they did in the hypothetical discrete choice experiment. However, 24 respondents behaved in the opposite way to that stated in the hypothetical experiment. Of these 24, 21 were respondents who stated in the discrete choice experiment that they would have the screening option presented in choice 12 yet when they were offered this option by the doctor they declined to be screened. The McNemar test indicated that the response patterns were not significantly different across the hypothetical and actual choice (p=0.000). In the one-tailed binomial test the null hypothesis of random consistency was rejected i.e. we reject the hypothesis that the proportion of inconsistent responses is 0.5 (p-value = 0.00) (and our results are therefore not due to chance).

Table 4: Stated Preference and Actual Responses for Chlamydia Screening

		Stated Preference from DCE			
		Screening		Total	
Actual Preference	Screening	Yes	No		
		Yes	99	3	102
		No	21	3	24
Total		120	6	126	

From the results presented in Table 4 it is evident that most respondents took the test (102 of 126). This is what we would expect given that the constant of the regression was significant. However, one worry would be that respondents were not trading over the screening tests, instead they were always choosing to be screened rather than not screened. The results presented in Table 5 indicate that this is not the case. It is clear to see that for each choice

<sup>1</sup> The chi-Squared statistic for the hypothesis of equal parameters is  $-2[\ln(1123.0566 \cdot 580.80376 / 499.39625)] = 85.71318 \sim \chi^2(10)$

there are respondents stating they would have the screening option or they would not. It can also be seen that choice number 12 (the actual test) was the most preferred option in the discrete choice experiment. It is comprised of the preferred levels of each of the attributes as determined in table 3.

Table 5: Responses for each screening option.

Choice Number	Would have screening	Would not have screening	Total
1	85	45	130
2	71	66	137
3	97	33	130
4	93	47	140
5	94	39	133
6	58	78	136
7	87	41	128
8	116	24	140
9	86	47	133
10	58	80	138
11	76	53	129
12	134	8	142
13	58	70	128
14	66	67	133
15	115	17	132
16	94	39	133
Total	754	1388	2142

The impact of the inclusion of pictures upon the external validity of the experiment is considered by the cross tabulation of validity of the experiment and pictures in Table 6. From the results it appears that there was no difference in the external validity of respondents' answers if they received the information sheet with the pictures or not.

Table 6: Cross tabulation of Validity and Pictures

Pictures in Information (%)	Validity (%)	
	Yes	No
Yes	83	17
No	84	16

Due to limited number of observations and insufficient variation in independent variables it was not possible to run the logistic regression analysis on valid/invalid responses.

## 5. Discussion.

This paper presents a test of the external validity of a discrete choice experiments for Chlamydia screening, and also assesses the impact of pictures on both parameter estimates and external validity. The test of external validity has several advantages over previous tests. The experiment provides a within subject comparison, thus removing the possibility that differences between the hypothetical and actual results are due to differences in sample population. As with Blumenschein, et al (2001), the experiment was conducted in a realistic setting – a family planning clinic where the respondents asked to complete the hypothetical discrete choice experiment prior to an appointment with the doctor and these respondents were offered a chlamydia-screening test during their appointment.

There was no significant difference in the proportions who said they would take the test in the hypothetical and real contexts, providing support for the external validity of the DCE approach. Further, whilst pictures had an impact on parameter estimates, they did not influence external validity results.

Whilst the results provide encouraging support regarding external validity, they do indicate that whilst 81% of respondents provided valid responses, 19% changed their responses in the actual situation. Given the small number of responses that did not display external validity it was not possible to test for factors contributing to invalid responses. It is worth noting that the majority of these responses switched from a 'yes' to 'no' response. Following on from this finding, there is a general concern, raised in the psychological and sociological literatures, about the tendency of people towards 'yea-saying' (i.e. answering 'Yes' to a question) when this is thought to be the socially desirable response (Couch and Kenitson, 1960; Bachman and O'Malley, 1984). This has been applied within the contingent valuation literature to explain the observed higher willingness to pay estimates generated from dichotomous choices studies as opposed to open-ended or payment card studies (Kanninen, 1995; Ready et al., 1996). Applying this to the discrete choice experiment presented in this study, it can be hypothesised that surveys using the DCE approach in the format presented in this experiment (a yes or no response) may produce biased parameters. *Further analysis will investigate whether the 'yes-no' respondents were more likely than other respondents to say yes to all hypothetical choices.*

It may also be that those respondents who stated that they were prepared to be screened in the choice experiment and then stated 'no' when faced with the actual choice, had failed to take constraints into account when answering the hypothetical choice experiment. As the screening test was 'free' subjects did not have to consider their budget constraint. However, once the test was presented to them they would have to consider whether they had time to have the test that day.

Drawing on the stochastic errors literature, failure to answer both choices in the same way may be due to factors such as a trembling hand (Harless and Camerer, 1994), respondents making mistakes (Hey and Orme, 1994) or imprecise preferences (Loomes and

Sugden, 1998). Future research should clearly follow respondents up to investigate why they changed their mind.

By conducting a field experiment it was not possible to test whether respondents would have external validity if they were offered a test for which they had to pay, as Chlamydia screening is offered free on the NHS. One further future development of the experiment presented here would be to apply the test to health care interventions for which a private market exists. Possible applications include recent developments in genetics which may result in the private provision of genetic screening programmes or the MMR vaccine where parents are refusing the vaccine provided by the NHS and instead choosing to pay to have the child vaccinated against each disease separately. By applying the test to marketed interventions it would be possible to test the external validity of willingness to pay measure derived from choice experiments.

Overall the results provide evidence of theoretical validity. However, further research is required to investigate why a number of individuals changed their minds between the hypothetical and actual choice, as well as to whether the results concerning external validity would apply to interventions which incur a charge at the point of consumption.

## References

- Altman, D. G. Practical statistics for medical research, Chapman and Hall, London, 1991.
- Bachman, J.G., and O'Malley, P.M., 1984. Yea-saying, nay-saying, and going to extremes: black-white differences in response styles. *Public Opinion Quarterly*, 48, 491-509.
- Blumenschein, K., Johannesson, M., Blomquist, G.C., Liljas, B., and O'Connor, R. M., (1998), Experimental Results on Expressed Certainty and Hypothetical Bias in Contingent Valuation. *Southern Economic Journal*, **65(1)**, pp169-171.
- Blumenschein, K., Johannesson, M., Yokoyama, K. K., and Freeman, P.R., (2001), Hypothetical Versus Actual Willingness to Pay in the Health Care Sector: Results from a Field Experiment, *Journal of Health Economics*, **20**, pp441-457
- Carlsson, F., and Martinsson, P., (2001), Do Hypothetical and Actual Willingness to Pay Differ in Choice Experiments?, *Journal of Environmental Economics and Management*, **41**, pp179-192.
- Clarke, P. M. (2002), Testing the Convergent Validity of the Contingent Valuation and Travel Cost Methods in Valuing the Benefits of Healthcare, *Health Economics*, **11**, pp117-127.
- Couch, A., and Kenitson, K., 1960 Yeasayers and naysayers: agreeing response set as a personality variable. *Journal of Abnormal and Social Psychology*, 60, 151-174.
- Diamond, P., and Hausman, J., 1994. Contingent valuation: Is some number better than no number? *Journal of Economic Perspectives*, 8, 45-64.
- Frykblom, P., (1997), Hypothetical Question Modes and Real Willingness to Pay, *Journal of Environmental Economics and Management*, **34**, pp275-287.
- Getzner, M., (2000), Hypothetical and Real Economic Commitments, and Social Status, In Valuing A Species Protection Programme, *Journal of Environmental Planning and Management*, **43(4)**, pp541-559.
- Harless D. and Camerer C. (1994) The predictive utility of generalized expected utility theories. *Econometrica*, 62, 1251-89.
- Hey J. and Orme C. (1994) Investigating generalisations of expected utility theory using experimental data. *Econometrica*, 62, 1291-326.
- Johannesson, M., Liljas, B., Johansson, P. O., (1998), An Experimental Comparison of Dichotomous Choice Contingent Valuation Questions and Real Purchase Decisions, *Applied Economics*, **30**, pp643-647.



- Johannesson, M., Blomquist, G. C., Blumenschein, K., Johansson, P. O., Liljas, B., O'Connor, R.M., (1999), Calibrating Hypothetical Willingness to Pay Responses, *Journal of Risk and Uncertainty*, **8**, pp21-32
- Kanninen, B., 1995. Bias in discrete response contingent valuation. *Journal of Environmental Economics and Management*, 28, 114-125.
- Kennedy, C. A., (2002), Revealed Preference Valuation Compared to Contingent Valuation: Radon Induced Lung Cancer Prevention, *Health Economics*, **11**, pp585-598.
- List, J. A., Margolis, M., and Shogren, J. F., (1998), Hypothetical-Actual Bid Calibration of a Multigood Auction, *Economics Letters*, **60**, pp263-268.
- List, J. A., and Shogren, J. F., (1998), Calibration of the Difference Between Actual and Hypothetical Valuations in a Field Experiment, *Journal of Economic Behaviour and Organisation*, **37**, pp193-205
- Loomes G. and Sugden R. (1998) Testing different stochastic specifications of risky choice. *Economica*, 65, 581-598.
- Loomis, J., Brown, T., Lucero, B., Peterson, G., (1996), Improving Validity Experiments of Contingent Valuation Methods: Results of Efforts to Reduce the Disparity of Hypothetical and Actual Willingness to Pay, *Land Economics*, **72(4)**, pp450-461.
- Loomis, J., Brown, T., Lucero, B., Peterson, G., (1997), Evaluating the Validity of the Dichotomous Choice Question Format in Contingent Valuation, *Environmental and Resource Economics*, **10**, pp109-123.
- MacMillan, D. C, Smart, T. S., and Thorburn, A. P., (1999), Validation of the Contingent Valuation Method: A Comparison of Real and Hypothetical Donations to an Environmental Trust, *Environmental and Resource Economics*, **14(3)**, pp399-414.
- Onwujekwe, O., (2001), Searching for Better Willingness to Pay Elicitation Method in Rural Nigeria: The Binary Question with Follow up Method Versus The Bidding Game Technique, *Health Economics*, **10**, pp147-158
- Ready, R.C., Buzby, J.C., and Hu, D., 1996. Differences between continuous and discrete contingent value estimates, *Land Economics*, 72, 397-411.
- Smith, V. K., and Mansfield, C., (1998), Buying Time: Real and Hypothetical Offers, *Journal of Environmental Economics and Management*, **36**. pp209-224.
- Vriens M, Loosschilder G, Rosbergen E, Wittink D. 1998. Verbal versus realistic pictorial representations in conjoint analysis with design attributes. *Journal of Product Innovation*, 15:455-467.

## Appendix 1: Information sheet provided (with pictures)

### Information about Chlamydia & Screening

Testing for Chlamydia is not part of the routine care offered at the Golden Square Clinic, however it is recommended in the national guidelines that women under 20 be regularly tested. This study wishes to find out women's views on Chlamydia screening to improve the healthcare that is offered.

#### **What is Chlamydia?**

Chlamydia is the most common sexually transmitted infection and can affect both women and men. It is caused by bacteria and can go on to problems some time after initial infection. In Aberdeen 1 in 10 people under the age of 25 are infected with Chlamydia and most do not know they are infected.

#### **What are the Symptoms of Chlamydia?**

Chlamydia is often silent. Many people, around 8 in 10 women and 5 in 10 men have no symptoms. This means that people may be infected and not realise it, Chlamydia may remain undetected for months or years.

Although many people have no symptoms for those who do the symptoms are

##### In Men

- Pain or burning feeling on passing urine
- Discharge from their penis

##### In Women

- Pain or burning feeling on passing urine
- Vaginal Discharge
- Bleeding between periods
- Lower abdominal Pain

#### **How can Chlamydia be diagnosed?**

There are several ways to diagnose Chlamydia.

- **Urine test:**

You will be given a 30-mL container to fill up with the first urine that you pass. The urine will be tested only for chlamydia and not for other substances.

- **Perineal Swab**

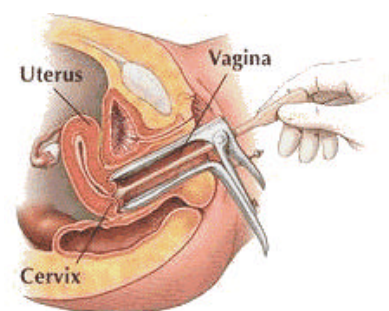
You will be asked to wipe your vagina with a cotton swab.

- **Full pelvic Examination:**

This complete "check-up" will take **more time** as it involves two parts:

1. Look inside the vagina with a speculum that holds the walls of the vagina apart (a little like a duck's bill to see clearly). A small sample of fluid will be also taken from the neck of your womb with a cotton swab.
2. The doctor will put two fingers inside the vagina to examine your womb

This pelvic examination is not painful and most women feel mild discomfort only.



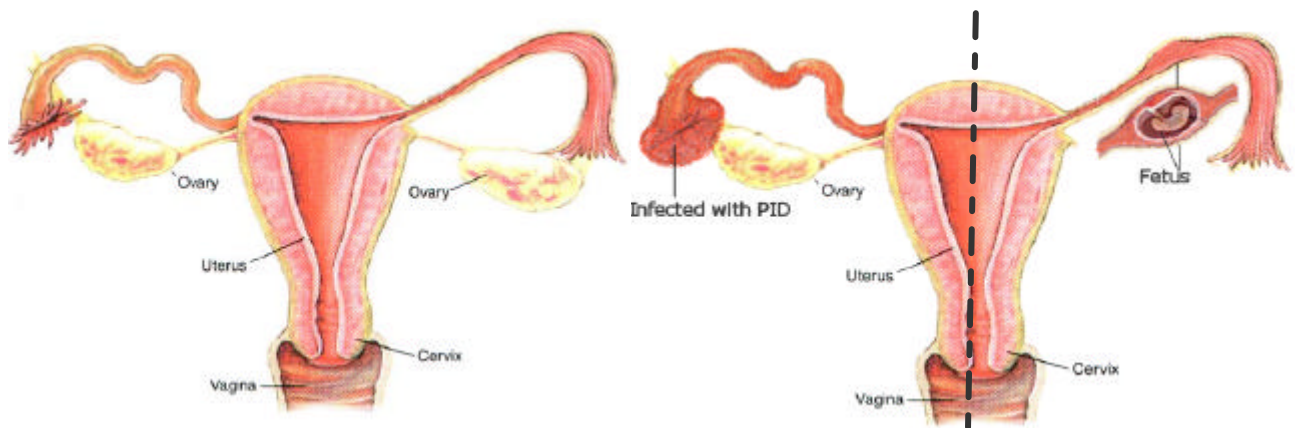
**FULL PELVIC EXAMINATION**

## Where are tests for Chlamydia carried out?

- The GP Clinic
- Family Planning Clinic in golden square.
- Genito Urinary Medicine (GUM) Clinic, this is located at Woolmanhill Hospital
- In your Home.

## What if Chlamydia is not treated?

Chlamydia is dangerous if it is not treated. It may permanently damage the reproductive organs causing infertility. The picture below shows the serious long terms complications that may result



This picture shows a healthy womb and fallopian tubes, eggs pass along the tubes to the womb.

Chlamydia is the main reason women develop **Pelvic Inflammatory Disease (PID)** a painful condition affecting the womb and fallopian tubes. Infected fallopian tubes become scarred as shown in left.

PID may lead to hospitalisation and can cause **infertility** or **ectopic pregnancy**. This is a serious condition where the fertilised egg grows in the fallopian tubes with no possibility of growing into a baby. This is shown in the right hand side fallopian tube

Chlamydia may be passed on to babies at birth causing eye infections or pneumonia.

## Can Chlamydia be treated?

If found early enough Chlamydia is easily cured with a dose of antibiotics.

## How can Chlamydia be prevented?

If you are negative for Chlamydia and want to prevent catching it – condoms will prevent infection.