VALUE IN HEALTH 21 (2018) 219–228

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

PREFERENCE-BASED ASSESSMENTS

Investigating the Heterogeneity in Women's Preferences for Breast Screening: Does the Communication of Risk Matter?

Caroline M. Vass, PhD[1], *Dan Rigby, PhD*[2], *Katherine Payne, PhD*[1,*]

[1]*Manchester Centre for Health Economics, The University of Manchester, Manchester, UK;* [2]*Department of Economics, The University of Manchester, Manchester, UK*

ABSTRACT

Background: The relative benefits and risks of screening programs for breast cancer have been extensively debated. **Objectives:** To quantify and investigate heterogeneity in women's preferences for the benefits and risks of a national breast screening program (NBSP) and to understand the effect of risk communication format on these preferences. **Methods:** An online discrete choice experiment survey was designed to elicit preferences from female members of the public for an NBSP described by three attributes (probability of detecting a cancer, risk of unnecessary follow-up, and out-of-pocket screening costs). Survey respondents were randomized to one of two surveys, presenting risk either as percentages only or as icon arrays and percentages. Respondents were required to choose between two hypothetical NBSPs or no screening in 11 choice sets generated using a Bayesian D-efficient design. The trade-offs women made were analyzed using heteroskedastic conditional logit and scale-adjusted latent class models. **Results:** A total of 1018 women completed the discrete choice experiment (percentages-only version = 507; icon arrays and percentages version = 511). The results of the heteroskedastic conditional logit model suggested that, on average, women were willing-to-accept 1.72 (confidence interval 1.47–1.97) additional unnecessary follow-ups and willing-to-pay £79.17 (confidence interval £66.98–£91.35) for an additional cancer detected per 100 women screened. Latent class analysis indicated substantial heterogeneity in preferences with six latent classes and three scale classes providing the best fit. The risk communication format received was not a predictor of scale class or preference class membership. **Conclusions:** Most women were willing to trade-off the benefits and risks of screening, but decision makers seeking to improve uptake should consider the disparate needs of women when configuring services.
Keywords: breast screening, discrete choice experiment, risk, willingness-to-pay.

Introduction

Breast cancer is the most common cancer in women [1]. Because of the burden and high expense of cancer care, which is an estimated £15 billion annually in the United Kingdom, many Western countries now encourage participation in screening for common malignant diseases [2]. In England, the National Health Service currently invites all women between the ages of 50 and 70 years for screening using mammography every 3 years as part of the National Health Service national breast screening program (NBSP). This and similar NBSPs provided in other countries are based on the premise that regular screening can identify tumors of the breast and ensure that therapy commences as soon as possible [3].

Screening for breast cancer via mammography has been proven to detect cases of breast cancer earlier [4], and women who participate in NBSPs have been shown to have improved mortality rates because of earlier intervention [5]. Nevertheless, because the mammogram produces an image that is interpreted by a radiographer, there is a chance of a cancer being missed (a false-negative) [6]. There is also a risk that the image will locate either dense breast tissue that is not cancerous (a false-positive) or a true cancer but one that is so slow-growing that it would never have been harmful in the woman's lifetime (termed "overdiagnosis") [7]. The potential for overdiagnosis means women may be recalled for unnecessary tests and biopsies [8]. Whether an NBSP causes more harm than good has been extensively debated by clinicians and academics [3,9–12]. Despite this, few studies [13,14] have quantified women's preferences for the benefits and risks associated with breast cancer screening.

Discrete choice experiments (DCEs) are a commonly used method of quantifying preferences for health care programs [15,16]. DCEs are a survey-based method, underpinned by economic theories [17,18], in which respondents choose their hypothetically preferred option from a choice set comprising a series of discrete options (typically products, programs, or policies), defined in terms of attributes that differ in their levels. Respondents are assumed to trade-off the levels of the attributes in

Conflicts of interest: The authors have no conflicts of interest.

*Address correspondence to: Katherine Payne, Manchester Centre for Health Economics, The University of Manchester, Oxford Road, Manchester M13 9PL, UK.

E-mail: katherine.payne@manchester.ac.uk.

1098-3015$36.00 – see front matter Copyright © 2018, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

http://dx.doi.org/10.1016/j.jval.2017.07.010

choosing the option yielding the greatest satisfaction or "utility" with a degree of randomness due to unobserved factors [19]. The choices made can be analyzed to reveal the strength of preference they have for the attributes [20].

Systematic reviews of the health care literature have shown that DCEs are increasingly used to elicit preferences for benefits and risks [16,21]. There has also been recent acknowledgment of their usefulness for regulatory decision making about the levels of risk that consumers of health care interventions will tolerate for an associated benefit [22,23]. Nevertheless, numerical probabilistic information is a notoriously difficult concept to communicate [24]. If DCE respondents do not understand the choice task, they may use simplifying heuristics, such as ignoring confusing attributes, violating the axiom of continuity in preferences resulting in biased preference estimates. Reviews of health care DCEs have found risk to be a neglected attribute [16,21]. A systematic review [21] of this literature identified that risk attributes were most commonly communicated quantitatively, most often as a numerical percentage. This is in contrast to recommendations in the risk communication literature that advocate pictures and/or graphics [25,26], which were used in 27 (23%) health care DCEs.

Quantification of women's preferences for benefits and risks may provide a useful contribution to the debate about the relative merits and harms of an NBSP. In the analysis of DCE data there is an increasing focus on heterogeneity in preference—both its determinants and its implications [27]—because the preferences of the "average" person can be of limited value. Decision makers presented with average preference data, from a sample of respondents, without any idea of the proportion of individuals feeling that way or that the range of these values may have an incomplete view that could hamper generalizations from the study results to the relevant population. This study aimed to examine the degree of heterogeneity in women's preferences for the benefits and risks associated with an NBSP.

To collect reliable preference data, the elicitation method used must be robust to formatting effects. In breast screening, the communication of risk in invitation leaflets has been controversial [28,29], with some suggesting that the format of risk may affect uptake. A secondary aim of this study was to understand whether preferences were affected by the risk communication format used.

Methods

This study used an online DCE to elicit women's preferences for a hypothetical NBSP. Approval for the study was obtained from The University of Manchester's Research Ethics Committee (AJ/ethics/1809/13/ref13178). The study was designed and reported in line with published recommendations [30,31].

Attributes and Levels

Attributes were identified through an iterative process of interviews with clinical experts (n = 4), a patient representative (n = 1) and female members of the public (n = 4), and reviews of the breast screening and DCE literatures. Levels were assigned through literature reviews and consultation with experts (n = 4) to determine a plausible and clinically relevant range. Table 1 presents the attributes and the levels used in the DCE.

The levels for out-of-pocket cost were chosen to reflect the costs associated with taking time off from work and traveling to a screening center (personal communication, Ian Jacob, 2013) and a realistic maximum based on the price of a private mammogram in the United Kingdom [32]. The attribute "probability of detecting a cancer" was assigned a range of 20 years on the basis of a study that found that the average woman entering screening at age 50 years had about a 3.5% probability of detecting a cancer [33]. Other levels were chosen to reflect detection rates achievable through stratified or more frequent screening [34]. Discussions with experts (n = 4) identified "unnecessary follow-up" to be the most pertinent and accurate representation of the downside risk of screening, rather than "overdiagnosis" or "unnecessary treatment." The attribute was assigned levels on the basis of the results of the Independent Review of Breast Cancer Screening [35], which estimated that just over 1% of women invited for screening would receive unnecessary follow-up, and a review of the Norwegian screening program that estimated that false recalls after mammography could be as high as 20% [8].

Experimental Design and Questionnaire

Fractional factorial designs can be used when there are too many possible profile combinations of attribute levels [36]. In this study, a fractional factorial design for the DCE was chosen to reduce the number of choice sets and, in turn, respondent fatigue. An experimental design minimizing the D error was generated using Ngene (ChoiceMetrics) [37] originally with conjectured priors, updated after a pilot study. The alternatives created were split into four blocks containing 11 choice sets, guided by the pilot study, including a check for monotonic preferences to verify that the respondents were answering in line with economic theory. Because screening is voluntary and uptake to the NBSP

Table 1 – Attributes and levels used in the DCE.

Label	Attribute	Definition	Levels for programs	Levels for opt-out of "no screening"
Detect	Probability of detecting a cancer	The chance of detecting a cancer from screening over a 20-y period	3%, 7%, 10%, 14%	None: no cancers detected (0%)
Risk	Risk of unnecessary follow-up	The probability of being recalled for a procedure or procedures when no harm existed	0%, 1%, 5%, 10%, 20%	None: no unnecessary follow-ups (0%)
Cost	Out-of-pocket cost of screening over a lifetime	The costs of attending the program including original screens and recalls; these could include transport, time off from work, and carer costs	£100 (£20 per screen); £250 (£50 per screen); £750 (£150 per screen); £1000 (£200 per screen)	No cost to you (£0)

DCE, discrete choice experiment.

in the United Kingdom is considerably less than 100% [38], a forced choice experiment would be inappropriate and so respondents were asked to choose between two formulations (A and B) of the screening program and an opt-out alternative of "no screening."

To understand whether the means by which risk information is presented affects preferences for screening, two versions of the DCE were created, with risk attributes framed either as percentages only (PO) or as icon arrays and percentages (IAP). Examples of the choice sets are shown in Figure 1 depicting the PO and IAP versions. The icon array graphics were created by the Risk Science Center at the University of Michigan [39]. The images used to present the levels in either icon arrays or percentages were identical in size (to the nearest pixel), allowing for comparison in a subsequent eye-tracking study [40].

Constructing the Survey

The survey was designed and presented online using SSI Web 8.3.8 (Sawtooth Software, Orem, UT) [41]. A pilot study comprising 1)

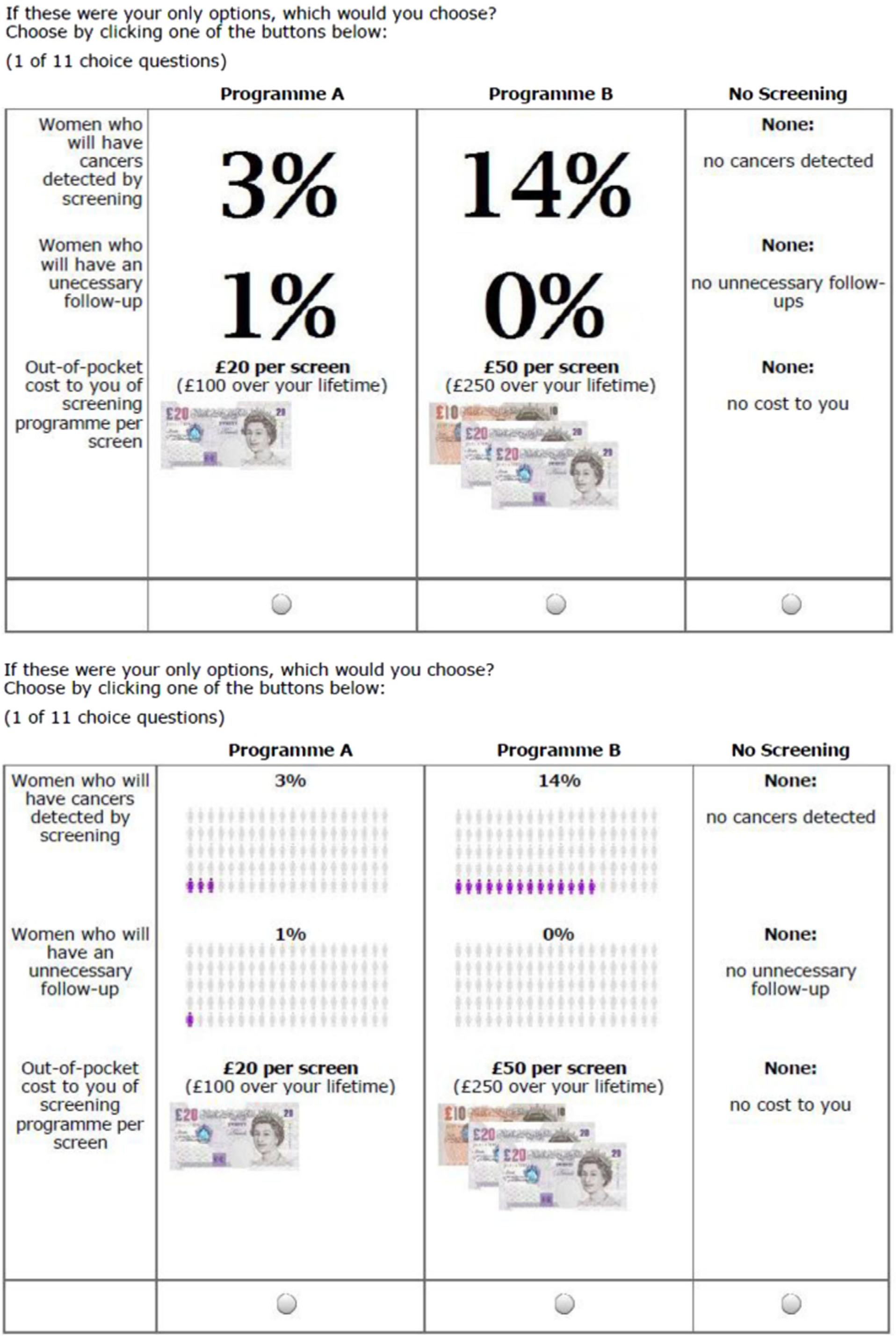

Fig. 1 – Example choice questions with risk communicated as a percentage only or as icon arrays and percentages.

in-depth face-to-face interviews using the think-aloud method with female members of the public (n = 4) as well as a patient representative (n = 1) and 2) an online survey with female members of the public (n = 56), recruited using ResearchNow® (an internet panel provider), was conducted. The experimental design was updated using priors from the pilot study data. The final survey comprised five sections containing training materials to explain the purpose of the survey, descriptions of each attribute, and a visual representation and video explaining the NBSP; choice set questions; sociodemographic questions; feedback on their experience of making choices in the survey; and validated questions to understand the respondents' risk attitudes and numeracy skills [42–44]. Women were randomly allocated to receive one of the two communication formats upon clicking the link to enter the survey.

Study Sample

The link to the online survey was sent to female members of the public, aged 18 to 70 years, recruited through ResearchNow. The views of women in screening (who may have better formed preferences) and women about to enter screening (with regard to upcoming changes in policy) were felt to be of particular interest to the interviewed experts (n = 4). Therefore, age bands (45–49 years and older than 50 years) were oversampled.

Statistical Analyses

The choice data were analyzed within a random utility maximization framework [17]. The aim of the analysis was to quantify the relative importance of each attribute and establish whether this was affected by the risk communication format. A secondary aim was to investigate whether heterogeneity in preferences was present and identify subgroups or "latent" classes driving any heterogeneity. There are two main types of heterogeneity in DCEs: preference heterogeneity (the degree to which preferences vary across respondents) and scale heterogeneity (the variation in the error term of responses) [45]. The analysis took account of both types of heterogeneity.

Scale heterogeneity

The data generated by this DCE came from two survey versions (PO and IAP). Standard conditional logit models assume that error variance (scale) is constant across individuals [46] because the scale parameter cannot be separated from some other weights (such as preference weights) in the utility function. Assuming it is constant allows interpretation of the estimated coefficients as pure preference weights [47]. Because the two samples were randomized to different risk communication formats, the estimated parameters may not reflect true differences in preferences if the preference coefficients are confounded with the potentially different scale parameters. The choice data in this study were therefore analyzed using models that allowed for scale heterogeneity.

First, the heteroskedastic conditional logit (HCL) model [48] was estimated on the basis of the following utility function:

$$U_{nj} = \beta_{\text{none}} + \lambda_n\beta_1\,\text{Detect}_{nj} + \lambda_n\beta_2\text{Risk}_{nj} + \lambda_n\beta_3\text{Cost}_{nj} + \lambda_n\beta_4\,\text{Detect}_{nj}\text{IAP}_n + \lambda_n\beta_5\text{Risk}_{nj}\text{IAP}_n + \varepsilon_{nj}, \tag{1}$$

where U represents an individual's (n) indirect utility for an alternative (j); β_{none} is an alternative specific constant (ASC) for the opt-out that captures differences in the mean of the distribution of the unobserved effects in the random component, ε_{nj}, between the opt-out and the other alternatives; and β_{1-3} are preference weights associated with each attribute. IAP_n is a dummy variable that equaled unity if individual n received the IAP version of the DCE. $\text{Detect}_{n\,j}\text{IAP}_n$ and $\text{Risk}_{n\,j}\text{IAP}_n$ represent interaction terms between the probability of detecting a cancer and the probability of unnecessary follow-up, with the dummy variable indicating the risk format received, respectively. Testing for significance of these interaction terms is a test of the effect of the risk communication format on preference weights.

λ is the scale parameter, which is inversely proportional to the variance of the error process, σ_ε^2, by:

$$\lambda_n = \frac{\pi}{\sqrt{6\sigma_\varepsilon^2}}. \tag{2}$$

Icon arrays are hypothesized to reduce cognitive load, which could therefore improve the DCE respondents' choice consistency by making the task simpler and therefore reducing the error variance. The scale parameter is therefore permitted to vary by the communication format allocated to, and it is modeled as follows:

$$\lambda_n = \exp(\gamma\text{IAP}_n). \tag{3}$$

Testing the significance of γ is therefore a test of whether risk communication format affects choice consistency.

Continuous and effects-coded specifications for all attributes were examined to evaluate the assumption of linearity in preferences.

Preference heterogeneity

Scale-adjusted latent class (SALC) analysis [49] was used in this study to allow investigation of both scale heterogeneity and preference heterogeneity. The SALC model allowed for analysis of heterogeneity in scale from latent sources in addition to the identification of latent preference classes. Within each latent preference class, different scale classes (each with relatively different error variances) were present and each associated with a scale membership probability. The SALC model also allowed for covariates (collected in the background questions in the survey) to enter preference and scale class assignment capturing observed factors that might influence preferences and/or choice consistency. The number of scale and preference classes was selected through an iterative process involving comparisons of the Akaike information criterion, the Bayesian information criterion (BIC), and the consistent Akaike information criterion [50]. The HCL and SALC models were estimated using Stata (StataCorp, College Station, TX) [51] and Latent GOLD (Statistical Innovations, Belmont, MA) [52] software, respectively.

Quantifying the Balance between Benefits and Risks

The balance between benefits and risks was quantified through estimating marginal rates of substitution (MRS), representing how much more of one attribute respondents are ready to tolerate in exchange for higher levels of another. For example, the ratio of coefficients of the probability of detection and risk of unnecessary follow-up, $\left(\frac{\beta_{\text{Detect}}}{-\beta_{\text{Risk}}}\right)$, represents the number of unnecessary follow-ups women were willing-to-accept for an additional cancer detected per 100 women screened. Similarly, the marginal willingness-to-pay (MWTP) for an additional cancer detected was calculated as $\frac{\beta_{\text{Detect}}}{-\beta_{\text{Cost}}}$ and the MWTP for an additional unnecessary follow-up as $\frac{\beta_{\text{Risk}}}{-\beta_{\text{Cost}}}$. The MWTP values were then used to estimate the relative value of each attribute.

Predicting Uptake

To simulate how uptake differs across different groups and to understand how uptake might change as the levels of risk and detection in an NBSP change, the probability of choosing screening (P_i) over no screening was calculated as follows:

$$P_i = \frac{e^{V_i}}{e^{V_i} + e^{V_j}} = \frac{e^{(\beta_1 \text{Detect} + \beta_2 \text{Risk} + \beta_3 \text{Cost})}}{e^{(\beta_1 \text{Detect} + \beta_2 \text{Risk} + \beta_3 \text{Cost})} + e^{\beta_{\text{none}}}}. \quad (4)$$

In Equation 4, V_i and V_j represent the deterministic components of utility for alternatives of breast screening and no screening, respectively.

Results

The final study sample (see Appendix Table A2 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2017.07.010) comprised 1018 women who completed the DCE (PO [n = 507] or IAP [n = 511]). The response rate, measured in terms of survey loads as a proportion of emails delivered, was 9%. In this study, 8.9% (n = 90) of the sample failed the check for monotonic preferences. Out of the respondents who failed, 42 received the PO version and 48 received the IAP version; this difference was not statistically significant between survey versions ($P = 0.533$). There was also no statistically significant difference between the two formats with regard to self-reported task difficulty ($P = 0.640$).

Results of the HCL Model

All attribute coefficients were statistically significant and had signs consistent with a priori expectations about the expected direction of impact of the attribute on preferences (Table 2). The IAP interaction terms were not statistically significant, suggesting that the risk format presented did not lead to differences in the marginal utility of the detect or risk attributes. The insignificant scale term also suggested that the method of communicating risk had no effect on choice consistency in the sample. A likelihood ratio test showed that allowing for nonlinearity in the attributes did not increase fit ($P > 0.05$), and thus the HCL and SALC models assumed a linear and continuous representation of the attributes, which was compatible with the data. This specification also allows for an easier interpretation of the estimated results. The ASC was large, negative, and statistically significant, suggesting that women derived utility from screening over and above that derived from the attributes.

Table 2 – Results of the heteroskedastic conditional logit model.

Attribute label	Estimates	
	Coefficients	SE
Utility		
Detect	0.081*	0.00
Risk	−0.047*	0.00
Cost†	−0.103*	0.01
IAP detect	−0.010‡	0.01
IAP risk	0.005‡	0.00
ASC (on none)	−1.497*	0.07
Scale term		
IAP	0.094	0.06

Note. The number of respondents is 1,018 and the number of observations is 33,594. **ASC represents the baseline utility from participating in screening beyond what is explained by the attributes presented.**

ASC, **alternative specific constant**; IAP, **icon arrays and percentages**; SE, standard error.

* $P < 0.001$.

† Cost attribute scaled to £1 = £100, so coefficient represents the effect of a £100 change in the cost of the program.

‡ $P < 0.05$.

Heterogeneity in Preferences

SALC models were estimated to explore heterogeneity in the preferences of the sample. Although several variables were hypothesized to influence preferences and scale (education, religion, age, children, experience of mammography, and experience of breast cancer with either friends and/or family), the only significant preference class covariates were dummy variables for concern about breast cancer risk, employment status, and ethnicity. The only significant scale class covariates were self-reported task difficulty and failure of the dominance test. The definitions of the variables are presented in Appendix Table A1 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2017.07.010.

In Appendix Table A3 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2017.07.010, a summary of latent class models with different numbers of preference classes and their associated information criteria and log likelihood values is presented. Allowing for different preference classes improved fit over a pooled HCL, which can be seen as a "one class" model (BIC 19877.922 v 14595.29). SALC models were always preferred to standard preference class models on the basis of the information criteria. Adding scale covariates improved the model fit further.

The final model selected included three scale classes and six preference classes. Table 3 describes parameters of the selected SALC model with the preference and scale covariates that were statistically significant in at least one preference class or one scale class. Relative to the base class (scale class 1), the scale factor in scale class 2 was 0.14 (high error variance) and in scale class 3 was 0.01 (very high error variance). Three-quarter of the respondents fell into scale class 2 or 3, exhibiting relatively low choice consistency. Unsurprisingly, the women in scale class 2 were significantly less likely to have reported the task as being very easy and the women in scale class 3 were significantly more likely to have failed the check for monotonic preferences and reported the task as hard. The preference parameters in Table 3 are presented for scale class 1, and so these estimates must be multiplied by the respective scale factors of 0.14 and 0.01 for scale classes 2 and 3, respectively.

Apart from the probability of detecting a cancer, which was statistically insignificant in preference class 5, the coefficients for each attribute had the expected signs and were statistically significant across all preference classes. The only significant preference covariates were employment status, unconcerned about risk of breast cancer, and ethnicity.

Most women (80%) fell into preference class 1, 2, or 4. Preference class 1 was the largest and accounted for almost a third of the sample (32.3%). Women in this class treated the attribute "probability of detecting a cancer" as the most important. The magnitude of the ASC in this class suggests that the utility acquired from participation outweighs the utility from any other attributes, meaning that women in this class would always participate in screening if detection, risk, and cost were within the range of the levels specified in this DCE. The women in this group were significantly less likely to be unconcerned about their risk of breast cancer and statistically significantly more likely to be employed. Women in preference class 2 (accounting for 29% of the sample) were also significantly less likely to be unconcerned about their risk of breast cancer. Another large class (accounting for 18.7% of the sample) was preference class 4, which contained women exhibiting similar preferences to classes 1 and 2 with large negative ASCs for the option of "no screening."

Table 3 – Preference classes in scale class 1 of a three-scale class 6 preference class model.

	Preference class 1	Preference class 2	Preference class 3	Preference class 4	Preference class 5	Preference class 6
Preference class proportions	32.35%	29.15%	7.48%	18.70%	7.73%	4.60%
ASC (on none)	–171.997[*]	–153.514[*]	–9.971[*]	–18.993[*]	1.189	–6.159[*]
	(65.932)	(53.775)	(2.253)	(3.869)	(1.366)	(1.437)
Detect	2.785[*]	0.846[*]	0.279[†]	0.152[*]	–0.104	0.956[*]
	(0.518)	(0.184)	(0.110)	(0.055)	(0.268)	(0.288)
Risk	–0.331[*]	–0.999[*]	–0.504[*]	–0.155[*]	–0.723[*]	–1.704[*]
	(0.083)	(0.234)	(0.140)	(0.041)	(0.234)	(0.421)
Cost[‡]	–1.003[*]	–0.303[*]	–5.354[*]	–2.268[*]	–2.417[*]	–0.638[*]
	(0.229)	(0.086)	(1.208)	0.478)	(0.770)	(0.173)
Preference covariates						
Unconcerned dummy[†]	–0.337[*]	–0.202[§]	0.231	0.117	0.120	0.071
	(0.128)	(0.115)	(0.150)	(0.126)	(0.148)	(0.199)
Employed dummy[§]	0.190[†]	0.029	–0.183	0.127	0.095	–0.257[§]
	(0.080)	(0.079)	(0.120)	(0.096)	(0.123)	(0.155)
White dummy[†]	0.227	0.289	0.048	–0.283	–0.508[*]	0.228
	(0.189)	(0.192)	(0.241)	(0.174)	(0.188)	(0.353)
IAP dummy	–0.064	0.111	0.054	0.025	–0.082	–0.045
	(0.074)	(0.075)	(0.115)	(0.091)	(0.116)	(0.152)
	Scale class 1	**Scale class 2**	**Scale class 3**			
Scale class proportions	23.84%	64.87%	11.29%			
Scale factor	1	0.139	0.014			
Scale covariates						
Failed monotonicity check dummy[†]		–0.088	6.630[†]			
		(4.133)	(3.101)			
Task difficulty[*]						
Very easy (1)	.	–0.815[*]	–3.358			
		(0.292)	(2.225)			
Easy (2)	.	–0.129	–3.072			
		(0.312)	(2.349)			
Neither easy/hard (3)	.	0.139	2.197[†]			
		(0.340)	(0.954)			
Hard (4)	.	0.759	1.606			
		(0.490)	(1.175)			
Very hard (5)	.	0.046	2.627[§]			
		(0.958)	(1.496)			
IAP dummy		0.042	0.220			
		(0.106)	(0.305)			

Note. SEs are given in parentheses.
ASC, alternative specific constant; IAP, icon arrays and percentages; SE, standard error.
[*] $P < 0.01$.
[†] $P < 0.05$.
[‡] Cost attribute scaled to £1 = £100, so coefficient represents the effect of a £100 change in the cost of the program.
[§] $P < 0.1$.

In preference class 5, in which women were significantly more likely to be from ethnic minorities (nonwhite), the ASC on "no screening" was positive although statistically insignificant. The only significant attributes were the cost of screening and the risk of unnecessary follow-up, suggesting that women in this class took account of the downsides of an NBSP. The coefficient on the attribute "probability of detecting a cancer" was negative and not statistically significant, suggesting that it was not an important factor of screening choice for women in preference class 5. To test whether women in this preference class ignored the attribute (exhibited attribute "nonattendance"), the model was restricted, constraining the coefficient on the Detect parameter to 0. Restricting the model to allow for non attendance to the detection attribute in preference class 5 improved model fit (BIC reduced to 14023.53), suggesting that women in this class completely ignored this attribute.

Willingness-to-Pay and Willingness-to-Accept Risk

MRS across the preference classes and on average from the HCL model are presented in Table 4. The average MWTP for an additional cancer detected was estimated to be £79.17 (confidence interval [CI] £66.98–£91.35) and willingness-to-accept a case of unnecessary follow-up was estimated to be £46.01 (CI £51.19–£40.84). The results of the HCL model also showed that, on average, women were willing-to-accept nearly 1.72 (CI 1.47–1.97) additional unnecessary follow-ups for an additional cancer detected per 100 women screened.

Table 4 – Marginal rates of substitution (including MWTP).

	Heteroskedastic conditional logit	Preference class 1	Preference class 2	Preference class 3	Preference class 4	Preference class 5*	Preference class 6
MWTP detect	£79.17 (£66.98 to £91.35)	£277.81	£279.48	£5.20	£6.71	–£4.28	£149.83
MWTP risk	–£46.01 (–£51.19 to –£40.84)	–£33.04	–£330.15	–£9.42	–£6.83	–£29.90	–£267.01
Willingness-to-accept risk	1.72 (1.47 to 1.97)	8.41	0.85	0.55	0.98	–0.14	0.56

MWTP, marginal willingness-to-pay.
* Insignificant coefficient on detection.

Predicted Uptake

Assuming a 10% probability of detecting a cancer, a 10% risk of unnecessary follow-up, and a lifetime screening cost of £100, on average, the probability of a woman participating in screening was estimated to be 85% (CI 84.87%–85.13%). In the best-case scenario with no risk of unnecessary follow-up and 14% probability of detecting a cancer, uptake was predicted to be 93% (CI 92.57%–92.72%).

As shown in the large ASCs in preference classes 1, 2, and 4, women in this class strongly dislike the option of "no screening." As a consequence, in the worst-case screening formulation (a 3% probability of detecting a cancer and a 20% risk of unnecessary follow-up) two-third of women would still attend (66.8%; CI 66.47%–66.80%). Only when the risk of unnecessary follow-up was 47% (given a 10% probability of cancer detection) would screening uptake drop to 50%. In a catastrophic scenario of no cancers being detected and all women in the program receiving some unnecessary follow-up, 4% of women would still participate in an NBSP. Figure 2 shows the effect of an increasing risk of unnecessary follow-up on the probability of a woman participating in an NBSP for different preference classes and the pooled HCL model. Appendix Figures B1 and B2 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2017.07.010 show the effect of increasing detection rates and cost on the probability of a woman participating in an NBSP.

Discussion

The results of the DCE showed many women trading the attributes presented, with results showing many significant coefficients with most signs in line with a priori expectations. In most analyses, the ASC was large and negative, suggesting that women derive utility from screening over and above that derived from the attributes included in the experiment. One possible reason for this could be the current policies that encourage participation, which could suggest that screening is inherently beneficial. No screening could imply that the detection of cancer would be dependent on self-examination, which women may feel is inferior.

The SALC analysis revealed that women's preferences for breast screening were highly heterogeneous, with six distinct preference classes. Allowing for different preference classes improved fit over a pooled HCL. This is in line with other research that has found that accounting for preference heterogeneity can improve model fit [27,53]. The study also highlighted the importance of accounting for scale, in addition to preference, heterogeneity in respondents' choices. People who failed the check for monotonic preferences and reported finding the task hard were more likely to be in a more random scale class. This result also suggests that task difficulty and failure of the monotonicity check may be correlated. This finding was in line with existing evidence to suggest that scale heterogeneity can affect the interpretation of coefficient estimates and is particularly prevalent in health care DCEs [45,54].

The largest class, preference class 1, comprised women who tended to be more concerned about their risk of breast cancer and were significantly more likely to be employed. There was a high utility associated with participating in a program and detection was the most valued attribute in this class. In preference class 5, however, women appeared to be uninterested in the NBSP with the estimate of the ASC and the detection attribute both being positive and insignificant. Women in this class tended to be from ethnic minorities. These findings are consistent with a large literature acknowledging low uptake rates from minority ethnic populations [55–57]. There is some evidence that better communication of the benefits through general practitioners' endorsement letters and multilingual leaflets [58] may increase uptake from ethnic minorities. Alternatively, women in this group could be compensated (or could receive a subsidy for their incurred out-of-pocket costs) for screening. A study found that paying for transport improved attendance by 16% in ethnic minority groups [58].

The estimated uptake rate of about 85% contrasted to the estimate that 75% of women invited attend an NBSP [59]. If generalizability is defined as the extent to which results can be transferred by a decision maker to the relevant population [10], this study could be seen as producing generalizable findings matching other literature and uptake estimates.

The results of the HCL model also indicated that the interaction terms between IAP and the DCE attributes were insignificant and that IAP had an insignificant effect on the scale term. These results were further confirmed in the SALC analysis that found that the risk format was not a significant predictor of either preference class or scale class membership. We therefore conclude that, in this example, risk communication format had no effect on preferences or choice consistency. Another DCE [60] that investigated the format of risk attributes concluded that graphics or icons have not aided respondents' choice making. In health risk communication more generally, research has shown that for some groups the format of risk is less important because all communication methods are difficult to understand [61]. In this DCE, the two formats could have been equally challenging for respondents and therefore resulted in no difference.

Recent changes to breast screening policy have included an extension of the screening ages by 6 years, inviting women between 47 and 73 years old in some areas of England (National Institute for Health Research Trial ISRCTN33292440), changing the benefit-risk ratio. Policymakers should be aware of women's preferences for the increased risk of unnecessary follow-up associated with screening younger and older women. For women between 50 and 70 years old, how preferences change through the program (because the probability of detecting a cancer reduces and the risk of unnecessary follow-up rises) should also be considered by policymakers [62].

A qualitative study investigating Australian women's views on overdiagnosis found that "the lower and intermediate estimates (1%–10% and 30%) had limited impact on attitudes and intentions, with many women remaining committed to screening" [63p1]. Although this is in line with the results of the HCL model, the SALC analysis indicated substantial heterogeneity around these views. At a 30% risk of unnecessary follow-up, the

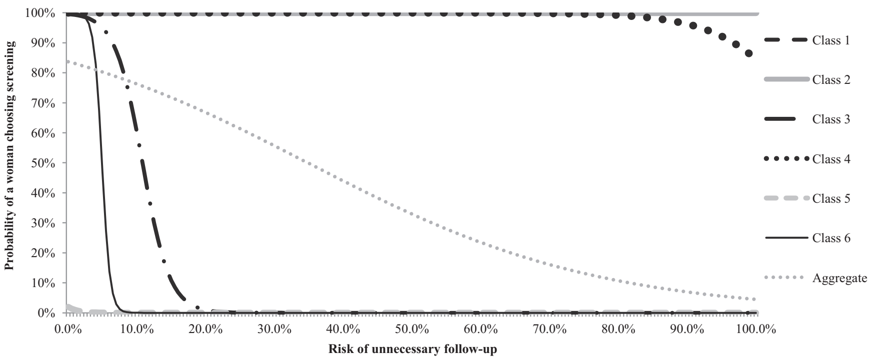

Fig. 2 – The effect of increasing risk of unnecessary follow-up on uptake for breast screening by preference class assuming a detection rate of 3% and a £100 lifetime cost of screening.

results of our DCE would predict uptake to be very low (almost 0) in preference classes 3, 5, and 6 (accounting for around a fifth of the women sampled).

This DCE possesses characteristics of generalizability, with results in line with a priori expectations and comparable with other results in the literature. Nonlinearity in preferences was investigated in the preliminary stages of model selection; nevertheless, investigations of nonlinearity could not be conducted in the SALC analysis because these failed to converge because of the low sample size in some classes. There is a possibility that some classes may contain individuals with nonlinear preferences and this should be considered when interpreting the results presented. Furthermore, although there were no statistically significant two-way interactions between the attributes in this study, estimating demand was not an analytical aim and readers should interpret the predicted uptake calculations with caution. Models that focus on preferences or preference heterogeneity can risk overfitting data, limiting the usefulness of the results for estimating demand [64]. This study presents predicted uptake to illustrate the degree of preference heterogeneity rather than providing an accurate analysis of demand.

Internet panels have been criticized because of potential selection bias and the conditioning of respondents whose preference and behavior may change because of participation in the panel [65]. Therefore, the views found in this study may not necessarily be representative of the general public's. The survey source may limit the interpretation of the quantified benefit-risk trade-offs and even the existence of the distinct latent preference classes identified. Nevertheless, internet panels allowed a large sample size, paramount for the investigation of heterogeneity.

The conclusion that IAP offered no advantage to respondents in terms of reduced cognitive burden and had no effect on their preferences may be specific to the context of this study (breast screening) or the magnitude of the levels used (percentages were whole numbers). In other scenarios, the results may differ, and further research is required to investigate the generalizability of this finding. Furthermore, we described the MRS with the risk and detect attributes as a willingness-to-accept risk. In the willingness-to-pay literature there is evidence that people do not feel the same when asked to accept an identical amount in compensation [66]. Further research is required to understand whether the finding from this study is robust to framing effects.

Conclusions

The results of this DCE suggested that most women were willing to trade-off the probability of detecting a cancer to avoid unnecessary follow-up. Nevertheless, there exists significant preference heterogeneity in women's preferences for an NBSP. For some women, particularly those from ethnic minorities, no amount of additional cancers detected would sufficiently compensate for the risk of unnecessary follow-up. The study also found that women's preferences were robust to the communication format used to present the risk attributes. This study contributes to the debate about the relative harms and merits of NBSPs by highlighting the drivers of screening attendance and quantifying the degree of heterogeneity in preferences.

Acknowledgments

We are grateful for feedback received at the Society for Medical Decision Making's 36th and 37th annual meetings, International Society for Pharmacoeconomics and Outcomes Research, and the International Academy of Health Preference Researcher's inaugural meeting. We are also grateful to experts Professor Gareth Evans, Professor Tony Howell, Dr. Michelle Harvie, and Ms. Paula Stavrinos of the Nightingale Centre at Wythenshawe Hospital and Professor Stephen Campbell of the Centre for Primary Care at The University of Manchester for providing their thoughts and comments on our idea and drafts of the survey.

Source of financial support: C.M. Vass was in receipt of a National Institute for Health Research School for Primary Care Research PhD Studentship between October 2011 and 2014. Preparation and revision of this article was made possible with a grant from Mind the Risk, a project funded by Riksbanken Jubileumsfond. The views expressed in this article are those of the authors and not of the funding bodies.

Supplemental Materials

Supplemental material accompanying this article can be found in the online version as a hyperlink at http://dx.doi.org/10.1016/j.jval.2017.07.010 or, if a hard copy of article, at www.valueinhealthjournal.com/issues (select volume, issue, and article).

REFERENCES

[1] McPherson K, Steel CM, Dixon JM. ABC of breast diseases. Breast cancer—epidemiology, risk factors and genetics. BMJ 2000;321:624–8.

[2] von Karsa L, Anttila A, Ronco G, et al. Cancer screening in the European Union. Report on the implementation of the Council Recommendation on cancer screening 2008.

[3] Torjesen I. How much is too much breast screening? BMJ 2015; 350:h139.
[4] Bleyer A, Welch HG. Effect of three decades of screening mammography on breast-cancer incidence. N Engl J Med 2012;367:1998–2005.
[5] Moss SM, Wale C, Smith R, et al. Effect of mammographic screening from age 40 years on breast cancer mortality in the UK Age trial at 17 years' follow-up: a randomised controlled trial. Lancet Oncol 2015;16:1123–32.
[6] Liston JC, Dall BJG. Can the NHS Breast Screening Programme afford not to double read screening mammograms? Clin Radiol 2003;58: 474–7.
[7] Welch HG, Black WC. Overdiagnosis in cancer. J Natl Cancer Inst 2010;102:605–13.
[8] Hofvind S, Thoresen S, Tretli S. The cumulative risk of a false-positive recall in the Norwegian Breast Cancer Screening Program. Cancer 2004;101:1501–7.
[9] Baum M. Harms from breast cancer screening outweigh benefits if death caused by treatment is included. BMJ 2013;346:f385.
[10] Kirwan CC. Breast cancer screening: What does the future hold? BMJ 2013;346:f87.
[11] Pace LE, Keating NL. A systematic assessment of benefits and risks to guide breast cancer screening decisions. JAMA 2014;311: 1327–35.
[12] Gøtzsche P, Nielsen M. Screening for breast cancer with mammography (review). Cochrane Database Syst Rev 2011: CD001877.
[13] Mansfield C, Tangka FK, Ekwueme DU, et al. Stated preference for cancer screening: a systematic review of the literature, 1990–2013. Prev Chronic Dis 2016;13:E27.
[14] Gerard K, Shanahan M, Louviere J. Using stated preference discrete choice modelling to inform health care decision-making: a pilot study of breast screening participation. Appl Econ 2003;35:1073–85.
[15] Clark M, Determann D, Petrou S, et al. Discrete choice experiments in health economics: a review of the literature. Pharmacoeconomics 2014;32:883–902.
[16] de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. Health Econ 2012;21:145–72.
[17] McFadden D. Conditional logit analysis of qualitative choice behaviour. In: Zarembka P,ed. Frontiers in Econometrics. New York, NY: Academic Press, 1974. p. 105–42.
[18] Lancaster KJ. A new approach to consumer theory. J Polit Econ 1966;74:132–57.
[19] Thurstone L. A law of comparative judgment. Psychol Rev 1927;34:273–86.
[20] Lancsar E, Donaldson C. Discrete choice experiments in health economics: distinguishing between the method and its application. Eur J Health Econ 2005;6:314–6.
[21] Harrison M, Rigby D, Vass C, et al. Risk as an attribute in discrete choice experiments: a critical review. Patient 2014;7:151–70.
[22] Hauber A, Fairchild AO, Johnson F. Quantifying benefit-risk preferences for medical interventions: an overview of a growing empirical literature. Appl Health Econ Health Policy 2013;11:319–29.
[23] Ho MP, Gonzalez JM, Lerner HP, et al. Incorporating patient-preference evidence into regulatory decision making. Surg Endosc 2015;29: 2984–93.
[24] Lipkus I. Numeric, verbal, and visual formats of conveying health risks: suggested best practices and future recommendations. Med Decis Making 2007;27:696–713.
[25] Price M, Cameron R, Butow P. Communicating risk information: the influence of graphical display format on quantitative information perception—accuracy, comprehension and preferences. Patient Educ Couns 2007;69:121–8.
[26] Galesic M, Garcia-Retamero R, Gigerenzer G. Using icon arrays to communicate medical risks: overcoming low numeracy. Health Psychol 2009;28:210–6.
[27] Hole AR. Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment. J Health Econ 2008;27:1078–94.
[28] Mayor S. Critics attack new NHS breast screening leaflet for failing to address harms. BMJ 2010;341:c7267.
[29] Gøtzsche PC, Jørgensen KJ. The breast screening programme and misinforming the public. J R Soc Med 2011;104:361–9.
[30] Bridges JF, Hauber AB, Marshall D, et al. Conjoint analysis applications in health—a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. Value Health 2011;14: 403: 13.
[31] Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. Pharmacoeconomics 2008;26:661–77.
[32] BreastHealth UK. Private breast screening and diagnostic service prices. 2014. Available from: www.breasthealthuk.com.
[33] Welch HG, Frankel B. Likelihood that a woman with screen-detected breast cancer has had her "life saved" by that screening. Arch Intern Med 2011;171:2043–6.
[34] Hall AE, Chowdhury S, Hallowell N, et al. Implementing risk-stratified screening for common cancers: a review of potential ethical, legal and social issues. J Public Health (Oxf) 2014;36:285–91.
[35] Independent UK. Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. Lancet 2012;380:1778–86.
[36] Johnson F, Lancsar E, Marshall D, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. Value Health 2013;16:3–13.
[37] ChoiceMetrics. Ngene User Manual 1.1.1. ChoiceMetrics, 2012.
[38] Taylor-Phillips S, O'Sullivan E, Kearins O, et al. The effects of a UK review of breast cancer screening on uptake: an observational before/after study. J Med Screen 2013;20:86–90.
[39] Center for Bioethics and Social Sciences in Medicine. University of Michigan. Available from: www.iconarray.com.
[40] Vass C, Rigby D, Campbell S, et al. Investigating the framing of risk attributes in a discrete choice experiment: an application of eye-tracking and think aloud. Med Decis Making 2014;35:E99.
[41] Sawtooth Software. Sawtooth Software S.S.I. Web 8.3.8. Orem, UT: Sawtooth Software, 2012.
[42] Evans G, Blair V, Greenhalgh R, et al. The impact of genetic counselling on risk perception in women with a family history of breast cancer. Br J Cancer 1994;70:934–8.
[43] Evans G, Burnell LD, Hopwood P, Howell A. Perception of risk in women with a family history of breast cancer. Br J Cancer 1993;67: 612–4.
[44] Schwartz L, Woloshin S, Black W, Welch HG. The role of numeracy in understanding the benefit of screening mammography. Ann Intern Med 1997;127:1023–8.
[45] Fiebig D, Keane M, Louviere J, Wasi N. The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. Mark Sci 2010;29:393–421.
[46] Swait J, Louviere J. The role of the scale parameter in the estimation and comparison of multinomial logit models. J Mark Res 1993;30:305–14.
[47] Louviere J, Hensher D, Swait J. Stated Choice Methods: Analysis and Application. Cambridge, UK: Cambridge University Press, 2000.
[48] Hole AR. Small-sample properties of tests for heteroscedasticity in the conditional logit model. University of York, York, UK: Working Paper 06/04, Health Economics Data Group, 2006.
[49] Magidson J, Vermunt J. Removing the scale factor confound in multinomial logit choice models to obtain better estimates of preference. Sequim, WA: Presented at: Proceedings of the Sawtooth Software Conference, 1–18.
[50] Dziak J, Coffman D, Lanza S, Li R. Sensitivity and specificity of information criteria. Methodol Cent Tech Rep Ser Pennsylvania State University 2012.
[51] StataCorp. Stata Statistical Software: Release 12. College Station, TX: StataCorp, 2011.
[52] Statistical Innovations. Latent GOLD. Belmont, MA: Statistical Innovations, 2013.
[53] Shen J. Latent class model or mixed logit model? A comparison by transport mode choice data. Appl Econ 2009;41:2915–24.
[54] Flynn T, Louviere J, Peters T, Coast J. Using discrete choice experiments to understand preferences for quality of life: variance-scale heterogeneity matters. Soc Sci Med 2010;70:1957–65.
[55] Hoare T. Breast screening and ethnic minorities. Br J Cancer Suppl 1996;29:S38–41.
[56] Renshaw C, Jack R, Dixon S. Estimating attendance for breast cancer screening in ethnic groups in London. BMC Public Health 2010;10:1–8.
[57] Bansal N, Bhopal RS, Steiner MFC, Brewster DH. Major ethnic group differences in breast cancer screening uptake in Scotland are not extinguished by adjustment for indices of geographical residence, area deprivation, long-term illness and education. Br J Cancer 2012;106:1361–6.
[58] Bell T, Branston L. Interventions to improve uptake of breast screening in inner city Cardiff general practices with ethnic minority lists. Ethn Health 1999;4:277–84.
[59] Screening and Immunisations Team Health and Social Care Information Centre. Breast Screening Programme, England Statistics for 2014-15. 2016. http://content.digital.nhs.uk/catalogue/PUB20018/bres-scre-prog-eng-2014-15-rep.pdf. accessed 21 Aug 17.
[60] Veldwijk J, Essers BAB, Lambooij MS, et al. Survival or mortality: Does risk attribute framing influence decision-making behavior in a discrete choice experiment? Value Health 2016;19:202–9.
[61] Peters E, Hart PS, Tusler M, Fraenkel L. Numbers matter to informed patient choices. Med Decis Making 2014;34:430–42.

[62] Barratt A. Overdiagnosis in mammography screening: a 45 year journey from shadowy idea to acknowledged reality. BMJ 2015;350: h867.
[63] Hersch J, Jansen J, Barratt A, et al. Women's views on overdiagnosis in breast cancer screening: a qualitative study. BMJ 2013;346:1–13.
[64] Hauber A, Gonzalez JM, Groothuis-Oudshoorn C, et al. Statistical methods for the analysis of discrete-choice experiments: a report of the ISPOR Conjoint Analysis Good Research Practices Task Force. Value Health 2016;19:300–15.
[65] Dennis J. Are internet panels creating professional respondents? Mark Res 2001;13:34–8.
[66] Grutters JPC, Kessels AG, Dirksen CD, et al. Willingness to accept versus willingness to pay in a discrete choice experiment. Value Health 2008;11:1110–9.