View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Research Papers in Economics

CHERE
CENTRE FOR HEALTH ECONOMICS RESEARCH AND EVALUATION

UNIVERSITY OF TECHNOLOGY SYDNEY

UTS:CHERE

The Centre for Health Economics Research and Evaluation (CHERE) was established in 1991. CHERE is a centre of excellence in health economics and health services research. It is a joint Centre of the Faculties of Business and Nursing, Midwifery and Health at the University of Technology, Sydney, in collaboration with Central Sydney Area Health Service. It was established as a UTS Centre in February, 2002. The Centre aims to contribute to the development and application of health economics and health services research through research, teaching and policy support. CHERE's research program encompasses both the theory and application of health economics. The main theoretical research theme pursues valuing benefits, including understanding what individuals value from health and health care, how such values should be measured, and exploring the social values attached to these benefits. The applied research focuses on economic and the appraisal of new programs or new ways of delivering and/or funding services. CHERE's teaching includes introducing clinicians, health services managers, public health professionals and others to health economic principles. Training programs aim to develop practical skills in health economics and health services research. Policy support is provided at all levels of the health care system by undertaking commissioned projects, through the provision of formal and informal advice as well as participation in working parties and committees.

University of Technology, Sydney
City campus, Haymarket
PO Box 123 Broadway NSW 2007
Tel: +61 2 9514 4720
Fax: + 61 2 9514 4730
Email: mail@chere.uts.edu.au
www.chere.uts.edu.au

# Decisions about Pap tests: What influences women and providers?

Denzil G. Fiebig[2], Marion Haas[1], Ishrat Hossain[1] and Rosalie Viney[1]

CHERE WORKING PAPER 2007/11

1. Centre for Health Economics Research and Evaluation
Faculty of Business
University of Technology, Sydney

2. School of Economics
University of New South Wales

A preliminary version of this paper was presented at the Australian Health Economics Society Conference held in Perth, September 2006. The research was supported by the National Health and Medical Research Council through a Program Grant (Grant No: 254202).

Corresponding Author:
Rosalie Viney
Centre for Health Economics Research and Evaluation
University of Technology, Sydney
PO Box 123
Broadway NSW 2007
Australia

Email: rosalie.viney@chere.uts.edu.au
Ph: +61 2 9514 4720
Fax: +61 2 9514 4730

First Version: September 2006
Current Version: October 2007

**Abstract**

Despite the success internationally of cervical screening programs debate continues about optimal program design. This includes increasing participation rates among under-screened women, reducing unnecessary early re-screening, improving accuracy of and confidence in screening tests, and determining the cost-effectiveness of program parameters, such as type of screening test, screening interval and target group. For all these issues, information about consumer and provider preferences and insight into the potential impact of any change to program design on consumer and provider behaviour are essential inputs into evidence-based health policy decision making. This paper reports the results of discrete choice experiments to investigate women's choices and providers' recommendations in relation to cervical screening in Australia. Separate experiments were conducted with women and general practitioners, with attributes selected to allow for investigation of interaction between women's and providers' preferences and to determine how women and general practitioners differ in their preferences for common attributes. The results provide insight into the agency relationship in this context. Our results indicate a considerable commonality in preferences but the alignment was not complete. Women put relatively more weight on cost, chance of a false positive and if the recommended screening interval were changed to one year.

Key words: Cervical Screening; Discrete choice experiments; Agency relationships, Consumer preferences

# 1 Introduction

Cervical cancer is one of the most preventable and curable forms of cancer. A cervical cancer may take 10 or more years to develop, during which time pre-cancerous changes can be detected by a Pap smear, also called a Pap test, allowing for early treatment with an excellent chance of a full recovery. It is estimated that regular cervical screening can prevent more than 90% of cervical cancers. As a result, many countries have implemented cervical screening programs, such as the National Cervical Screening Program (NCSP) in Australia. Since the introduction of the NCSP, cervical cancer has fallen from the eighth to the fourteenth most common cancer among Australian women. Deaths from cervical cancer fell by 40% between 1986 and 1998. While the NCSP has been largely successful in achieving its objectives, there remain considerable health policy challenges in relation to cervical screening. These include increasing participation rates among under-screened women, reducing unnecessary early re-screening, improving the accuracy of and confidence in cervical screening tests, and determining the most cost-effective parameters for the screening program, such as type of screening test, screening interval, age to commence and age to cease screening.

The impact of any changes to the screening program, such as an introduction of new screening technology or changes in screening interval, on costs and outcomes depends on how consumers and providers change their behaviour in response to the policy change. In Australia's NCSP, cervical screening is most commonly provided at a primary care encounter. The cervical screening decision occurs as a result of an interaction between a provider giving information and making a recommendation to a patient, who may or may not have been seeking a screening test. It is important to understand the context in which women are provided with advice by their general practitioner (GP) and ultimately make decisions about cervical screening. This is a complex interaction that will depend on characteristics of the woman and the GP, as well as the characteristics of different tests.

Information about women's and providers' preferences and insight into the potential impact of any change to the current program on the behaviour of women and providers are essential inputs into evidence-based health policy decision making.

Existing data sources provide limited information for analysis of consumer and provider preferences for screening, or the likely behavioural responses to changes in policy parameters. Information about screening is collected by the NCSP through the registries, but these data provide aggregate level information only, and are limited to screened women who participate in the registries. The National Health Survey provides some information about screening behaviour, and allows analysis of screening choice based on personal characteristics (Belkar, et al., 2006). However, a limitation of such data is that they provide a single self-report observation of whether a woman has had a screening test within a particular interval, with no information about the context in which the decision to screen is made. Therefore, they have limited value in predicting behaviour in relation to different policy parameters, and limited scope to model interaction with providers.

In this paper we use stated preference data from a discrete choice experiment (DCE) to evaluate consumers' and providers' preferences for attributes of alternative tests, proposed changes to policy recommendations and potential new technologies in relation to cervical screening. Stated preference data are useful in settings and contexts where market or revealed preference data are not available, or where there is limited variability in market or revealed preference data. Both these situations frequently arise for health programs, particularly where there is innovation, or where policies change as is the case here. To predict demand for screening in different policy settings accurately, it is most valid to model jointly the preferences of consumers and providers. Typically, revealed preference data provide information about the ultimate choice made, but are less informative about the consumer-provider interaction that was the basis of this choice. Thus, a further advantage of the use of a choice experiment is that it provides the capacity to model and estimate these interactive relationships.

Separate experiments were conducted with women and GPs to determine the impact of a common set of attributes on their choices and recommendations respectively. Further, the approach allows for characteristics of women to be incorporated in the choice situation presented to GPs and for characteristics of GPs, and their recommendations, to be incorporated in the choice situation presented to women. By

collecting stated preference data it is possible to compare the preferences of each group and to learn more about how they interact. This approach has been applied elsewhere in modelling consumer demand where agency relationships are important (Bartels, et al., 2006), and we demonstrate here that DCEs provide a powerful tool to model the interaction between a doctor and their patient and to investigate the relevant policy issues in the specific context of cervical screening.

# 2 Screening for cervical cancer

## *2.1 Screening decision*

Factors that have been found to be relevant to women's decisions about cervical screening include doctor's recommendation, previous experience with detection of abnormalities and their follow-up, having a female doctor available and women's knowledge of screening recommendations and perceptions of risk (Bush, 2000, Hennig and Knowles, 1990, McKie, 1993, Milburn and MacAskill, 1994). A key question is how important the provider's recommendation is to the decision about whether to have a screening test.

When making the decision to undertake cervical screening, the consumer faces choices about when to have the test and what type of test to have. This decision is made in the context of the advice from the provider, who is better informed about the health consequences of the choice. There is information asymmetry in this situation, and the incentives facing the provider may not be the same as those facing the consumer – for example, the provider may be influenced by factors such as risk of litigation or the provision of incentive payments to increase screening. Such a scheme has existed in Australia. Within this framework it is possible to ask whether the GP acts as a perfect agent for the consumer in the sense described by Culyer (1989). That is, does the GP recommend the choice that the consumer would make if she had full information? Given the role of the provider in the screening decision, it is also important to examine the response of providers to policy parameters.

### *2.2 Screening interval*

The NCSP recommends that all women between the ages of 20-69 should have regular Pap smears every two years. Information about a women's screening history and reminders to women are currently provided through cervical cytology registers as part of the NSCP. However, there is debate internationally about the appropriate screening interval and age range for screening (Bjorge, et al., 1994, Boyce, et al., 1990, Cruickshank, et al., 1997, Grant, 1999, Law, et al., 1999, Van Wijngaarden and Duncan, 1993). The screening interval is three yearly in the United Kingdom. There have been suggestions that the screening interval could increase to five yearly for some women, particularly if such a change was combined with the introduction of HPV (human Papilloma virus) testing or, more recently, the HPV vaccine. Changing the screening interval would affect the cost-effectiveness of the NCSP(Anderson, et al., 2004), but the impact depends in part on how any change in recommendations would affect participation in screening. Given that the 2 yearly screening interval has been long established and widely accepted by the community and providers, one possible concern is that a change in the recommended screening interval may undermine provider or consumer confidence in the screening interval, affecting participation rates.

Australian studies suggest at least one third of women in Australia who had a negative smear have at least one further Pap smear before the recommended 24 months before re-screening has elapsed (Australian Institute of Health and Welfare, 2004, Australian Institute of Health and Welfare (AIHW), 1998, Mitchell, et al., 2000). Participation rates throughout Australia appear to be plateauing at 60-70% (DHAC, 2000). In 2002-03 the age standardised screening rate in the target population of women aged 20-69 was 60.7% but rates of screening vary with factors such as age, socioeconomic status, indigeneity and country of origin (Australian Institute of Health and Welfare, 2005). Thus, confidence in the NCSP may already be fragile, and could be eroded further if additional changes to recommendations were made. Information about how changes to recommendations about screening interval and age range might influence choices and subsequent behaviour of providers and consumers is thus relevant to the assessment of cost-effectiveness and important to any decisions to change the current program.

### *2.3 New technologies*

While manual reading of a Pap smear is the current recommended screening technology, there have been a number of recent developments in cervical screening technology which aim to improve the detection of abnormalities and enhance the overall accuracy and value of screening. These new technologies are at different stages of development and none has yet been fully evaluated in terms of its potential role in a national screening program, either in Australia or overseas. Nonetheless, some are currently available in Australia and have been actively marketed to women and providers. Liquid based cytology and automated screening both have been available in Australia for a number of years and are sometimes used as an adjunct to conventional manual reading. Women presenting for cervical screening may currently be offered one or both of these technologies in addition to a conventional Pap smear. As these are not covered by Medicare, the services attract an additional out-of-pocket fee. Applications have been made to the Medical Services Advisory Committee for listing of liquid based cytology and HPV testing on the Medicare Benefits Schedule, but neither has at yet been recommended.

Liquid based cytology is the most widely available adjunct technology in Australia. It involves the use of liquid-based sample collection and automated slide preparation, designed to provide more representative samples of evenly dispersed cells. In the past decade, there has been growing concern from providers about litigation arising from cervical screening, in part due to perceptions about the accuracy of cervical screening (Mitchell, 1997). Estimates of the sensitivity and specificity of the conventional Pap smear vary from 55% to 80% for sensitivity and from 65% to 99.4% for specificity (Fahey, et al., 1995, Soost, et al., 1991, van Oortmarssen and Habbema, 1991). Systematic reviews have reported that liquid based cytology can increase the sensitivity of the Pap smear but there are no accurate estimates for specificity (Payne, et al., 2000). One reason for the increased use of adjunct technologies in Australia is the possibility that they may provide greater accuracy in screening, and that this may be attractive to providers concerned about litigation.

However, there is some evidence that, while the new technologies increase the rate of detection of cervical abnormalities (Austin and Ramzy, 1998, Australian Health Technology Advisory Committee, 1998, Cuzick, 1999), these appear more likely to be

minor lesions such as atypia, minor non-specific changes and low grade CIN (ie CIN1), which are less likely to progress to cancer than higher grade lesions. Ideally, new technologies should significantly enhance the detection of high grade CIN (CIN2 and CIN3) but without undue over-detection of low grade lesions. It is important to assess the tradeoffs made by women and providers in assessing the value of increased accuracy in detecting high grade lesions relative to the increased possibility of detecting low grade lesions.

### 2.4 *Increasing participation consistent with the national recommendations*

As an adjunct to the NCSP, the Australian Government Department of Health and Ageing has introduced a range of incentives via the Practice Improvement Program (PIP) to encourage general practices to increase participation in cervical screening consistent with the nationally recommended guidelines, that is, to increase screening of under-screened women, and to maximize the proportion of the target population that are screened at the recommended interval. The PIP Cervical Screening Incentive offers financial incentives to encourage GPs to take cervical smears from unscreened and under-screened women aged between 20 and 69 years (http://www.cervicalscreen.health.gov.au/internet/screening/publishing.nsf/Content/professionals, accessed September 13, 2006). The PIP comprises a Sign-on Payment, a Service Incentive Payment (SIP) for each screen of a woman in the target age range who has not been screened in the past four years, and an Outcomes Payment that rewards practices that adopt a systematic approach to cervical screening, as measured by reaching the specified target screening rate. These incentives are all aimed at changing the behaviour of GPs in relation to cervical screening, for example, by increasing the rate of opportunistic screening.

## 3 Discrete choice experiment

### *3.1 Methods*

In DCEs the stated preferences of individuals are collected via surveys in which respondents are asked to make choices from a set of hypothetical but realistic alternatives. Respondents are asked to indicate which option among those presented in each scenario is preferred (that is, which they would choose if the options presented represented the available choice set, with a “none of these” option often allowed).

Each alternative is described in terms of its underlying attributes and these can be varied across the range of plausible and policy relevant levels. Attribute levels are varied independently, and so respondents are forced to make trade-offs between attributes; for example between a higher priced, more accurate test and a less costly but lower quality test. Because the combination of attributes and levels is potentially very large, experimental design principles are used to generate a sample of choice sets with the appropriate statistical properties to allow the quantification of the effect of the attributes, independently and together, and to predict how people will choose under different circumstances. In order to increase the sample size in a cost-effective manner, each respondent is asked to perform not just one choice task as described above, but rather several such choice tasks, in each of which the respondent faced a new set of hypothetical alternatives.

There are many advantages and opportunities associated with stated preference data generated by DCEs but their collection also comes with added responsibilities (Viney et al, 2002). The selection of the respondent sample and the development and testing of the survey instrument are issues in all survey-based research. In addition, conducting a DCE survey requires attention to the framing of choices in a realistic and useful way, the selection of attributes and levels that are credible and meaningful to respondents, and designing the choice sets in accordance with the principles of experimental design.

### *3.2 Development of the choice experiment*

As cervical screening is undertaken at intervals that may be two years or more in length, it is appropriate to conceptualise the interaction as a "one-shot" interaction, where the GP makes the recommendation and the women makes the consumption decision in the same encounter. For the women, the decision is whether or not to have a cervical screening test and if so whether to have a standard Pap smear or an alternative test technology which may be available. For the provider, the decision to be modelled is whether to recommend a test at this encounter, and which combination of testing modalities to recommend.

The choice experiment involved the collection of stated preference data in experiments conducted on two independent samples: women in the target population, and GPs. The samples were independent in that the sample of general practitioners was not selected to be the actual providers to the women who responded to the survey, or vice versa. Although the nature of the relationship between the provider and consumer (e.g. is this her usual GP?) is important in the analysis of interactive effects, this can be addressed within each survey by inclusion of consumer characteristics in the survey to providers and provider characteristics in the survey for consumers. To address issues relating to the decision to have a first cervical screening test (an important policy question given the persistence of a group of women in the population who have never been screened), two separate experiments were designed for the women; for women who had and had not had a cervical screening test previously. In the current paper, analysis and results are only presented for the previously screened women.

A feature of these particular choice experiments is that it was necessary to consider not only attributes of the alternatives (the screening tests), but also of the decision making context. That is, for a woman, the decision to screen depends on factors relating to the encounter that will be common across all screening tests available at the encounter, such as familiarity with the GP or the time since her last screening test. These factors are important because they may be the key determinants of participation in screening, and if they are not explicitly included in the experiment the responses may lead to inaccurate estimates of the impact of attributes of the screening test. They are also important because of their relevance to policy options available. Similar considerations apply for the GP in making a screening recommendation. Thus, the experiment included a set of context attributes which were common across all screening tests offered, and a set of alternative specific attributes, that varied across the tests offered. The doctor attributes in the patient choice task are context attributes as they vary over scenarios but not over options. That is, for any particular scenario, women are asked to choose between alternative tests but are given advice by the same doctor. Similarly patient attributes that appear in the doctor choice task vary over scenarios but not over alternative tests. Other context variables are provided by attributes such as time since last screening test and the recommended screening interval.

Tables 1 and 2 present examples of the choice sets faced by women and GPs respectively. Different variants of the choice tasks are generated by systematically varying attribute levels. From Tables 1 and 2, there is a choice between standard Pap tests and liquid based Pap tests, which potentially provide greater accuracy, but at a greater cost. Conditional on choosing one of the two different Pap tests, women are also asked whether they would have an additional HPV test. For both choice questions and for both sets of respondents, the recommendation of (Carson, 2000) is followed by including a reference alternative that is constant across all choice occasions and allows the respondents to choose not to be tested, in the case of women, and, not to recommend any test in the case of GPs. In general this adds to the realism of the choice tasks but, because of the importance of the regularity of screening and the potential impact of recommended intervals between screens, the presence of this "opt-out" or "none" choice is an essential component of the choice task.

The choice of attributes and levels in the experiment was determined on the basis of a review of literature relating to cervical screening decisions and provider recommendations, and consideration of the current policy context for the NCSP. The final set of attributes and levels was informed by the conduct of pilot studies conducted with 79 women in the target screening age range and 12 GPs, which tested the comprehensibility of the choice scenarios and attributes and the appropriateness of the range of levels. The stated preference analogue of poor-quality revealed preference data occurs when respondents are offered attribute levels over which they are unwilling to trade. From the pilots it was evident that too many levels had been chosen in defining the attribute for how overdue (according to the screening interval) a Pap test was for a woman was to be tested, and the number of levels was reduced. A further change was to reword this attribute so that it was described in terms of the number of years since the last screening test, rather than the actual screening interval relative to the recommended screening interval. Apart from this change the pilots confirmed the viability of the planned study (Fiebig and Hall, 2005).

Tables 3 and 4 present the full list of attributes and levels in the experiment. For the women, the context attributes included characteristics of the GP (her familiarity with the GP, the sex of the GP, and whether the GP would receive any incentive payment

for a cervical screening test), the time (in years) since her last Pap test, the current national recommended screening interval (in years) and the doctor's recommendation. There were two attributes that related to HPV testing, but these are not considered in the current analysis. For the GPs the context attributes included the reason for the consultation (specifically for a Pap test, for a general check-up, or for a minor or serious health problem), the familiarity of the doctor with the patient (eg whether they were a regular patient or a new patient), the age and socioeconomic status of the patient, the time in years since the patient's last Pap test, the recommended screening interval and whether the GP would receive an incentive payment. The levels of incentive payment were no incentive payment, an incentive payment if the woman was overdue for a screen, an incentive payment for screening at the recommended interval and an incentive payment for reaching a screening target amongst the eligible population.

Most attributes are self-explanatory from the tables, but the cost attribute requires some explanation. In the Australian context, the liquid based test is not currently covered by the Medicare Benefits Schedule, and thus, is not refundable via Medicare. It is also widely known to be an add-on to the standard Pap test. Therefore, it appeared unrealistic that the liquid based test would be cheaper than the standard Pap test, so the attribute was constructed such that the levels defined the additional cost of the liquid based test relative to the cost of a standard test. Given the construction of the design, this approach resulted in three levels of cost appearing in the survey for the liquid based test ($20, $40 and $60), although these cost levels were paired with the cost of the standard test such that the additional cost ranged across all four levels ($10 -$40).

From Table 3 it can be seen that for the previously screened women there were 8 context attributes (4 with 4 levels and 4 with 2 levels) and 3 alternative specific attributes, each with 4 levels. For the providers, there were 8 context attributes, all with four levels, and the same 3 alternative specific attributes. Thus, overall the full factorial for the women's survey includes $4^4*4^2*4^{3*2}$ alternatives, and $4^8*4^{3*2}$ alternatives. The design for this study was constructed using systematic techniques that have been developed to find optimal or near-optimal designs for choice

experiments; see for example Burgess and Street(Burgess and Street, 2003, Street and Burgess, 2004a, Street and Burgess, 2004b)

The need to design for both the context and the alternative tests created an additional experimental design complexity. The optimal design for this specific choice problem is not known. Therefore a design strategy was devised that would ensure that the effect of the alternative specific attributes could be estimated independently of the context attributes (which are common to all alternatives offered in a choice set). The designs were constructed by finding separate optimal designs for the common attributes and the alternative specific attributes, and then combining these to create an overall design. In effect, this breaks the design problem down into two stages. In the first stage, the choice problem is characterized as the decision to screen/recommend screening, given the attributes of the choice context. In the second stage, the choice problem is characterized as which screening test to choose/recommend. Thus, each choice set from the design for the common attributes was combined with every choice set for the alternative specific attributes. This approach led to a final design of 512 choice sets for each experiment (comprising 32 scenarios for the common attributes, and 16 choice sets for the alternative specific attributes. Each design was blocked into 16 versions, each with 32 choice sets. Allocation of choice sets to versions was systematic to ensure that each version included all 32 scenarios for the common attributes.

In addition to the choice sets, the women's survey also included questions about socio-demographic characteristics and the woman's cervical screening history. The GP survey also included questions about socio-demographic characteristics, practice characteristics, knowledge of and familiarity with cervical screening tests, and a series of attitudinal questions relating to the current screening guidelines, medico-legal concerns in relation to cervical screening and opportunistic screening.

### *3.3 Recruitment and data*

Respondents for the women's survey were randomly sampled from the NSW population, aged 18-69 and recruited via door to door recruitment with random start points. Recruitment of respondents and conduct of interviews was carried out by an external recruitment and data collection firm, Surveys Australia. Participants were

provided with a double movie pass in recognition of the time taken to complete the survey.

Respondents for the GPs' survey were randomly sampled from the Australian Medical Association contact list for GPs in NSW. Recruitment was via mail out of a self-completed survey, with telephone follow-up. Participants in the GP survey were provided with a gift voucher in recognition of the time taken to complete the survey. Participants were given an information sheet and provided written consent to participate. Responses were anonymous. The study was approved by the University of Technology, Sydney Human Research Ethics Committee. Respondents were randomly allocated to versions of the survey.

The total sample of women comprised 234 women, of whom 167 were in the sample of women who had previously had a cervical screening test (analysed in this paper), giving 5344 choice observations. The total sample of general practitioners comprised 215 GPs, giving 6880 choice observations. Sociodemographic variables for each sample are summarised in Table 5.

Table 6 summarises the choice frequencies in the survey. Women chose no test in 37.3% of the choice sets, the standard Pap test in 39.1% of choice sets and the liquid based test in 23.7% of choice sets. GPs chose no test in 42.8% of choice sets, the standard Pap test in 32.4% of choice sets and the liquid based test in 24.7% of choice sets. Apart from twelve women (ten of whom always chose the standard Pap test and two always chose the liquid based Pap test) and two GPs (one of whom always chose no test, and the other always chose the liquid based test), all other respondents were responsive to changes in the attributes in the choice sets. However, 13.8%, 10.2% and 26.3% of women never chose no test, the standard Pap test or the liquid based test respectively. The corresponding figures for general practitioners were 1.4%, 5.6% and 22.3%. Thus, particularly for the liquid based test, the raw data suggest that a relatively high proportion of respondents in both samples had strong preferences against this option.

# 4 Estimation methods

The statistical analysis of choice data relies on the random utility model (McFadden and Train, 2000) where each respondent faces a choice amongst *J* alternatives repeated under *S* scenarios or choice situations. The utility that individual *i* derives from alternative *j* in scenario *s* is composed of systematic and random components denoted by

$$(1) \qquad U_{isj} = X'_{isj}\beta_i + \varepsilon_{isj}$$

where $X_{isj}$ is a *K* x 1 vector of explanatory variables and $\beta_i$ is a conformable vector of coefficients.

Conditional on $\beta_i$, and assuming the disturbance terms $\varepsilon_{isj}$ to be identically and independently distributed (IID) as extreme value, the standard multinomial logit (MNL) specification results (Train, 2003). The probability that individual *i* chooses *j* in scenario *s* is then given by:

$$(2) \qquad P_{isj} = \frac{\exp(X'_{isj}\beta_i)}{\sum_h \exp(X'_{ish}\beta_i)}.$$

Simplicity of estimation and interpretation are among the main advantages of this model but these come at the cost of some restrictive assumptions that may be unrealistic in many situations (Train, 2003).

In general, variability (heterogeneity) among respondents is expected, for example because of differences in tastes and decision making processes. Therefore, respondents with the same observed characteristics may value and weight attributes of a product differently when making a decision. The MNL specification can be generalized to account for this heterogeneity by allowing components of coefficients ($\beta$) to randomly vary over individuals but not over the repeated choices made by an individual by setting:

(3) $\beta_{ki} = \overline{\beta}_k + \varpi_{ki}$ $k = 1,\ldots,K$

where $\overline{\beta}_k$ is the mean parameter vector for the population and $\omega_{ki}$ is the individual specific deviation from the mean. In this random parameter framework there is flexibility in the choice of the distribution of $\omega_{ki}$. Here they are assumed to follow standard normal distributions, independent of each other and of the $\varepsilon_{isj}$. This specification introduces error correlation across choice situations, accounting for the dependence structure in unobserved utility among the repeated choices of an individual which comes from the panel structure of the data. This would be expected, since the same unobserved factors affect a specific respondent, to a certain degree, over the repeated choices. MNL would not capture that dependence. This correlation is not perfect because of the presence of the independent extreme value terms $\varepsilon_{isj}$. Even though the $\omega_{ki}$ are assumed to be independent, this specification also induces correlation across the alternatives in each choice situation as long as generic attributes appear in the utility specifications for these alternatives.

Advances in computer power and simulation based methods have made the resultant random parameter or mixed logit (MXL) model computationally feasible to estimate and popular in empirical work (Hall, et al., 2006, Revelt and Train, 1998, Train, et al., 1999). Estimation by maximum simulated likelihood (MSL) is undertaken using a program downloaded from Kenneth Train's website (Train, 2004). All estimation results reported below were generated using 1000 Halton draws to simulate the likelihood functions to be maximized (Train, 2003);

We restrict our attention to the initial decision to choose or recommend a cervical screening test and which test to choose or recommend. Analysis of the second stage decision to choose or recommend an HPV test will be pursued in later work. Because the model specification is similar for both women and GPs (albeit with a different set of explanatory variables for each) they are discussed together below. In each of 32 choice tasks, the respondent was asked either to choose (for the women) or recommend (for the GPs) between three different alternatives: a standard Pap test, a

liquid based Pap test or no test. The utility that individual *i* derives from choosing/recommending any test:

$$(4) \qquad U_{isj} = DL_j \alpha_{i1} + DT_j \alpha_{i2} + X'_{isj} \beta + \varepsilon_{isj} \quad ; j = liquid, standard.$$

The utility from the third alternative (*j = no test*) is normalized to zero.

In (4), *DL* is a dummy variable for the liquid based Pap test and *DL* is a nesting dummy for recommending *either test* (where the base case is no test). Modelling intercepts as the only random coefficients is a common way of capturing heterogeneity in repeated measures or panel data. How the alternative specific intercepts are entered makes no difference in MNL but in MXL different specifications imply different covariance structures across choice alternatives (Hall, et al., 2006). The specification in (4), making the coefficients associated with the *either test* dummy and the *liquid* dummy random, induces correlation across the two test choices and choosing the shift dummy for *liquid* rather than *standard* is consistent with the hypothesis that there is more variability associated with the less commonly used test. In the estimation, this assumption is checked by running the alternative model with *either test* and *standard* dummies.

The $X_{isj}$ are the other attributes specified in the choice task (both the context attributes and the alternative specific attributes). Although there is overlap in the variables included in the two experiments, the approach taken to consider the interactive nature of decision making is to analyse the decisions of the women and the GPs separately, but to assess the extent to which they had similar responses to variables common to both choice decisions, and the extent to which attributes of the GP impacted on the choices made by women and vice versa.

## 5 Estimation results

Log-likelihood values for each of the MXL specifications are compared with the standard multinomial logit model in Table 7. Either of the MXL models with the intercepts as random coefficients results in a dramatic improvement in fit over the

MNL models. While these models are nested, the hypothesis tests are non-standard because the parameter space is restricted under the alternative. In such situations the LR test statistic does not have the usual chi-square asymptotic distribution (Andrews, 1998). However, in this case the appropriate critical value for the LR test will be smaller than the usual chi-square value. Therefore, the LR test statistic for the comparison of the MNL and MXL models will lead to rejecting MNL in favor of the MXL model at every reasonable significance level. As expected the MXL specification with *standard* as the shift dummy is dominated in terms of fit by the specification with *liquid* as the shift dummy and hence is consistent with our hypothesis of higher variability for liquid Pap tests. Tables 8 and 9 present the results for this second MXL model specification for the women and the GPs respectively.

For the women, the mean coefficient for the random intercept for *either test* was positive but not significant, but the mean coefficient for the random intercept for the liquid based Pap test was negative and significant. This replicates, for the base case with all continuous attributes and demographic variables set to zero and all qualitative variables set at their omitted level, the choices of women in the survey where they tended to choose the standard Pap test and no test in preference to the liquid based test. The standard deviations of both random coefficients are significant, indicating heterogeneity among respondents in their preferences for both tests and the presence of significant persistence in choices across choice situations. The expected correlation between the two tests is captured by the standard deviation associated with the *either test* variable.

In general, the variables included in the experiment were significant and had expected signs. Women were less likely to choose a test if the GP was male, or if the GP was not their regular GP with the two effects having a similar magnitude. The variables for the nationally recommended screening interval suggest that women are responsive to the recommended policy – that is, they were more likely to choose a cervical screening test if the recommended screening interval was one year (relative to the base case of the current screening interval of two years) and less likely to choose a test if the recommended screening interval was three or five years. Consistent with this, women were responsive to the time since their last test, and were more likely to

choose to test if their last screening test was 2 years, 3 years or 5 years ago relative to the base case of one year ago. The doctor's recommendation was also significant. Women were more likely to choose to have a cervical screening test if the doctor recommended it and to choose a particular test if the doctor recommended it. Women's decisions to screen were not affected by whether their doctor would receive an incentive payment or not - this variable had a coefficient that was small and not significant in the analysis.

In terms of the alternative specific attributes, all of these were significant and had the expected signs. Women were less likely to choose a specific test the more expensive it was, and the higher the false positive and false negative rates.

The estimated model also included sociodemographic characteristics: age, education, smoking status, country of birth and dummy variables for high and medium household incomes. The latter were interacted with the choice to determine if women with higher incomes were more likely to choose the liquid based test. Inclusion of the sociodemographic characteristics improved the fit of the MXL model indicating significant sources of heterogeneity associated with observable differences between women in addition to the unobserved heterogeneity captured by the random coefficients. The income variables show that women with higher incomes (relative to the base case of lower income group) were more likely to choose the liquid based test, as the coefficients for both high and medium household incomes when interacted with the liquid based test were large and precisely estimated. Those women who did not report their income were much more likely to choose no test in preference to either the standard or liquid based test. The remaining sociodemographic characteristics had estimated coefficients with signs that were typically as expected but none were significant at the 5% level.

For the GPs, the pattern of estimates for both random coefficients is similar to that for the women. What is somewhat different is that relative to the estimated means, the estimated standard deviations are much smaller for the GPs. This is an indication that there is less variability in the GPs recommendations compared to the women's

choices; a result that is not surprising given asymmetry in knowledge and experience in making these decisions.

Relative to the base case of a woman who had consulted the doctor specifically for a cervical screening test, doctors were less likely to recommend a test if the woman had consulted for any other reason. This was the case even if the reason was a general check up or a minor health problem, but much less likely if the reason for the consultation was a serious health problem. Interestingly, relative to the base case of a woman who the GP sees regularly for most of her primary health care, doctors were more likely to recommend a cervical screening test at this consultation for a less familiar patient. This suggests that doctors are willing to undertake opportunistic screening, but are less likely to do so for a woman that they see regularly, perhaps because they consider that they are likely to have other opportunities to recommend a cervical screening test to women they see regularly for primary health care, or because they are more confident about the screening history and behaviour of these women.

GPs were highly responsive to the time since a woman last had a cervical screening test. Relative to the base case of a year ago, the odds ratio for recommending a test for a woman who had had a Pap test 2 years ago, 3 years ago or never were 5.2, 17.8 and 45.9 respectively.

Relative to the base case of a woman who is less than 20, doctors were significantly less likely to recommend a cervical screening test for a woman aged 20-29, a woman aged 60-69 and a woman aged 70 or more. The coefficient for women aged 30-59 was not significantly different from the base case. Thus, overall, doctors seem more likely to screen very young women (who may be more likely to have never screened before) and less likely to screen women in the older age groups. Relative to the base case of a woman in the lowest socioeconomic group, doctors were less likely to recommend screening to women who they perceived to be in higher socioeconomic groups. This may be consistent with a perception that opportunistic screening is more important for this group, or with perceptions about risk factors for cervical cancer.

GPs appeared to be willing to adhere to nationally recommended guidelines. Relative to the base case of the current screening interval (two years), doctors were more likely to recommend a test if the recommended screening interval was one year, and less likely to if the recommended screening interval was three years or five years. By contrast, GPs did not appear to respond in the expected manner to practice incentive payments. The coefficients for the incentive payment for screening a woman who was overdue for a Pap test, or for screening at the recommended screening interval were not significant, and the coefficient for an incentive payment for reaching a target screening rate of greater than 70% was significant but negative. This counter-intuitive result may suggest that doctors responded negatively to the implication that their screening recommendations would be influenced by financial incentives.

In terms of the alternative specific attributes, all of these were significant and had the expected signs. GPs were less likely to recommend a specific test the more expensive it was, and the higher the false positive and false negative rates.

The estimated model also included self-reported characteristics of the GP. The gender of the GP was not a significant effect indicating no difference between males and females in whether they recommended either test compared to not testing. GPs who had been practising for less than one year were less likely to recommend a cervical screening test, but otherwise years of practice was not a significant effect. Variables were included to reflect the doctor's own perception of their usual practice in making recommendations between different cervical screening tests. These were interacted with the recommendation. Doctors who indicated that they mostly recommended the standard test were significantly more likely to recommend the standard test, and doctors who indicated they mostly recommended the liquid based test were more likely to recommend it. However, for doctors who reported that they recommended both tests equally, the coefficient for the standard test was not significant, but they were significantly more likely to recommend the liquid based test. One possible explanation for this is that the scenarios in the choice experiment were more weighted towards situations in which these doctors would recommend a liquid based test.

In order to further compare the preferences of women in making their choices and the GPs when they make their recommendations, Table 10 collects the estimates for attributes common to both experiments. Because these estimates come from separate estimations it is important to acknowledge the potential for differences due simply to scaling. The scale is normalized on the extreme value errors in each case and there is no reason why the degree of heterogeneity reflected in the extreme value errors should be the same for the women and the GPs. Table 10 controls for this possibility by reporting ratios of coefficients. Ratios that were similar in magnitude would indicate the presence of the scaling phenomena and, that after correcting for scale differences due to inherent differences in heterogeneity, would indicate women and GPs were making the same trade-offs across the common attributes. Another way to control for scale is to calculate marginal rates of substitution. These are also provided for the two sets of results using the chance of a false negative as the numeraire.

Five of the coefficient ratios are similar in magnitude, which is consistent with the hypothesis that scaling explains the differences in the estimated coefficients. Moreover, in these cases the ratios are less than one, consistent with the plausible result that women exhibit more heterogeneity in their choices compared to that associated with the GPs. Thus for both levels of when the patient last had a cervical screening test, the chance of a false negative and for the two levels of the recommended screening interval greater than two years both women and GPs are on average making similar tradeoffs, indicating that in these cases their preferences are aligned. Similarly the rates of substitution show how close the tradeoffs of women and GPs are for these attributes.

However, the fact that the ratios of estimates for cost, chance of a false positive and if the recommended screening interval is one year are all greater than unity indicates that this alignment of preferences is not complete. In terms of rates of substitution, that for the one year interval is 1.87 times larger for women with the discrepancy even more pronounced for the other two attributes. We stress that our results are drawn from separate samples and so do not reflect the interaction of a specific GP and

patient. Instead we can say that on average, women put relatively more weight on these attributes than GPs when making their choices.

## 6 Discussion

To our knowledge this is the first discrete choice experiment in the health setting to examine the interaction of preferences of experts and consumers in a setting where the expert (provider) is in an agency relationship with the consumer. A previous paper has examined providers' and women's preferences for cervical screening programs (Arana, et al., 2006) but there, both the women and the providers were being asked about their own screening decisions, rather than seeking information about the providers' recommendations. Similarly, other studies have examined issues of agency by considering patients' preferences for characteristics of providers (Vick and Scott, 1998) but not the interaction of provider and patient preferences.

Our results indicate a considerable commonality in preferences but the alignment was not complete. Women put relatively more weight on cost, chance of a false positive and if the recommended screening interval is one year. One possible explanation is that GPs have inaccurate perceptions of the preferences of their patients. Another possibility is that they do correctly perceive the preferences of their patients but shade their recommendations in line with their own preferences. Also because of the asymmetric information that characterizes the doctor-patient relationship, a GP might not recommend the alternative preferred by the patient not because of any misperceptions or because of their own self-interest but because they feel, based on their superior knowledge, that it is in the best interests of the patient.

It is difficult to determine from our analysis which of these explanations is driving the observed differences in preferences. However it is possible to make some plausible conjectures. In order to argue that it is inaccurate perceptions by GPs one would have to argue why it only occurs for three of the attributes and is not more systematic across attributes. Such arguments seem difficult to mount.

In the survey, GPs were asked to rate their level of concern about legal action being taken against them. On a 7-point Likert scale, only 13% of GPs rated their concern about over-screening at one of the three highest levels, while 34% did so for their concern about under-screening. Thus the GPs may not be trading off cost and chance of false positives as much as women because concerns about litigation lead them on average to recommend screening more than women would choose screening when faced with the same scenario. But if this is the reason why didn't it manifest itself in other common attributes where, according to our results, women and GPs are making similar tradeoffs. More telling is the inconsistency for the attribute associated with the recommended screening interval being one year. The fact that women are placing relatively higher weight on this attribute indicates they are more likely to choose to screen than the GPs are to recommend screening in response to a change from two years to one year in the recommended screening interval.

For false positive results and the preferred frequency of screening, GPs are likely to be better informed and even though they have accurate perceptions of the preferences of women they might shade their recommendations in accord with their view of what is in the best interests of their patient. For cost, it is reasonable to expect some doctors to consider that it is up to women themselves to decide whether they can afford a particular test and that the GP recommendation should give more weight to the health and clinical aspects of the decision.

The results also show that there are important aspects of the interaction between the patient and the provider that are likely to affect the outcomes of a program such as the NCSP. In particular, opportunistic screening is a potentially important way of increasing the screening rate, particularly among unscreened or under-screened women. Doctors appear willing to engage in opportunistic screening in the sense that they are responsive to whether or not they expect to have future opportunities to screen the woman (because she is a regular patient), but are unwilling to recommend screening to a woman who is consulting for another reason. By contrast, women are more likely to take up screening if it is recommended by the doctor. This appears to

provide an opportunity to increase screening rates by encouraging doctors to recommend screening even when the woman is attending for another reason, particularly a general health check or a minor health problem. However, an interesting challenge is created by the fact that women are less likely to take up a screening opportunity if they are unfamiliar with the doctor, whereas doctors are more likely to recommend screening to women who are not their regular patients.

While the results suggest that providers are not responsive to financial incentives (and indeed may respond perversely to them) caution needs to be applied in interpreting this result. It may suggest that a choice experiment, which relies on self report of intended behaviour, is not an effective mechanism to assess the impact of financial incentives of this type.

The results suggest that changes to the recommended screening interval are not likely to result in perverse behaviour (contrary to recommendations) for either women or general practitioners. Thus, the concern that a change to the screening interval may undermine confidence in the screening program appears unfounded. Both doctors and patients appear likely to follow national recommendations although, as we have noted, the doctors were less likely to screen than women when the recommendation was one year.

While there has been considerable discussion over the past decade about the merits of a more accurate screening test, these results do not indicate strong support among either women or GPs for the liquid based test, especially if it is more expensive. While both groups have a preference for increased accuracy, this does not translate into a preference for the liquid based test per se, in the sense that the labelled alternative of the liquid based test was chose in less than 25% of choice sets for both women and general practitioners.

At present we have only considered the preferences of women who have previously participated in screening. An important extension of the current work will be to examine the preferences of the never-screened women.

**References**


1. Anderson R, Shanahan M, Haas M, De Abreu Lourenco R, Saville M. Cost effectiveness evaluation of the national cervical screening program. CHERE: Sydney, 2004.
2. Arana JE, Leon CJ, Quevedo JL. The effect of medical experience on the economic evaluation of health policies. A discrete choice experiment. Social Science & Medicine 2006; 63: 512-524.
3. Austin RM, Ramzy I. Increased detection of epithelial cell abnormalities by liquid-based gynecologic cytology preparations. A review of accumulated data. Acta Cytol 1998; 42: 178-184.
4. Australian Health Technology Advisory Committee. Review of automated and semi-automated cervical screening devices. Commonwealth Department of Health and Family Services: Canberra, 1998.
5. Australian Institute of Health and Welfare. Cervical screening in Australia 2002-2003. (aihw cat. No. 26; cancer series no. 32). Canberra, 2005.
6. Australian Institute of Health and Welfare. Cervical screening in Australia 2001-2002. AIHW cat. no. CAN 22. Australian Institute of Health and Welfare: Canberra, 2004.
7. Australian Institute of Health and Welfare (AIHW). Breast and cervical cancer screening in Australia 1996-1997. AIHW: Canberra, 1998.
8. Bartels R, Fiebig DG, Soest A. Consumers and experts: An econometric analysis of the demand for water heaters Empirical Economics 2006; 31: 369-391.
9. Belkar R, Fiebig DG, Haas M, Viney R. Why worry about awareness in choice problems? Econometric analysis of screening for cervical cancer. Health Econ 2006; 15 1: 33-47.
10. Bjorge T, Gunbjorud AB, Langmark F, Skare GB, Thoresen SO. Cervical mass screening in norway--510,000 smears a year. Cancer Detection & Prevention 1994; 18: 463-470.
11. Boyce JG, Fruchter RG, Romanzi L, Sillman FH, Maiman M. The fallacy of the screening interval for cervical smears. Obstetrics & Gynecology 1990; 76: 627-632.
12. Burgess L, Street D. Optimal designs for asymmetric choice experiments. Department of Mathematical Sciences, UTS: Sydney, 2003.
13. Bush J. "It's just part of being a woman": Cervical screening, the body and femininity. Social Science & Medicine 2000; 50: 429-444.
14. Carson RT. Contingent valuation: A user's guide. Evnvironmental Science & Technology 2000; 34: 1413-1418.
15. Cruickshank ME, Angus V, Kelly M, McPhee S, Kitchener HC. The case for stopping cervical screening at age 50. British Journal of Obstetrics & Gynaecology 1997; 104: 586-589.
16. Culyer AJ. The normative economics of health care finance and provision. Oxford Review of Economic Policy 1989; 5: 34-58.
17. Cuzick J. Screening for cancer: Future potential. Eur J Cancer 1999; 35: 1925-1932.
18. Fahey MT, Irwig L, Macaskill P. Meta-analysis of pap test accuracy. Am J Epidemiol 1995; 141: 680-689.
19. Fiebig D, Hall J. Discrete choice experiments in the analysis of health policy. In Quantitative tools for microeconomic policy analysis. Productivity Commission: Canberra, 2005; 119-136.

20. Grant CM. Cervical screening interval: Costing the options in one health authority. J Public Health Med 1999; 21: 140-144.
21. Hall J, Fiebig DG, King MT, Hossain I, Louviere JJ. What influences participation in genetic carrier testing? Results from a discrete choice experiment. J Health Econ 2006; 25: 520-537.
22. Hennig P, Knowles A. Factors influencing women over 40 years to take precautions against cervical cancer. Journal of Applied Social Psychology 1990; 20: 1612-1621.
23. Law MR, Morris JK, Wald NJ. The importance of age in screening for cancer. J Med Screen 1999; 6: 16-20.
24. McFadden D, Train K. Mixed mnl models for discrete response. Journal of Applied Econometrics 2000; 15: 447-470.
25. McKie L. Women's views of the cervical smear test: Implications for nursing practice--women who have had a smear test. J Adv Nurs 1993; 18: 1228-1234.
26. Milburn K, MacAskill S. Cervical screening: Continuing concerns in the 1990s. Health Educ J 1994; 53: 201-213.
27. Mitchell H. Report disclaimers and informed expectations about papanicolaou smears: An Australian view. Arch Pathol Lab Med 1997; 121: 327-330.
28. Mitchell H, Higgins V, Burrows C. Victorian cervical cytology report 1999. Victorian Cervical Cytology Registry: Melbourne, 2000.
29. Payne N, Chilcott J, McGoogan E. Liquid-based cytology in cervical screening: A rapid and systematic review. Health Technol Assess 2000; 4.
30. Revelt D, Train K. Mixed logit with repeated choices: Households' choices of appliance efficiency level. Rev Econ Stat 1998; 80: 647-657.
31. Soost HJ, Lange HJ, Lemacher W, Ruffing-Kullmann B. The validation of cervical cytology. Sensitivity, specificity and predictove values. Acta Cytol 1991; 35: 8-14.
32. Street D, Burgess L. Optimal and near-optimal pairs for the estimation of effects in 2-level choice experiments. Journal of Statistical Planning and Inference 2004a; 118: 185-199.
33. Street D, Burgess L. Optimal stated preference choice experiments when all choice sets contain a specific option. Statistical Methodology 2004b; 1: 37-45.
34. Train K. Discrete choice methods with simulation. Cambridge University Press: Cambridge, 2003.
35. Mixed logit estimation for panel data using maximum simulated likelihood. http://elsa.berkeley.edu/Software/abstracts/train0296.html[December 15].
36. Train KE, Herriges JA, Kling CL. Mixed logit models for recreation demand. In Valuing recreation and the environment: Revealed preference methods in theory and practice, Herriges JA, Kling CL, (eds). Edward Elgar: Cheltenham, UK, 1999; 121-140.
37. van Oortmarssen G, Habbema J. Epidemiologocal evidence for age-dependent regression of pre-invasive cancer. Br J Cancer 1991; 64: 559-565.
38. Van Wijngaarden WJ, Duncan ID. Rationale for stopping cervical screening in women over 50. BMJ 1993; 306: 967-971.
39. Vick S, Scott A. Agency in health care. Examining patients' preferences for attributes of the doctor-patient relationship. J Health Econ 1998; 17: 587-605.

## Table 1: Example of a choice task for the women's sample

You are visiting the GP who gives you some information about Pap tests and raises the issue of having a Pap test.

**About this GP:**

| | |
|---|---|
| This GP is | your regular GP who you usually see for most care, including Pap tests |
| This GP is | Female |
| This GP's practice will receive a special incentive payment if you have a Pap test at this visit | No |

**About the tests available:**

| | Standard Pap test | Liquid based Pap test |
|---|---|---|
| The out of pocket costs to you for this test will be | $0 | $20 |
| The chance that this test will give you a false negative result is | 1 in 20 | 1 in 33 |
| The chance that this test will give you a false positive result is | 1 in 1000 | 1 in 500 |

**Other information the GP gives you about cervical screening:**

| | |
|---|---|
| The GP tells you that you had your last Pap test | about 1 year ago |
| The national recommendation is that women should have a Pap test | every 1 year |
| If you have either Pap test you can at the same time have an HPV test at an additional out-of-pocket cost to you of | $50 |
| The GP recommends that | you do not have a Pap test at this visit |
| The GP recommends that you | do not have the HPV test |

At this visit to the GP what would you choose to do?

| | |
|---|---|
| *Circle the number next to your choice*<br><br>**I would not have a cervical cancer screening test**<br><br>**I would have a standard Pap test**<br><br>**I would have the liquid based Pap test** | <br><br>**1**<br><br>**2**<br><br>**3** |
| *Circle Yes or No to show your choice*<br><br>**If you chose to have a Pap test, would you also have the HPV test at this visit?** | **Yes**<br><br>**No** |

## Table 2: Example of choice task for the general practitioners sample

**A female patient is attending your practice for a primary care consultation**

**About this patient:**

| | |
|---|---|
| This patient is attending the consultation | for a minor health problem |
| This patient is | a patient who has previously consulted your practice but has not consulted you |
| This patient last had a Pap test | about 3 years ago |
| This patient is aged | Less than 20 |
| In your perception this patient is in | the middle income/SES range |

**About the tests available:**

| | Standard Pap test | Liquid based Pap test |
|---|---|---|
| The out-of-pocket costs to the patient for this test will be | $0 | $20 |
| The chance that this test will give a false negative result is | 1 in 10 | 1 in 10 |
| The chance that this test will give a false positive result is | 1 in 150 | 1 in 100 |

**Other information about cervical screening:**

| | |
|---|---|
| The national recommendation is that women should have a Pap test | every 3 years |
| If the patient has a Pap test at this consultation, your practice will receive | a standard consultation fee and an incentive payment if the patient has a Pap test at the recommended screening interval |
| At the same time that the patient has a Pap test it is possible for her to have an HPV test at an additional cost of | $150 |

**What would you recommend to this patient with regard to a Pap test at this consultation?**

| | |
|---|---|
| *Circle the number next to your choice*<br>**I would not recommend the patient have a cervical cancer screening test at this consultation** | **1** |
| **I would recommend the patient have a standard Pap test at this consultation** | **2** |
| **I would recommend the patient have a liquid based Pap test at this consultation** | **3** |

| | |
|---|---|
| *Circle Yes or No to show your choice*<br>**If you recommended that the patient have a Pap test, would you also recommend she have the HPV test at this visit?** | **Yes**<br><br>**No** |

*Table 3: Attributes, levels and corresponding names for the women's survey*

| **Context Attributes** | **Levels** | **Names** |
|---|---|---|
| The recommended screening interval | • 1 year | recint1 |
| | • 2 years | Base |
| | • 3 years | recint2 |
| | • 5 years | recint3 |
| This general practitioner is | • Your regular GP who you usually see for most care, including Pap tests; | Base |
| | • A GP you do not usually see, or have not seen before | Knowgp |
| This general practitioner is | • Female | Base |
| | • Male | Sexgp |
| Your doctor tells you had your last cervical screening test | • About 1 year ago | Base |
| | • About 2 years ago | testdue1 |
| | • About 3 years ago | testdue2 |
| | • About 5 years ago | testdue3 |
| Your doctor recommends that | • You do not have a cervical screening test today | Base |
| | • You have the standard Pap test today | drrrec1 |
| | • You have the liquid based cytology Pap test today | drrrec1 |
| | • You have a cervical screening test today but does not recommend one test over the other | drrrec2 |
| Your doctor will receive a special payment if you have a cervical screening test | • No | Base |
| | • Yes | Drink |

| Alternative specific attributes | **Levels** | | **Names** |
|---|---|---|---|
| | **Standard Pap test** | **Liquid based Pap test** | |
| Cost of the test to the patient | \$0, \$10, \$20, \$30 (A) | A+\$10, A+\$20, A+\$30, A+\$40 | Cost |
| The chance that this test will give a false negative result | 1 in 20 (5%), 1 in 15 (6.67%), 1 in 10 (10%), 1 in 5 (20%) | 1 in 100 (1%), 1 in 33 (3.03%), 1 in 20 (5%), 1 in 10 (10%) | Fneg |

| | | | |
|---|---|---|---|
| The chance that this test will give a false positive result | 1 in 1000 (0.1%), 1 in 250 (0.4%), 1 in 150 (0.67%), 1 in 100 (1%) | 1 in 2000 (0.05%), 1 in 500 (0.2%), 1 in 150 (0.67%), 1 in 100 (1%) | Fpos |

Notes:

- The attributes and levels pertaining to the HPV test are not reported in this table as they are not used in the analysis presented in this paper. For completeness, they appear in the example choice task (table 1).
- All variables, except cost ($), fneg (%) and fpos (%), have been dummy coded in the estimation.

*Table 4: Attributes, levels and corresponding names for the GP survey*

| Context Attributes | Levels | Names |
|---|---|---|
| The patient is attending this consultation | • Specifically for a cervical screening test | Base |
| | • For a general check-up | consult1 |
| | • For a minor health problem | consult2 |
| | • For a serious health problem | consult3 |
| The recommended screening interval is every | • 1 year | recint1 |
| | • 2 years | Base |
| | • 3 years | recint2 |
| | • 5 years | recint3 |
| This patient is | • a patient who consults you for most of her primary health care including Pap tests | Base |
| | • a patient who has consulted you regularly but not previously for a Pap test | knowpat1 |
| | • a patient who has previously consulted your practice, but has not consulted you | knowpat2 |
| | • a new patient to your practice | knowpat3 |
| This patient last had a cervical screening test | • About 1 year ago | Base |
| | • About 2 years ago | testdue1 |
| | • About 3 years ago | testdue2 |
| | • Never | testdue3 |
| The age of this patient is | • Less than 20 years old | Base |
| | • 20-29 | agepat1 |
| | • 30-59 | agepat2 |
| | • 60-69 | agepat3 |
| | • 70 or more years old | agepat4 |
| Your perception of this patient's household income | • In the lowest income/SES range | Base |
| | • In the low-middle income/SES range | sespat1 |
| | • In the middle income/SES range | sespat2 |
| | • In the highest income/SES range | sespat3 |
| Payment to your practice for a Pap test is | • A standard consultation fee only | Base |
| | • A standard consultation fee, and an incentive payment if the woman is overdue for a Pap test by 12 months or more | payment1 |
| | • A standard consultation fee, and an incentive payment if the woman has a Pap test at the recommended screening interval | payment2 |
| | • A standard consultation fee, and an incentive payment if the practice achieves a screening rate of 70% or more of the eligible patient population | payment3 |

| Alternative specific attributes | **Levels** | | **Names** |
|---|---|---|---|
| | **Standard Pap test** | **Liquid based Pap test** | |
| Cost of the test to the patient | $0, $10, $20, $30 (A) | A+$10, A+$20, A+$30, A+$40 | Cost |
| The chance that this test will give a false negative result | 1 in 20 (5%), 1 in 15 (6.67%), 1 in 10 (10%), 1 in 5 (20%) | 1 in 100 (1%), 1 in 33 (3.03%), 1 in 20 (5%), 1 in 10 (10%) | Fneg |
| The chance that this test will give a false positive result | 1 in 1000 (0.1%), 1 in 250 (0.4%), 1 in 150 (0.67%), 1 in 100 (1%) | 1 in 2000 (0.05%), 1 in 500 (0.2%), 1 in 150 (0.67%), 1 in 100 (1%) | Fpos |

Notes:

- The attributes and levels pertaining to the HPV test are not reported in this table as they are not used in the analysis presented in this paper. For completeness, they appear in the example choice task (table 2).
- All variables, except cost ($), fneg (%) and fpos (%), have been dummy coded in the estimation.

**Table 5: Sociodemographic variables, levels, and corresponding names, means or percentages by sample**

| Variables | Levels | Names | Means/percentages |
|---|---|---|---|
| | **Women** | | |
| Age | In years | age | 42 |
| Education | • Some primary to completed secondary | Base | 54% |
| | • Trade | ed_trade | 16% |
| | • Some university | ed_some uni | 10% |
| | • Completed university | ed_high | 20% |
| Household Income | • Up to $50,000 | Base | 63% |
| | • $50,001 to $80,000 | inc_med | 16% |
| | • Over $80,000 | inc_high | 14% |
| | • Income not reported | inc_mis | 7% |
| Country of birth | • Australia | Base | 86% |
| | • Other | cob | 14% |
| Smoking | • Never regular | Base | 54% |
| | • Current smoker | smoke_curr | 23% |
| | • Ex-smoker | smoke_ex | 23% |
| | | | |
| | **GP** | | |
| Recommendation of tests to patients having a cervical screening test in the past year | • Mostly liquid based pap test | recom_liq | 14% |
| | • Mostly standard pap test | recom_std | 55% |
| | • Both tests about equally | recom_eq | 22% |
| | • No preference of one test over another | Base | 9% |
| Gender | • Female | Base | 54% |
| | • Male | gender | 46% |
| GP practice years | • Less than 1 year | prac1 | 7% |
| | • 1-5 years | prac2 | 20% |
| | • 6-10 years | Base | 18% |
| | • 11-20 years | prac3 | 28% |
| | • More than 20 years | prac4 | 27% |

Note: All variables, except age, have been dummy coded in the estimation.

**Table 6: Choice response frequencies**

| | **Women** | | **GPs** | |
|---|---|---|---|---|
| | Freq | % | Freq | % |
| No test | 1991 | 37.3% | 2947 | 42.8% |
| Standard pap test | 2088 | 39.1% | 2232 | 32.4% |
| Liquid based test | 1265 | 23.7% | 1701 | 24.7% |
| Total | 5344 | 100.0% | 6880 | 100.0% |

**Table 7: Log-likelihood values for MNL and MXL Models**

| **Model** | **Log-Likelihood** | |
|---|---|---|
| | *Women* | *GPs* |
| MNL | -5240.97 | -5491.35 |
| MXL (test & standard Pap intercepts) | -3844.58 | -4935.48 |
| MXL (test & liquid Pap intercepts) | -3842.12 | -4886.78 |

**Table 8: Estimated results for the Mixed logit specification for women**

| Variables | | Estimates | p-value |
|---|---|---|---|
| *Intercepts* | | | |
| either test | Mean | 0.7054 | 0.2738 |
| | Std. Deviation | 2.3273 | <0.0000 |
| liquid test | Mean | -1.1300 | 0.0005 |
| | Std. Deviation | 2.5850 | <0.0000 |
| *Context attributes* | | | |
| sexgp | | -0.9210 | <0.0000 |
| knowgp | | -0.9846 | <0.0000 |
| recint1 | | 0.4111 | <0.0000 |
| recint2 | | -0.4234 | <0.0000 |
| recint3 | | -1.0818 | <0.0000 |
| testdue1 | | 1.1063 | <0.0000 |
| testdue2 | | 1.4281 | <0.0000 |
| testdue3 | | 2.0718 | <0.0000 |
| drrec1[a] | | 0.8869 | <0.0000 |
| drrec2 | | 0.7633 | <0.0000 |
| drinc | | -0.0656 | 0.2844 |
| *Alternative specific attributes* | | | |
| cost | | -0.0268 | <0.0000 |
| fneg | | -0.0321 | <0.0000 |
| fpos | | -0.4504 | <0.0000 |
| *Sociodemographic attributes* | | | |
| age | | 0.0169 | 0.1364 |
| ed_trade | | -0.7230 | 0.0718 |
| ed_some uni | | -0.6057 | 0.2447 |
| ed_high | | -0.3050 | 0.5217 |
| inc_med | Standard test[b] | 0.4845 | 0.4389 |
| inc_med | Liquid test | 1.8528 | <0.0000 |
| inc_high | Standard test | -0.0591 | 0.9047 |
| inc_high | Liquid test | 1.8208 | 0.0079 |
| inc_mis | Standard test | -1.2123 | 0.0393 |
| inc_mis | Liquid test | -1.9796 | <0.0000 |
| Cob | | 0.1262 | 0.7786 |
| smoke_curr | | -0.1959 | 0.7173 |
| smoke_ex | | -0.4472 | 0.2003 |

Notes:

- Refer to the tables 3 and 5 for more descriptions of the variables.
- [a] For the utility function of the standard test, drrec = 1 if the doctor recommended standard test and for the utility function of the liquid test, drrec = 1 if the doctor recommended liquid test.
- [b] Modeling the effect of the corresponding income level on this specific test.

**Table 9: Estimated results for the Mixed logit specification for GPs**

| Variables | | Estimates | p-value |
|---|---|---|---|
| *Intercepts* | | | |
| either test | Mean | 1.6301 | <0.0000 |
| | Std. Deviation | 1.1312 | <0.0000 |
| liquid test | Mean | -0.9906 | <0.0000 |
| | Std. Deviation | 1.5690 | <0.0000 |
| *Context attributes* | | | |
| consult1 | | -2.2893 | <0.0000 |
| consult2 | | -1.9908 | <0.0000 |
| consult3 | | -3.7900 | <0.0000 |
| knowpat1 | | 0.6137 | <0.0000 |
| knowpat2 | | 0.3018 | 0.0060 |
| knowpat3 | | 0.3003 | 0.0006 |
| testdue1 | | 1.7285 | <0.0000 |
| testdue2 | | 2.7981 | <0.0000 |
| testdue3 | | 3.7663 | <0.0000 |
| agepat1 | | -0.5294 | 0.0001 |
| agepat2 | | 0.1007 | 0.3752 |
| agepat3 | | -0.9762 | <0.0000 |
| agepat4 | | -2.3698 | <0.0000 |
| recint1 | | 0.4013 | 0.0001 |
| recint2 | | -0.7646 | <0.0000 |
| recint3 | | -1.6122 | <0.0000 |
| payment1 | | 0.1132 | 0.2344 |
| payment2 | | -0.0909 | 0.3930 |
| payment3 | | -0.4290 | <0.0000 |
| sespat1 | | -0.2895 | 0.0042 |
| sespat2 | | -0.0842 | 0.4134 |
| sespat3 | | -0.3030 | 0.0038 |
| *Alternative specific attributes* | | | |
| Cost | | -0.0134 | <0.0000 |
| Fneg | | -0.0586 | <0.0000 |
| fpos | | -0.3471 | <0.0000 |
| *Sociodemographic attributes* | | | |
| gender | | 0.0055 | 0.9753 |
| prac1 | | -1.2120 | 0.0006 |
| prac2 | | -0.1190 | 0.6051 |
| prac3 | | -0.0387 | 0.8528 |
| prac4 | | -0.2888 | 0.2480 |
| recom_std | Standard test [a] | 0.6425 | 0.0013 |
| recom_liq | Liquid test | 2.4135 | <0.0000 |
| recom_eq | Standard test | 0.3551 | 0.2082 |
| recom_eq | Liquid test | 1.7783 | <0.0000 |

Notes:

- Refer to the tables 4 and 5 for more descriptions of the variables.
- [a] Modeling the effect of the corresponding recommendation level on this specific test.

## Table 10: Comparing MXL estimates of common attributes

| | Estimates | | | Rates of substitution | | |
|---|---|---|---|---|---|---|
| **Attribute** | **GP** | **Women** | **Ratio*** | **GP** | **Women** | **Ratio*** |
| False negative | -0.06 | -0.03 | 0.55 | 1.00 | 1.00 | 1.00 |
| False positive | -0.35 | -0.45 | 1.30 | 5.92 | 14.02 | 2.37 |
| Cost | -0.01 | -0.03 | 1.99 | 0.23 | 0.83 | 3.64 |
| Interval 1 yr | 0.40 | 0.41 | 1.02 | -6.84 | -12.80 | 1.87 |
| Interval 3 yrs | -0.76 | -0.42 | 0.55 | 13.04 | 13.18 | 1.01 |
| Interval 5 yrs | -1.61 | -1.08 | 0.67 | 27.49 | 33.68 | 1.23 |
| Last test 2 yrs | 1.73 | 1.11 | 0.64 | -29.47 | -34.44 | 1.17 |
| Last test 3 yrs | 2.80 | 1.43 | 0.51 | -47.71 | -44.46 | 0.93 |

* Women's value divided by GP value