HAL
open science

What are GPs' preferences for financial and non-financial incentives in cancer screening? Evidence for breast, cervical, and colorectal cancers

Jonathan Sicsic, Nicolas Krucien, Carine Franc

To cite this version:

Jonathan Sicsic, Nicolas Krucien, Carine Franc. What are GPs' preferences for financial and non-financial incentives in cancer screening? Evidence for breast, cervical, and colorectal cancers. Social Science & Medicine, 2016, 167, pp.116-127. 10.1016/j.socscimed.2016.08.050 . halshs-02951764

HAL Id: halshs-02951764

https://shs.hal.science/halshs-02951764v1

Submitted on 28 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

What are GPs' preferences for financial and non-financial incentives in cancer screening? Evidence for breast, cervical, and colorectal cancers[1]

Jonathan Sicsic[1,*], PhD, Nicolas Krucien[2], PhD, Carine Franc[1], PhD

Authors' affiliations:

[1]CESP, University Paris-Sud, UVSQ, INSERM, University Paris-Saclay, Hôpital Paul Brousse, 16, Avenue Paul Vaillant-Couturier, 94801 Villejuif. France

[2]Health Economics Research Unit, University of Aberdeen, Institute of Applied Health Sciences, Foresterhill, Aberdeen, AB25 2QN. United Kingdom

*__Corresponding author__: Jonathan Sicsic. E-mail: jonathan.sicsic@inserm.fr. Phone: +330145595258

Highlights

- General practitioners (GPs) play an important role in cancer screening activities.
- A DCE was used to assess the importance of financial and non-financial incentives.
- Both types of incentives are required to boost GPs' cancer screening provision.
- The relative impact of incentives differs across screening contexts.
- Additional payment seems relatively more incentivizing for colorectal cancer.

[1] This is the last version of the article before proofs. The final version is available at: http://www.sciencedirect.com/science/article/pii/S0277953616304865. Please cite this article as: Sicsic, J., Krucien, N., Franc, C., 2016. What are GPs' preferences for financial and non-financial incentives in cancer screening? Evidence for breast, cervical, and colorectal cancers. Soc. Sci. Med. 167, 116–127. doi:10.1016/j.socscimed.2016.08.050.

ABSTRACT

General practitioners (GPs) play a key role in the delivery of preventive and screening services for breast, cervical, and colorectal cancers. In practice, GPs' involvement varies considerably across types of cancer and among GPs, raising important questions about the determinants of GPs' implication in screening activities: what is the relative impact of financial and non-financial incentives? Are GPs' preferences for financial and non-financial incentives cancer-specific? Is there preference heterogeneity and how much does it differ according to the screening context? This study investigates the determinants of GPs' involvement in cancer screening activities using the discrete choice experiment (DCE) methodology. A representative sample of 402 GPs was recruited in France between March and April, 2014. Marginal rates of substitution were used to compare GPs' preferences for being involved in screening activities across three types of cancers: breast, cervical, and colorectal. Variability of preferences was investigated using Hierarchical Bayes mixed logit models. The results indicate that GPs are sensitive to both financial and non-financial incentives, such as a compensated training and systematic transmission of information about screened patients, aimed to facilitate communication between doctors and patients. There is also evidence that the level and variability of preferences differ across screening contexts, although the variations are not statistically significant on average. GPs appear to be relatively more sensitive to financial incentives for being involved in colorectal cancer screening, whereas they have higher and more heterogeneous preferences for non-financial incentives in breast and cervical cancers. Our study provides new findings for policymakers interested in prioritizing levers to increase the supply of cancer screening services in general practice.

Keywords: France; Cancer screening; Discrete choice experiment; Economic incentives; Hierarchical bayes; Mixed logit; Preference heterogeneity.

1. Introduction

Cancer is a major health issue worldwide (WHO, 2008). Preventive care, such as screening, is important for reducing the mortality and burden of cancer (Gellad and Provenzale, 2010; Kerlikowske et al., 1995; Quinn et al., 1999). In many healthcare systems, general practitioners (GPs) play an important role in delivering preventive care and appropriate recommendation for the screening of breast, cervical, and colorectal cancer (Carrieri and Bilger, 2013; Jensen et al., 2012; Weller, 1997). In a study analysing decision-making processes for several cancers among a national sample of US adults, providers such as GPs were cited as the most highly rated information source for initiating screening discussion and recommending screening (Hoffman et al., 2010). However, GPs' involvement in cancer screening varies considerably across types of cancer and among GPs (Federici et al., 2005; Ganry and Boche, 2005). In a representative survey conducted by the French National Institute of Cancer (INCa), GPs reported that they "routinely" check breast, cervical, and colorectal cancer screening in 56%, 45% and 34% of consultations (Bungener et al., 2010). This somewhat low level of screening performance and the between-cancer variability raise important questions about the determinants of GPs' participation in cancer screening activities. For instance, what makes GPs willing to be involved in cancer screening activities? Do GPs differ in their attitudes towards different types of cancer?

Previously, it has been suggested that financial incentives such as pay-for-performance (P4P) would greatly influence GPs' decisions to deliver preventive services for cancer (Armour et al., 2004; Town et al., 2005). For instance, in France, GPs earn financial bonuses in addition to their fee-for-service (FFS) remuneration if they reach (or draw closer to) a target rate of 80% of women aged between 50 and 74 years having been screened for breast cancer in the past two years. In France, GPs are mostly paid on a FFS base whereas in the Scandinavian countries and in the UK, doctors are primarily paid through a per capita mechanism. The latter has been shown to be more favorable to the supply of prevention while FFS provides no incentive to supply preventive care if not compensated (Franc and Lesur, 2004; Hennig-Schmidt et al., 2011). Economic theory predicts that monetary incentives may induce optimal provision of healthcare services but the empirical analyses show mixed results (Eijkenaar et al., 2013; Mannion and Davies, 2008) and nonsignificant results for preventive care (Kiran et al., 2014; Li et al., 2014; Sicsic and Franc, 2016). Initially, the low performance of P4P has been explained by inadequate level of financial incentives (Town et al., 2005) and their temporal nature, as prevention may have long or mid-term returns of investment. Indeed, P4P remuneration is based on an annual measure of GPs' activity in terms of cancer screening provision, whereas the benefits of

screening are expected in much longer term. Other studies have suggested potential unintended consequences of financial incentives, particularly the crowding out of doctors' intrinsic (non-financial) motivations by extrinsic (financial) rewards (Janus, 2010; Sicsic et al., 2012), which may contribute to explain the low impact of P4P on GPs' prevention activities.

Other non-financial factors are likely to play a significant role in GPs' decisions regarding cancer screening. These non-financial factors include physicians training, receiving feedback, and assistance from other non-health professionals (McIlfatrick et al., 2013; Sabatino et al., 2008). Previous studies investigated doctors' preferences for key job attributes concerning location choices in general practice by focusing on the role of both pecuniary and non-pecuniary incentives (Günther et al., 2010; Holte et al., 2015; Scott, 2001). Holte et al (2015) found that additional income had smaller impact on GPs' choices than improvements in non-monetary attributes such as opportunity for professional development. To the best of our knowledge, there is no similar evidence for the role of pecuniary and non-pecuniary incentives in GPs' preferences for cancer-screening activities.

This study offers to bridge this gap by investigating how GPs trade financial and non-financial incentives when making decisions to be involved in screening activities for three types of cancers, namely breast, cervical, and colorectal cancer. For these cancers, the effectiveness of screening is recognized and guidelines are available (Saslow et al., 2012), despite recent debates about the benefits and harms balance for breast cancer screening (Gøtzsche and Nielsen, 2011). A better understanding of how financial and non-financial incentives influence GPs' decisions and interact with each other will help to improve effectiveness, quality, and sustainability of screening programmes. Given the considerable variability in GP's involvement in cancer screening activities, it is important to understand how GPs' preferences differ across both cancers and GPs themselves. We address this issue by investigating heterogeneity in GPs' preferences for cancer screening programmes by fitting Hierarchical Bayes mixed logit models and by comparing preferences across breast, cervical, and colorectal cancers.

In the next section, we summarize the literature on interventions to improve delivery of preventive services and review the French context for cancer screening. In section 3, we present the discrete choice experiment (DCE) survey and the statistical methodology used to respond to the different research questions. The results are presented in section 4 and discussed in section 5.

2. Literature

2.1/Interventions to increase delivery of preventive services

Much of the research to date has focused on evaluating the efficacy of interventions aimed at promoting behaviour change among healthcare providers (Ellis et al., 2005; Grimshaw et al., 2001; Sabatino et al., 2012, 2008; Zapka and Lemon, 2004). In their literature review, Grimshaw et al (2001) found that active interventions, such as reminders and educational outreach, were effective in changing healthcare provider behaviour, whereas less active interventions (e.g. attending conferences, reading medical journals) were not effective. Another study found that no single intervention was effective across the cancer continuum (Ellis et al., 2005). Interventions that were effective in several topic areas included the use of office systems (reminders and prompts), health care provider advice, removal of financial barriers, and multi-component interventions. In their literature review, Sabatino et al (2008) identified ten studies that reported the use of provider assessment and feedback to increase recommendation for breast, cervical, and colorectal cancer screening. They concluded that assessment and feedback interventions produced positive effects in both trainee and non-trainee physician groups but financial incentives alone were not effective. The result was confirmed four years later in an updated literature review (Sabatino et al., 2012).

The conclusions obtained in these various studies, although different, are not necessarily contradictory. It is possible that some screening incentives would act as complements and then their valuation would differ depending on whether they are combined or not. Besides, the context in which the various screenings are enrolled may influence how the incentive is perceived by GPs. It is thus essential to analyze more precisely the French screening context and the role played by the GP in each context.

2.2/The French screening context for cancer

In France, a national program for breast cancer screening has been implemented since 2004: women aged 50-74 years are mailed an invitation to perform a free mammogram (free at the point of use) in a radiological centre. They can choose a doctor (e.g. a GP or a gynaecologist) who will be informed of the results of the mammogram. Thus, referral GPs might not always be informed of the realization of a mammogram by their eligible patients, and this lack of information could constitute a barrier and hinder their involvement in breast cancer screening

(Liberalotto, 2012). Accordingly, systematic communication of the screening results to the referral GP could be one interesting method to promote.

To increase the take-up of colorectal cancer screening, a national program was implemented in 2009: men and women between 50 and 74 years are invited by mail to perform a free faecal occult blood test (FOBT) in a biological centre, and the referral GP should always be informed of the results. The GP takes a leading role in facilitating patient adherence to the national programmes: he/she is supposed to propose the test and explain the modalities of implementation and the consequences in case of positivity. Yet, GPs report being the initiators of a discussion about colorectal cancer screening in less than half of the cases (Bungener et al., 2010). One possible explanation could be related to the time required for its proposal in consultation and explanation of how it works. We assume that GPs could be sensitive to qualified staff assistance and/or additional compensation to offset the effort.

Cervical cancer screening (based on smear tests) has not yet been included in a national program (only experimentations are ongoing). Sociological studies have investigated GPs' opinions and attitudes concerning the effectiveness of experimental programs conducted locally (Liberalotto, 2012). Two main mechanisms implemented by local associations promoting cervical cancers screening were often cited by the GPs: practical training for the implementation of smear tests intended to the GP and the provision of information leaflets to the patient. Even if cervical cancer screening is mostly performed during gynaecologist consultations, GPs' role is recognized by the profession: 45% of GPs declare to routinely check cervical cancer screening, half of which declare that "performing this act is part of [their] job" (Bungener et al., 2010). In this context, a P4P programme was extended to cervical cancer screening in 2012 through the *Rémunénation sur Objectif de Santé Publique* (ROSP).

3. Discrete choice experiment

A stated preferences discrete choice experiment (DCE) was conducted to measure GPs' preferences for cancer screening incentives in three different contexts, namely breast, cervical, and colorectal cancer. DCEs are widely used in health to investigate patients, public, and health professionals' preferences for treatments or medical procedures (Clark et al., 2014; de Bekker-Grob et al., 2012). Discrete choice experiments are underpinned by Lancaster's consumer theory (Lancaster, 1966) which assumes that the utility of a product or service is derived from its characteristics (attributes). DCEs ask participants to make choices between several hypothetical

scenarios offering different combinations of attributes, in order to infer their preferences for each attribute or combination of attributes, independently. The first step consist in selecting attributes and levels, the second step is choosing an appropriate design for the choice tasks, the third step is sampling respondents and collecting data, and the last step is analysing data using econometric models.

3.1/ Selection of attributes and levels

The selection of attributes and levels was based on a literature review (see section 2) and interviews with ten GPs as well as members of two local structures managing the screening programmes at the departmental level. A thematic analysis of interviews was conducted to determine the most important aspects of screening programmes delivery (e.g. communication of information, skills, rewards) to include in the DCE study. For each cancer screening context, five attributes were selected: the first four attributes were non-financial (or transfers in kind) with two levels and the last attribute was a financial incentive with four levels of payment (see Table 1 for a summary of attributes, and Supplementary file A for a detailed description of attributes).

The Leaflet (*LEAF)* attribute consisted of cancer screening leaflets intended for the patient designed to facilitate doctor-patient communication and transmission of information, in order to possibly improve the screening acceptability and patient adherence. The Leaflet incentive is in line with the results of McIlfatrick et al (2013), i.e. the notion of empowering individuals to take responsibility for their health issues.

The Training (*TRAIN)* attribute was a compensated training for the GP, aimed to improve his/her feelings of autonomy and competence encountered in the doctor-patient relationship and ultimately increase his/her motivation and involvement in cancer screening. This training included the analysis of the difficulties in the implementation of screening, strategies of conviction and practical application, and coordination with other specialists. The training slightly differed according to the screening context, focusing on the benefits/risks of mammography in breast cancer screening, and on the realisation of pap smears in cervical cancer screening.

The Listing (*LIST)* attribute was a bi-annual listing of screened patients destined to the GPs, updated every 6 months, designed to promote GPs' access to information and save time during regular consultations schedules in the process of updating patient's records. According to the results of the literature, we assumed that this incentive might be more valued by GPs in the context of breast and cervical cancers screening, because GPs are not always informed of the results of the test (Liberalotto, 2012).

The Assistance (*ASSIST)* attribute was a qualified staff assistance designed to help GPs with cancer screening follow-up. It was thought as a 'material' support to doctors who, for instance, could run out of time in their daily practice and/or require additional assistance or technical staff.

The Payment (*PAY)* attribute was a financial device in line with P4P systems. It offered different levels of additional remuneration as a percentage increase of the year fees (0%, 1%, 3%, 5%) based on reaching a specific target of screened patients, patients for which the doctor was the referral GP. We chose the same targets as proposed by the National Health Authority (HAS, France's equivalent to NICE in the UK), i.e. 80% for breast and cervical cancer screening, and 50% for colorectal cancer screening.

Assuming the majority of GPs would prefer receiving support for promoting cancer screening, we expected all 5 attributes to have significant and positive effects on GPs' decisions. For example, offering training (*TRAIN*) would increase the probability of GPs being involved in cancer screening programmes.

3.2/ Design of the choice tasks

We used NGENE software (ChoiceMetrics) to generate a 12 choice tasks D-efficient design (94% D-efficiency) with non-informative (null) priors and allowing estimation of all main effects as well as two pre-specified interactions effects (*LEAF* x *PAY; ASSIST* x *LIST)* that were judged particularly relevant following previous interviews with GPs. We wanted to investigate a potential interaction effect between the Leaflet and the Payment, because providing and explaining leaflets to patients is time consuming and then GPs would need to be compensated for the income loss due to immediate decrease in their productivity. Despite this initial opportunity cost, GPs may be willing to provide leaflets for free (i.e., without needing a final compensation) because it can be seen as an investment that would help GPs save time during future consultations by empowering (or enabling) patients.

The second interaction effect between the Assistance and the Listing was motivated by the assumption that Listing might be more valued when accompanied with trained staff to help identify and raise the awareness of individuals not up to date with their screening. Conversely, Assistance might be more valuable if combined with up-to-date information about screening achievement.

A binary response format was used to create the choice tasks, in which participants faced one hypothetical screening programme at a time and were asked to indicate whether they would change their usual screening practice for the proposed scenario ('yes/no'). This choice format closely mimics decisions problems faced by GPs in their daily practice. We also offered to the participants the possibility to not choose between their current situation and the hypothetical one by answering "*I don't know*". This option was used to avoid forced choices that could be unreliable, and it can be seen as a structured way of providing missing values. An example of a choice task is presented in Figure 1. The same design was used for all three screening contexts (i.e., breast, cervical, and colorectal cancer).

3.3/ Sampling and recruitment

The DCE was included in an online survey sent to a listing of representative French GPs using the quota sampling method. Respondents accessed the questionnaire by clicking on a link that was included in an e-mail containing little information about the survey, in order to limit the selection of respondents. GPs were recruited by a survey company. All precautions were taken to ensure anonymity of the data, in agreement with the CNIL (Commission Nationale de l'Informatique et des Libertés, French law no. 78-17). The questionnaire took about 15 minutes to be completed, and respondents were remunerated €28 for their participation to offset the opportunity cost of its completion. Before responding to the DCE, each GP was randomly assigned to one screening context (i.e., breast, cervical, or colorectal cancer). For each type of cancer screening, a quota was applied based on age, gender, and practice location distribution: if the GP entered the quota, he/she could respond to the remaining of the questionnaire.

The sampling objective was to obtain a total of 400 completed questionnaires. The number of participants was computed using the Louviere et al (2000) approximate formulae for DCE sample size. For a choice probability of 50%, accuracy level of 10%, confidence level of 5% and 12 experimental tasks, the minimum number of respondents is 33 per type of cancer screening. Then we multiplied this minimum requirement by 4 to obtain a sample robust to individuals' idiosyncrasies and allowing for investigation of variability in preferences. The survey was addressed to approximately 4,000 GPs between March and April, 2014, until the expected number of completed questionnaires was reached. Among the 4,000 internet links sent, 685 were consulted and 402 resulted in a completed questionnaire (59% incidence rate). Before launching the survey, a pre-test survey was conducted among ten GPs to check the respondents' understanding of the attributes and their levels, as well as the questionnaire in general. As a result

of their feedback, we added the possibility to check the exhaustive definition of attributes at any time during completion of the choice experiment, and added a coloured bar so that respondents could check their progression in the questionnaire.

3.4/Discrete choices modelling

Question 1: What are GPs' preferences for both financial and non-financial incentives?

The GPs' preferences were measured within the random utility maximization framework (Manski, 1977; Thurstone, 1927) using a binary logit model, which is a specific case of the traditional multinomial logit model that not hold the restrictive independence of irrelevant alternatives assumption (McFadden, 1974). In each choice task t ($t=1,\ldots,T$), the response to cancer screening programme i ($i=\{yes,no\}$) by GP n ($n=1,\ldots,N$) was related to a utility function composed of a deterministic part (V_{nti}) and a stochastic part (ϵ_{nti}). Different functional forms were tested for the deterministic part of the utility function: first a linear additive utility function with no interactions between attributes (equation 1); second a multiplicative utility function with the two above-mentioned interactions between attributes (equation 2).

$$U_{nti} = \alpha + \beta_1 LEAF_t + \beta_2 TRAIN_t + \beta_3 LIST_t + \beta_4 ASSIST_t + \beta_5 PAY_t + \varepsilon_{nti} \quad (1)$$

$$U_{nti} = \alpha + \beta_1 LEAF_t + \beta_2 TRAIN_t + \beta_3 LIST_t + \beta_4 ASSIST_t + \beta_5 PAY_t + \gamma_1 PAY_t \times LEAF_t + \gamma_2 ASSIST_t \times LIST_t + \varepsilon_{nti} \quad (2)$$

In equation 1 and 2, $LEAF_t$, $TRAIN_t$, ..., PAY_t, represent the levels of the attributes presented in choice task t, α is a constant term, $\beta_1, \ldots, \beta_5$, are the part-worth utility coefficients for the five attributes, and γ_1, γ_2 are the coefficients of the interaction effects. We assumed a positive impact of the financial and non-financial attributes on GPs' utility function, as well as positive interaction effects. The PAY attribute was coded linearly and the non-financial attributes were dummy coded. The standard errors of the estimates were corrected for group clustering.

Question 2: Are GPs' preferences cancer-specific?

We compared the preferences between the three types of cancer using marginal rates of substitution (MRS) between the non-financial attributes and the payment. The MRS were derived from the results of the binary logit models with no interaction (equation 1). Computation of MRS values allowed normalisation of coefficients by eliminating the scale coefficient (i.e. the variance of the error term ϵ_{nti}) thus allowing direct comparison of mean preferences between cancer screening contexts. The 95% confidence intervals around the MRS were calculated using the

Delta method (Hole, 2007). The MRS between the PAY attribute and each non-financial attribute indicate the percentage of *potential* additional payment (percentage of their annual turnover) the GPs were willing to give up in exchange of the non-financial attribute *k*. This interpretation is very similar to a conclusion in terms of willingness-to-pay (WTP), except that WTP are generally computed with a fixed price coefficient that is expected to have a negative value and to be independent of individual actions, whereas the payment is expected to be positive and conditional on the GP's performance. To avoid any confusion, we used the MRS terminology.

Question 3: Do GPs differ in their preferences for screening incentives?

The binary logit model implicitly assumes that all GPs' hold same preferences for the proposed attributes of cancer screening programmes. This assumption can be misleading since there is evidence in the literature of considerable variability in GPs' involvement in cancer screening activities whatever the type of cancer (Bungener et al., 2010). In this section we relax the assumption of preferences homogeneity by fitting a Hierarchical Bayes mixed logit model (labelled HB-MXL) estimated through the Markov Chain Monte Carlo (MCMC) algorithm. There are several benefits of using the Bayesian version of the mixed logit model over the more traditional (i.e. frequentist) version (see (Regier et al., 2009) for a comparison of both methodologies). First the Bayesian approach allows making a distinction between preferences "heterogeneity" (i.e. between-GP variability in estimated preferences) and preferences "uncertainty" (i.e. within-GP variability in estimated preferences) (Daziano and Achtnicht, 2014). Second the HB-MXL approach allows estimation of the whole distribution of preferences rather than only the mean and standard deviation, what will facilitate the comparison of GPs' preferences across screening contexts. Eventually, it has been argued that Bayesian inference performs better than traditional frequentist approach when working on moderate to low sample sizes (Train, 2003).

The HB-MXL model was specified in MRS-space in order to facilitate computation of MRS coefficients and comparison of preferences across cancer screening contexts (Train and Weeks, 2005). The MRS-space HB-MXL model applied to our context is written as following:

$$U_{nti} = \alpha_n + \beta_{5n} PAY_t + \beta_{5n}(w_{1n} LEAF + w_{2n} TRAIN_t + w_{3n} LIST_t + w_{4n} ASSIST_t) + \varepsilon_{nti} \quad (3)$$

Where ($w_{kn} = \beta_{kn}/\beta_{5n}$) is the ratio of the non-financial attribute *k*'s coefficient to the PAY coefficient being interpreted directly as a marginal rate of substitution. The six random parameters ($\alpha_n, w_{1n}, \dots, w_{4n}, \beta_{5n}$) are assumed to follow a multivariate normal distribution with

full covariance matrix to be estimated. Because of the panel nature of data (i.e., multiple choices per GP), the HB-MXL model can also be used to recover preferences for each GP, allowing thus for comparison of MRS distribution across the three cancer screening contexts (see Train, 2003, chapter 12, for a detailed description of Bayesian procedures applied to discrete choice data). Details about estimation of equation 3 using the MCMC algorithm are provided in Appendix 1. The HB-MXL models were estimated using the RSGHB package written in R software (Dumont et al., 2015).

4. Results

4.1/ Samples of respondents

The initial samples included 135 GPs for the breast cancer screening questionnaire, 133 GPs for the cervical cancer screening questionnaire, and 134 GPs for the colorectal cancer screening questionnaire, for a total of 402 respondents. After exclusion of serial non-traders (i.e. GPs systematically answering '*yes*' or '*no*' in all choice tasks), the estimation samples included 108 GPs (80%) for the breast cancer questionnaire, 111 GPs (83%) for the cervical cancer questionnaire, and 114 GPs (85%) for the colorectal cancer questionnaire. Non-traders were excluded from the analyses because they provide no useful information for the estimation of GPs' preferences. We verified that exclusion of non-traders did not introduce a selection bias by comparing their characteristics to those of the included respondents. The included and excluded respondents were not significantly different (at the 5% level) in terms of age, gender, practice location, group practice, activity and screening practices (see Appendix 3 for test results).

Descriptive statistics of the three surveys' populations are displayed in Table 2. In the three samples, the included GPs are representative of the French GP population in terms of gender (72% are males), age (the mean age is 52 years *vs.* 50.1 years in the total French GP population), and practice location. The distribution of GPs in rural or urban areas, type of practice (group *vs* solo) and screening practices is similar from one cancer to another: about 66% of GPs work in urban areas, about 49% of GPs work in a group and about 71% of GPs declare checking whether the patient has performed a screening for the studied cancer on a very regular basis (i.e. 'systematically' or 'very often').

4.2/ Discrete choices modelling

Question 1: What are GPs' preferences for both financial and non-financial incentives?

The results of the binary logit models are displayed in Table 3 (the results do not differ when including non-traders in the models, see Supplementary file B for robustness checks). Information about the number (%) of *I don't know* responses, which were discarded from the subsequent analyses, are presented in Supplementary file C. The results of the main effects model for breast cancer show that all the attributes but Assistance (ASSIST) are significant and with the expected sign. The most weighted attributes are the Listing (*LIST*=0.84 [0.55; 1.13]) and the Training (*TRAIN*=0.73 [0.48; 0.99]). The two interaction effects *LEAF* x *PAY* and *ASSIST* x *LIST* are positive but not significant at the 5% level. Results of S-estimates show that inclusion of at least 319 GPs in the survey would have been necessary in order to obtain a statistically significant impact of ASSIST at the 5% level. Nonsignificance of ASSIST could thus be partially explained by the relatively low sample size.

Regarding cervical cancer screening, all attributes are significant at the 5% level and, similar to breast cancer screening, the most weighted attributes are the Listing (*LIST*=0.94 [0.63; 1.25]) and the Training *(TRAIN*=0.80 [0.53; 1.08]*)*. The interaction effects are also not significant at the 5% level.

In the colorectal cancer screening context, all the attributes but Assistance are significant at the 5% level, and GPs are mainly sensitive to the Listing (*LIST*=0.57 [0.29; 0.85]), Training (*TRAIN*=0.54 [0.27; 0.80]), and Payment (PAY=0.45 [0.37; 0.53]). There is a significant interaction effect between Listing and Assistance (*LIST x ASSIST*= 0.46 [0.04; 0.88]) meaning that combined, these two attributes are more weighted in GPs' utility function than when proposed separately. However, the results of the likelihood ratio (LR) tests comparing statistical performance of the choice models with and without the interaction effects indicate that interaction effects do not significantly improve the data fit (LR test: Chi-2=1.67; DF=2; P-value=0.4338). We also used LR tests to investigate the relevance of using nonlinear coding for the PAY attribute, by comparing two rival models using 1) a categorical coding scheme and 2) a polynomial function. Modelling nonlinearities did not significantly improve model fit for cervical and colorectal cancers, though it increased the model likelihood for breast cancer (results of the LR tests are available upon request). For parsimonious reason and to better compare the results across cancer screening settings, we decided to keep the linear coding scheme.

Question 2: Are GPs' preferences cancer-specific?

The results of the mean marginal rates of substitution (MRS) derived from the binary logit models are presented in Table 3. The GPs are systematically more sensitive to non-financial attributes in the context of breast and cervical cancers as compared to colorectal cancer screening, although the differences in mean MRS are not statistically significant at the 5% level. For instance, in the context of breast and cervical cancer screening, the GPs are willing to give up potential increases of 2.07% (95% CI: [1.23; 2.90%]) and 2.47% (95% CI: [1.48; 3.47%]) of their turnover, respectively, to beneficiate from the Training, compared to 1.20% (95% CI: [0.60-1.80%]) in the context of colorectal cancer screening. The MRS are the highest for the listing: the GPs are willing to give up potential increases of 2.36% (breast cancer), 2.89% (cervical cancer) and 1.27% (colorectal cancer) to benefit from the listing. Assistance is the least desirable attribute, and the MRS for this attribute is not significant for breast cancer screening (MRS = 0.45% [-0.40; 1.30%]) and colorectal cancer screening (MRS=0.04% [-0.53; 0.62]).

The comparison of preferences between cancer screening contexts is detailed by relaxing the assumption of preference homogeneity and analysing the results of the HB-MXL models. For each cancer, the cumulative density functions (CDFs) of MRS estimates are displayed in Figure 2, and the quartiles of the posterior distribution are presented in Table 4.

A visual analysis of the CDFs shows that the distributions of MRS differ across cancer screening contexts and type of screening incentive. The CDFs are most shifted for the Leaflet, with the MRS being the highest for cervical cancer at all points of the distribution. The Kolmogorov-Smirnov (KS) tests confirm that for this attribute, the distribution of MRS estimates are not equal across cancer screening contexts ($p<0.0001$). For instance, for cervical cancer, half of GPs have an estimated MRS for the leaflet > 1.61% as compared to 1.25% for breast cancer and 0.99% for colorectal cancer.

Regarding the other non-financial attributes, we find that the CDF for Training, Listing and Assistance are systematically more shifted toward higher MRS values in the context of breast and cervical cancer screening as compared to colorectal cancer screening. This result supports the conclusion that GPs are more sensitive to non-financial incentives for breast and cervical cancers. For instance, for cervical cancer, half of GPs have an estimated MRS for Training > 2.10% (respectively, 1.95% for breast cancer), compared to 1.15% for colorectal cancer. Statistically speaking, the KS tests comparing the CDFs of MRS estimates between gynaecological cancers (either breast or cervical) and colorectal cancer systematically reject the null hypothesis of equal distributions ($p<0.0001$). However, the CDFs are not significantly different (at the 1% level) between breast and cervical cancer for the Training ($p=0.0743$) and for the Listing ($p=0.0454$), showing that GPs' preferences are more similar in breast and cervical cancers screening contexts.

Question 3: Do GPs differ in their preferences for screening incentives?

In order to assess the level of preference heterogeneity (between-GP variability), we plotted the individual MRS estimates for each GP and in each cancer screening context. In addition, as an indicator of the level of preference heterogeneity, we calculated the interquartile range (IQR) as the difference between the upper and lower quartile of the MRS distribution (see Table 4). In Figure 3, there is evidence of greater preference heterogeneity for Training, Listing and Assistance in the context of breast and cervical cancers screening as compared to colorectal cancer screening. For instance, the IQR for the listing is 2.28 (Q1=1.11, Q3=3.39) for cervical cancer as compared to only 0.76 (Q1=0.73, Q3=1.49) for colorectal cancer.

In order to assess the level of preference uncertainty (within-GP variability), we calculated the quartiles of the MRS's standard deviation distribution (results are reported in Table 4). Complementarily, preference uncertainty was assessed graphically by plotting the estimated MRS plus / minus their standard deviation, for each GP and in each cancer screening context (see Figure 3). There is evidence of greater preference uncertainty in the context of breast and cervical cancers screening: at all point of the distribution, the standard deviations are higher for breast and cervical cancers as compared to colorectal cancer. For instance, the mean standard deviation of MRS estimates for the Training is 1.30 for cervical cancer, 1.21 for breast cancer and 0.85 for colorectal cancer. Looking at Figure 3, the range of MRS draws represented by the dotted lines is higher for breast and cervical cancers as compared to colorectal cancer, indicating higher preference uncertainty.

5. Discussion

To the best of our knowledge, this study is the first to investigate GPs' preferences for both financial and non-financial incentives for taking part to cancer screening activities. This study contributes to the literature by addressing three research questions. First, we examined how GPs make trade-offs between financial and non-financial incentives. Second, we analysed variations in GPs' preferences according to the cancer screening context by focusing on breast, cervical, and colorectal cancer. Finally, we quantified the variability in GPs' preferences for cancer screening incentives.

In this study, we showed that GPs' trade-offs for screening incentives depend on the type of cancer: no single attribute dominates all others for all cancers (including the financial attribute). This result is interesting because it seems not to have been demonstrated before, while it could partly explain the contrasting conclusions obtained in the literature. Our results indicate that in the context of breast and cervical cancers, GPs are particularly sensitive to receiving compensated training and to receive a bi-annual listing of screened patients to identify eligible women not to date with their screening. The conclusion holds when we take into account preference heterogeneity. This result is consistent with previous studies that highlighted the need for training (McIlfatrick et al., 2013) and for transmission of information particularly in the context of gynaecological cancers (Liberalotto, 2012).

Regarding cervical cancer screening, GPs are in competition with other doctors such as gynaecologists, who perform most of pap smears in France (Bungener et al., 2010). They may be looking for training and feedback on their practices to feel more confident about this act, thus explaining their high marginal rates of substitution for these services. It is interesting to note that these non-financial incentives would be workable in practice and would not necessarily be as expensive as financial incentives.

Regarding colorectal cancer screening, the additional payment had relatively more impact, even if the differences between cancers are not statistically significant on average. The result could be explained by the fact that the GP is expected to play an important role in explaining the modalities of the test and/or convincing all eligible patients to perform the test, and time has an opportunity cost requiring compensation. Indeed, according to a recent survey from the French National Institute of Cancer (INCa), 27% of GPs declare that the explanation of FOBT is time consuming (Bungener et al., 2010). Another argument may explain the result: colorectal cancer screening is destined to a relatively large population in the GP patient base (men and women aged 50 to 74 years). Thus, providing prevention to all by explaining and justifying screening will be time consuming and the achievement of target rates will probably come at a cost. Contrary to breast and cervical cancers, colorectal cancer screening has not been included in a P4P scheme yet: our results indicate that GPs are probably expecting financial rewards particularly for this cancer.

The joint influence of both financial and non-financial incentives was investigated by including two interactions as arguments of GPs' utility function. The first interaction was between the Payment (financial) and the Leaflet (non-financial) incentive and was never significant in none of the screening contexts. This result may be explained by the fact that the Leaflet was not an attractive incentive *per se* and the association with financial rewards did not

change its intrinsic value. That the Leaflet was less desirable than the other attributes is consistent with it not being an innovative device, because doctors as well as patients could already download documents produced and available online by reliable institutions such as the National Cancer Institute or the League against Cancer. This non significant result might also be explained considering that two opposing effects were at work: a positive interaction effect for GPs considering that the additional payment offsets the opportunity cost of providing leaflets, and a null or negative effect for GPs considering that leaflet is a long-term investment used to decrease time spent in future prevention consultations.

Then we investigated a potential interaction effect between two non-financial incentives, namely the Listing and the Assistance. The interaction effect was always positive in all screening contexts, but it was significant only in the colorectal cancer screening context. One possible rationale for this result is that, as stated previously, providing prevention to all by explaining and justifying screening may be more time consuming for colorectal cancer, thus the listing was more valued when accompanied with trained staff providing help to identify and raise the awareness of individuals not up to date with their screening. This non significant result could also be explained by the low valuation of the Assistance in all cancer screening contexts.

The analysis of preference heterogeneity was achieved by fitting Hierarchical Bayes mixed logit models. To our knowledge, it is the first application of this methodology in a health context that resulted in estimating subject-specific preferences and comparing the distribution of preferences in different cancer screening contexts. Interestingly, we have shown that the level of between- and within-GP variability depends on the type of cancer: there is greater preference heterogeneity and also greater preference uncertainty for screening incentives in the context of breast and cervical cancers as compared to colorectal cancer. This result may be explained by the fact the environment underlying breast and cervical cancer screening is more complex, especially because GPs are in competition with other health care professionals such as gynaecologists who prescribe most of mammograms and pap smears in France. As a consequence, the perception of the usefulness of a particular incentive as well as the degree of accuracy in elaborating preferences for cancer screening programmes may depend on exogeneous factors (e.g. the density of gynaecologists in the region indicating greater competition) and/or GPs' characteristics such as gender, experience, training, or specialised activity. Indeed, sociological studies have shown that female GPs show greater commitment towards breast and cervical cancers screening compared to male GPs (Liberalotto, 2012). However, due to sample size issues, we were not able to relate variations of preferences to GPs' individual characteristics. Eventually, current debates about the benefits and harms of breast cancer screening could translate into different attitudes and

preferences towards mammography, thus explaining greater preference heterogeneity and uncertainty.

Study limitations

The design of the survey using the quota method ensured that our samples of respondents were representative of the French GP population in terms of age, gender, and geographical location. However, respondents represented a sample of individuals who voluntarily accessed the questionnaire. This opt-in nature of the sample reduces the generalizability of the results and constitutes an important issue to be considered for the sampling of respondents in discrete choices surveys. Yet, we believe that the selection of respondents was limited as we achieved a high incidence rate (60%), meaning that 6 GP out of ten completed the entire questionnaire after clicking on the survey link. Moreover, the fact that the e-mail contained little information about the survey and that GPs were financially rewarded to answer the questionnaire certainly enabled to mitigate selection issues of GPs who, for instance, could be more involved in cancer screening activities or more interested by the subject. Removal of about 15-20% of serial non-traders is another limitation of our study. Yet the included and excluded GPs were not significantly different from one another and sensitivity analyses showed that removal of non-traders did not change the results (see Supplementary file B).

Regarding the estimation of single profiles DCE models, we assumed that the utility associated with the reference situation (the GP usual practice context) was null. This is a strong assumption as the 'current situation' is likely to largely differ among GPs. Yet, such differences were captured by the constant parameter of the utility function that was allowed to vary over the sample in the HB-MXL models. Thus, any differences in GPs' reference situation were reflected in variability of estimates of the constant parameter. Eventually one limitation of our approach is that we explicitly assumed that GPs' systematic preferences for current situation were normally distributed over the sample, whereas it might be the case that sample preferences would be best described by other continuous/finite distributions. This design was preferred to a more traditional pairwise comparison between two (or more) alternatives, first because we anticipated that GPs would have more difficulties comparing and inferring utilities from two or more hypothetical scenarios compared to one, second because of the fact they already experienced a situation where some attributes were available (e.g. leaflets, P4P) so that they could find it difficult to completely extract from their current situation.

Conclusion

The study provides meaningful results for policymakers wishing to increase GPs' involvement in breast, cervical, and colorectal cancer screening. Screening programmes that would favour GPs implication are both financial and non-financial; they include training and systematic communication of information between doctors and patients. Non-financial incentives could be stronger motivators for breast and cervical cancers, and financial incentives are likely to have more impact for colorectal cancer. Additional research is needed to assess the performance of these incentives and their combination in GPs' practice.

Acknowledgements

We benefited for this research from grants provided by the French National Institute for Cancer (INCa) (INCA_7014). We would like to thank Dr Diane Skatun, Mary Kilonzo, and the three anonymous reviewers for their useful comments on the paper.

References

Allenby, G.M., Rossi, P.E., 1998. Marketing models of consumer heterogeneity. J. Econom. 89, 57–78. doi:10.1016/S0304-4076(98)00055-4

Armour, B.S., Friedman, C., Pitts, M.M., Wike, J., Alley, L., Etchason, J., 2004. The influence of year-end bonuses on colorectal cancer screening. Am. J. Manag. Care 10, 617–624.

Bungener, M., Eisinger, F., Aubin-Auger, I., 2010. Médecins généralistes et dépistage des cancers. Synthèse des résultats de l'enquête barométrique INCa/BVA Septembre 2010.

Carrieri, V., Bilger, M., 2013. Preventive care: underused even when free. Is there something else at work? Appl. Econ. 45, 239–253. doi:10.1080/00036846.2011.597729

Clark, M.D., Determann, D., Petrou, S., Moro, D., Bekker-Grob, E.W. de, 2014. Discrete Choice Experiments in Health Economics: A Review of the Literature. PharmacoEconomics 32, 883–902. doi:10.1007/s40273-014-0170-x

Daziano, R.A., Achtnicht, M., 2014. Accounting for uncertainty in willingness to pay for environmental benefits. Energy Econ. 44, 166–177. doi:10.1016/j.eneco.2014.03.023

de Bekker-Grob, E.W., Ryan, M., Gerard, K., 2012. Discrete choice experiments in health economics: a review of the literature. Health Econ. 21, 145–172. doi:10.1002/hec.1697

Dumont, J., Keller, J., Carpenter, C., 2015. RSGHB: Functions for Hierarchical Bayesian Estimation: A Flexible Approach.

Eijkenaar, F., Emmert, M., Scheppach, M., Schöffski, O., 2013. Effects of pay for performance in health care: A systematic review of systematic reviews. Health Policy Amst. Neth. 110, 115–130. doi:10.1016/j.healthpol.2013.01.008

Ellis, P., Robinson, P., Ciliska, D., Armour, T., Brouwers, M., O'Brien, M.A., Sussman, J., Raina, P., 2005. A Systematic Review of Studies Evaluating Diffusion and Dissemination of Selected Cancer Control Interventions. Health Psychol. 24, 488–500. doi:10.1037/0278-6133.24.5.488

Federici, A., Giorgi Rossi, P., Bartolozzi, F., Farchi, S., Borgia, P., Guasticchi, G., Guastcchi, G., 2005. Survey on colorectal cancer screening knowledge, attitudes, and practices of general practice physicians in Lazio, Italy. Prev. Med. 41, 30–35. doi:10.1016/j.ypmed.2004.11.010

Franc, C., Lesur, R., 2004. Systèmes de rémunération des médecins et incitations à la prévention. Rev. Économique Vol. 55, 901–922. doi:10.3917/reco.555.0901

Ganry, O., Boche, T., 2005. Prevention practices and cancer screening among general practitioners in Picardy, France. Public Health 119, 1023–1030. doi:10.1016/j.puhe.2005.02.004

Gellad, Z.F., Provenzale, D., 2010. Colorectal cancer: national and international perspective on the burden of disease and public health impact. Gastroenterology 138, 2177–2190. doi:10.1053/j.gastro.2010.01.056

Gøtzsche, P.C., Nielsen, M., 2011. Screening for breast cancer with mammography. Cochrane Database Syst. Rev. CD001877. doi:10.1002/14651858.CD001877.pub4

Grimshaw, J.M., Shirran, L., Thomas, R., Mowatt, G., Fraser, C., Bero, L., Grilli, R., Harvey, E., Oxman, A., O'Brien, M.A., 2001. Changing provider behavior: an overview of systematic reviews of interventions. Med. Care 39, II2-45.

Günther, O.H., Kürstein, B., Riedel-Heller, S.G., König, H.-H., 2010. The role of monetary and nonmonetary incentives on the choice of practice establishment: a stated preference study of young physicians in Germany. Health Serv. Res. 45, 212–229. doi:10.1111/j.1475-6773.2009.01045.x

Heidelberger, P., Welch, P.D., 1983. Simulation Run Length Control in the Presence of an Initial Transient. Oper. Res. 31, 1109–1144.

Hennig-Schmidt, H., Selten, R., Wiesen, D., 2011. How payment systems affect physicians' provision behaviour--an experimental investigation. J. Health Econ. 30, 637–646. doi:10.1016/j.jhealeco.2011.05.001

Hensher, D.A., Rose, J.M., Greene, W.H., 2015. Applied Choice Analysis: A Primer. Cambridge University Press, Cambridge.

Hoffman, R.M., Lewis, C.L., Pignone, M.P., Couper, M.P., Barry, M.J., Elmore, J.G., Levin, C.A., Van Hoewyk, J., Zikmund-Fisher, B.J., 2010. Decision-making processes for breast, colorectal, and prostate cancer screening: the DECISIONS survey. Med. Decis. Mak. Int. J. Soc. Med. Decis. Mak. 30, 53S–64S. doi:10.1177/0272989X10378701

Hole, A.R., 2007. A comparison of approaches to estimating confidence intervals for willingness to pay measures. Health Econ. 16, 827–840. doi:10.1002/hec.1197

Holte, J.H., Kjaer, T., Abelsen, B., Olsen, J.A., 2015. The impact of pecuniary and non-pecuniary incentives for attracting young doctors to rural general practice. Soc. Sci. Med. 128, 1–9. doi:10.1016/j.socscimed.2014.12.022

Janus, K., 2010. Managing motivation among health care professionals. Adv. Health Care Manag. 9, 47–77. doi:10.1108/S1474-8231(2010)0000009007

Jensen, L.F., Mukai, T.O., Andersen, B., Vedsted, P., 2012. The association between general practitioners' attitudes towards breast cancer screening and women's screening participation. BMC Cancer 12, 254. doi:10.1186/1471-2407-12-254

Kass, R.E., Carlin, B.P., Gelman, A., Neal, R.M., 1998. Markov Chain Monte Carlo in Practice: A Roundtable Discussion. Am. Stat. 52, 93–100. doi:10.1080/00031305.1998.10480547

Kerlikowske, K., Grady, D., Rubin, S.M., Sandrock, C., Ernster, V.L., 1995. Efficacy of screening mammography. A meta-analysis. JAMA J. Am. Med. Assoc. 273, 149–154.

Kiran, T., Wilton, A.S., Moineddin, R., Paszat, L., Glazier, R.H., 2014. Effect of payment incentives on cancer screening in ontario primary care. Ann. Fam. Med. 12, 317–323. doi:10.1370/afm.1664

Lancaster, K., 1966. A new approach to consumer theory. J Polit Econ 134–57.

Li, J., Hurley, J., DeCicca, P., Buckley, G., 2014. Physician response to pay-for-performance: evidence from a natural experiment. Health Econ. 23, 962–978. doi:10.1002/hec.2971

Liberalotto, N.A., 2012. L'engagement des médecins généralistes à l'égard du dépistage des cancers féminins: un révélateur de leurs positionnements face aux transformations de leur contexte d'exercice. Ph.D thesis. EHESS. 510 pages.

Mannion, R., Davies, H.T.O., 2008. Payment for performance in health care. BMJ 336, 306–308. doi:10.1136/bmj.39463.454815.94

Manski, C.F., 1977. The structure of random utility models. Theory Decis. 8, 229–254. doi:10.1007/BF00133443

McFadden, D., 1974. Conditional Logit Analysis of Qualitative Choice Behavior., in: Frontiers in Econometrics. Academic Press: New York, pp. 105–142.

McIlfatrick, S., Keeney, S., McKenna, H., McCarley, N., McElwee, G., 2013. Investigating the role of the general practitioner in cancer prevention: a mixed methods study. BMC Fam. Pract. 14, 58. doi:10.1186/1471-2296-14-58

Quinn, M., Babb, P., Jones, J., Allen, E., 1999. Effect of screening on incidence of and mortality from cancer of cervix in England: evaluation based on routinely collected statistics. BMJ 318, 904–908.

Regier, D.A., Ryan, M., Phimister, E., Marra, C.A., 2009. Bayesian and classical estimation of mixed logit: An application to genetic testing. J. Health Econ. 28, 598–610. doi:10.1016/j.jhealeco.2008.11.003

Sabatino, S.A., Habarta, N., Baron, R.C., Coates, R.J., Rimer, B.K., Kerner, J., Coughlin, S.S., Kalra, G.P., Chattopadhyay, S., 2008. Interventions to increase recommendation and delivery of screening for breast, cervical, and colorectal cancers by healthcare providers

systematic reviews of provider assessment and feedback and provider incentives. Am. J. Prev. Med. 35, S67-74. doi:10.1016/j.amepre.2008.04.008

Sabatino, S.A., Lawrence, B., Elder, R., Mercer, S.L., Wilson, K.M., DeVinney, B., Melillo, S., Carvalho, M., Taplin, S., Bastani, R., Rimer, B.K., Vernon, S.W., Melvin, C.L., Taylor, V., Fernandez, M., Glanz, K., 2012. Effectiveness of interventions to increase screening for breast, cervical, and colorectal cancers: nine updated systematic reviews for the guide to community preventive services. Am. J. Prev. Med. 43, 97–118. doi:10.1016/j.amepre.2012.04.009

Saslow, D., Solomon, D., Lawson, H.W., Killackey, M., Kulasingam, S.L., Cain, J., Garcia, F.A.R., Moriarty, A.T., Waxman, A.G., Wilbur, D.C., Wentzensen, N., Downs, L.S., Jr, Spitzer, M., Moscicki, A.-B., Franco, E.L., Stoler, M.H., Schiffman, M., Castle, P.E., Myers, E.R., 2012. American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology screening guidelines for the prevention and early detection of cervical cancer. Am. J. Clin. Pathol. 137, 516–542. doi:10.1309/AJCPTGD94EVRSJCG

Scott, A., 2001. Eliciting GPs' preferences for pecuniary and non-pecuniary job characteristics. J. Health Econ. 20, 329–347.

Sicsic, J., Franc, C., 2016. Impact assessment of a pay-for-performance program on breast cancer screening in France using micro data. Eur. J. Health Econ. HEPAC Health Econ. Prev. Care. doi:10.1007/s10198-016-0813-2

Sicsic, J., Le Vaillant, M., Franc, C., 2012. Intrinsic and extrinsic motivations in primary care: an explanatory study among French general practitioners. Health Policy Amst. Neth. 108, 140–148. doi:10.1016/j.healthpol.2012.08.020

Thurstone, L.L., 1927. A law of comparative judgment. Psychol. Rev. 273–286.

Town, R., Kane, R., Johnson, P., Butler, M., 2005. Economic incentives and physicians' delivery of preventive care: A systematic review. Am. J. Prev. Med. 28, 234–240. doi:10.1016/j.amepre.2004.10.013

Train, K., 2003. Discrete Choice Methods with Simulation. Cambridge University Press, New York.

Train, K., Weeks, M., 2005. Discrete Choice Models in Preference Space and Willingness-to-Pay Space, in: Scarpa, R., Alberini, A. (Eds.), Applications of Simulation Methods in Environmental and Resource Economics, The Economics of Non-Market Goods and Resources. Springer Netherlands, pp. 1–16.

Weller, D., 1997. Cancer screening in general practice. Aust. Fam. Physician 26, 517–519, 522–525, 527.

WHO, 2008. The Global Burden of Disease: 2004 Update. World Health Organization.

Zapka, J.G., Lemon, S.C., 2004. Interventions for patients, providers, and health care organizations. Cancer 101, 1165–1187. doi:10.1002/cncr.20504

Tables

Table 1. Summary of attributes, coding scheme, and expected effects

Attributes	Definition	Levels	Coding	Expected effect
Leaflet (LEAF)	Providing GPs with cancer screening leaflets	yes, no	dummy	$\beta 1 > 0$
Training (TRAIN)	Compensated targeted training for the GP	yes, no	dummy	$\beta 2 > 0$
Listing (LIST)	Bi-annual listing of screened patients	yes, no	dummy	$\beta 3 > 0$
Assistance (ASSIST)	Qualified staff assistance	yes, no	dummy	$\beta 4 > 0$
Payment (PAY)	Additional payment based on patients' screening rates improvement	0, 1, 3, 5 in %	linear	$\beta 5 > 0$

Table 2. Descriptive statistics of the three samples of respondents (after exclusion of serial non-traders)

	Breast cancer		Cervical cancer		Colorectal cancer		Overall	
	N	*%*	*N*	*%*	*N*	*%*	*N*	*%*
Gender								
Female	30	27.8	30	27.0	32	28.1	92	27.6
Male	78	72.2	81	73.0	82	71.9	241	72.4
Age								
<45 years	18	16.7	24	21.6	25	21.9	67	20.1
45-54 years	44	40.7	34	30.6	46	40.4	124	37.2
>54 years	46	42.6	53	47.8	43	37.7	142	42.6
Practice location								
Paris neighborhood	18	16.7	20	18.0	22	19.3	60	18.0
North	4	3.7	8	7.2	10	8.8	22	6.6
East	22	20.4	32	28.8	24	21.1	78	23.4
West	24	22.2	21	18.9	20	17.5	65	19.5
South	40	37.0	30	27.0	38	33.3	108	32.4
Type of location								
Rural (<=10,000 inhabitants)	34	31.5	36	32.4	44	38.6	114	34.2
Urban (>10,000 inhabitants)	74	68.5	75	67.6	70	61.4	219	65.8
Type of practice								
Solo	58	53.7	56	50.5	56	49.1	170	51.1
Group	50	46.3	55	49.6	58	50.9	163	48.9
Number of hours worked /week								
<= 50 hours	48	44.4	61	54.9	59	51.7	168	50.5
>50 hours	60	55.6	50	45.1	55	48.3	165	49.5
Screening practices								
Systematic	42	38.9	42	37.8	37	32.5	121	36.3
Very often	34	31.5	40	36.0	44	38.6	118	35.4
Often	28	25.9	22	19.8	25	21.9	75	22.5
Sometines/never	4	3.7	7	6.3	8	7.0	19	5.7
Total	108	100	111	100	114	100	333	100

Table 3. Estimation results for the binary logit models

	Breast cancer		Cervical cancer		Colorectal cancer	
	Main effects (1)	Interactions (2)	Main effects (1)	Interactions (2)	Main effects (1)	Interactions (2)
	Coef. [95% CI]	*Coef. [95% CI]*	*Coef. [95% CI]*	*Coef. [95% CI]*	*Coef. [95% CI]*	*Coef. [95% CI]*
1. Preferences estimates						
Intercept	-1.77 [-2.22; -1.32]	-1.61 [-2.09; -1.13]	-1.82 [-2.22; -1.41]	-1.75 [-2.18; -1.32]	-1.70 [-2.12; -1.29]	-1.49 [-1.93; -1.06]
LEAF	0.42 [0.21; 0.63]	0.27 [-0.03; 0.56]	0.51 [0.28; 0.74]	0.56 [0.26; 0.85]	0.30 [0.06; 0.54]	0.15 [-0.16; 0.46]
TRAIN	0.73 [0.48; 0.99]	0.67 [0.42; 0.93]	0.80 [0.53; 1.08]	0.79 [0.51; 1.06]	0.54 [0.27; 0.80]	0.46 [0.20; 0.73]
LIST	0.84 [0.55; 1.13]	0.71 [0.33; 1.08]	0.94 [0.63; 1.25]	0.84 [0.46; 1.21]	0.57 [0.29; 0.85]	0.36 [0.03; 0.70]
ASSIST	0.16 [-0.14; 0.46]	0.03 [-0.37; 0.43]	0.46 [0.17; 0.75]	0.35 [-0.05; 0.75]	0.02 [-0.24; 0.28]	-0.21 [-0.55; 0.15]
PAY	0.35 [0.27; 0.44]	0.33 [0.25; 0.42]	0.33 [0.24; 0.41]	0.33 [0.25; 0.42]	0.45 [0.37; 0.53]	0.43 [0.35; 0.51]
LEAF×PAY	-	0.05 [-0.07; 0.17]	-	-0.04 [-0.16; 0.08]	-	0.04 [-0.08; 0.15]
LIST×ASSIST	-	0.31 [-0.09; 0.71]	-	0.18 [-0.23; 0.59]	-	0.46 [0.04; 0.88]
2. MRS estimates (in %)						
LEAF	1.18 [0.61; 1.23]	-	1.57 [0.85; 2.28]	-	0.67 [0.14; 1.19]	-
TRAIN	2.07 [1.23; 2.90]	-	2.47 [1.48; 3.47]	-	1.20 [0.60; 1.80]	-
LIST	2.36 [1.44; 3.28]	-	2.89 [1.67; 4.10]	-	1.27 [0.60; 1.92]	-
ASSIST	0.45 [-0.40; 1.30]	-	1.41 [0.45; 2.37]	-	0.04 [-0.53; 0.62]	-
3. Model statistics						
Number of GPs	108	108	111	111	114	114
Number of observations	1180	1180	1222	1222	1251	1251
Log-Likelihood	-717.45	-716.45	-734.31	-734.07	-748.39	-746.72
Pseudo R2	0.122	0.123	0.126	0.126	0.137	0.139
S-estimates[a]	319	5683	46	11,577	17,477	29,806

[a] The S-estimate is the minimum sample size for estimating significant parameters. It is the maximum of all Sp-estimates, where Sp-estimates indicate the number of respondents needed to obtain a significant estimate of each parameter of the utility function (Hensher et al., 2015). For instance, in order to obtain a significant effect of ASSIST (the attribute with lowest t-ratio), it would have required 319 respondents for breast cancer, and 17,477 respondents for colorectal cancer. Similarly, in order to estimate a significant interaction between LEAF and PAY, it would have required 5683 respondents for breast cancer, 11,577 respondents for cervical cancer, and 29,806 respondents for colorectal cancer.

Table 4. Quartiles of the posterior distribution of individual MRS estimates – HB-MXL results

		Breast cancer		**Cervical cancer**		**Colorectal cancer**	
		MRS (%)		MRS (%)		MRS (%)	
Attribute	Quartiles	*Est* [a]	*SD* [b]	*Est*	*SD*	*Est*	*SD*
LEAF	Q1	1.06	0.67	1.36	0.79	0.79	0.71
	Mean	1.27	0.75	1.61	0.85	1.01	0.79
	Q2 (Median)	1.25	0.75	1.61	0.84	0.99	0.76
	Q3	1.42	0.79	1.91	0.90	1.28	0.82
	IQR [c] (Q3-Q1)	*0.36*		*0.55*		*0.49*	
TRAIN	Q1	1.49	1.04	1.63	1.17	0.81	0.77
	Mean	1.89	1.21	2.18	1.30	1.14	0.85
	Q2 (Median)	1.95	1.16	2.10	1.29	1.15	0.82
	Q3	2.30	1.30	2.87	1.36	1.46	0.90
	IQR (Q3-Q1) [c]	*0.81*		*1.24*		*0.65*	
LIST	Q1	1.25	1.37	1.11	1.75	0.73	0.98
	Mean	2.01	1.57	2.28	1.98	1.17	1.13
	Q2 (Median)	1.88	1.53	1.83	1.92	1.14	1.07
	Q3	2.62	1.65	3.39	2.10	1.49	1.20
	IQR (Q3-Q1) [c]	*1.37*		*2.28*		*0.76*	
ASSIST	Q1	-0.29	1.43	0.85	1.23	0.00	0.77
	Mean	0.9	1.70	1.36	1.34	0.26	0.85
	Q2 (Median)	0.98	1.60	1.47	1.33	0.28	0.82
	Q3	1.62	1.78	1.90	1.40	0.55	0.89
	IQR (Q3-Q1) [c]	*1.91*		*1.05*		*0.55*	
PAY	Q1	0.65	0.44	0.58	0.32	0.63	0.39
	Mean	0.86	0.48	0.73	0.36	0.83	0.42
	Q2 (Median)	1.02	0.49	0.79	0.37	0.95	0.43
	Q3	1.19	0.52	0.97	0.40	1.11	0.46
	IQR (Q3-Q1) [c]	*0.54*		*0.39*		*0.48*	

[a] Mean MRS values calculated for each GP over the 5,000 MCMC draws retained after convergence.

[b] Standard deviation from mean MRS values calculated for each GP over the 5,000 MCMC draws retained after convergence.

[c] Interquartile range : statistical measure of the dispersion of GPs' MRS values.

Lecture*:* For breast cancer, half of GPs have a mean MRS value for the Leaflet higher than 1.25 (median), and 25% of GPs have a mean MRS value higher than 1.42 (Q3). The difference between the upper quartile and the lower quartile (IQR) of the MRS distribution for the Leaflet is 0.36.

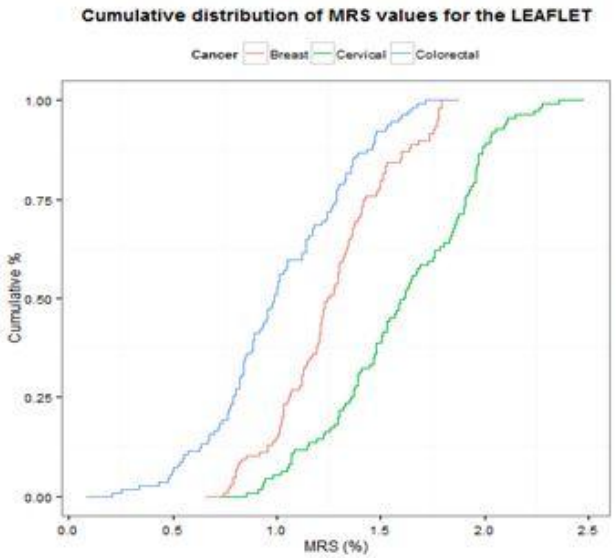
Figures

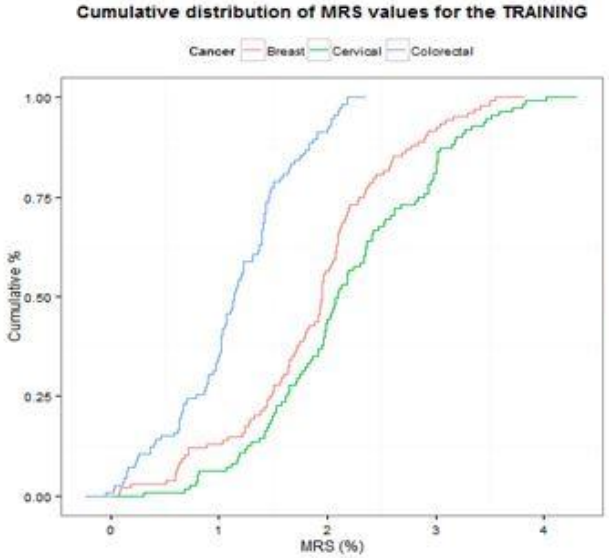
Figure 1. Example of a choice task

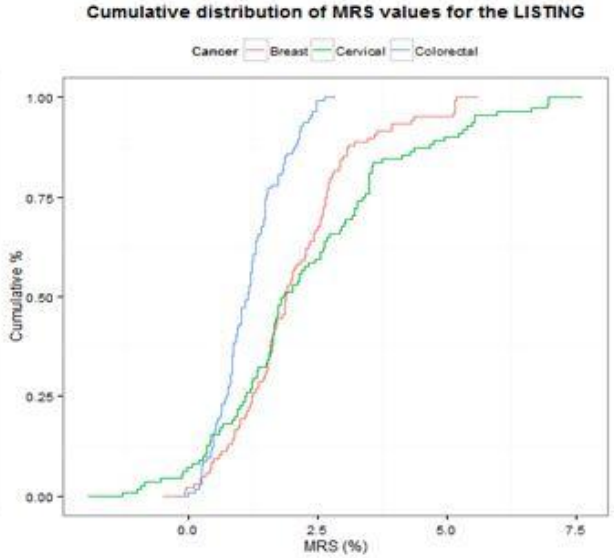
Cancer screening programme A
You are provided with [*] cancer screening leaflets for your patients
You are being financed a training for [*] cancer screening
You don't have access to a bi-annual listing of screened patients
You don't benefit from a qualified staff assistance
You benefit from a 3% additional payment based on targeted screening objective
Choice : Would you change your usual screening practice for the proposed cancer screening programme ?
□ Yes
□ No
□ I don't know

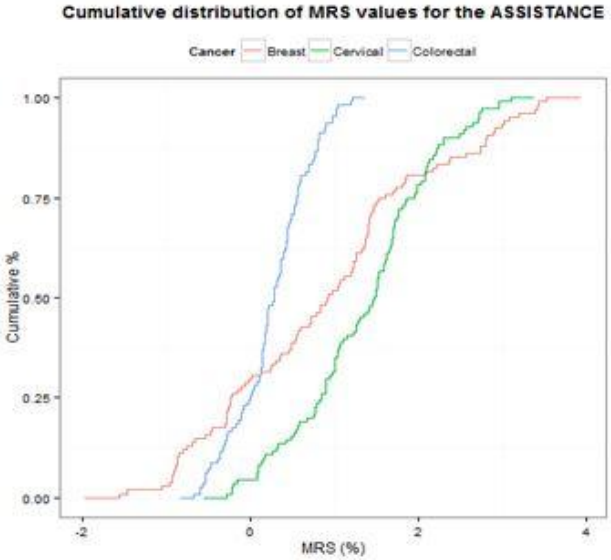
[*] was replaced by either 'breast', 'cervical, or 'colorectal' according to the questionnaire version

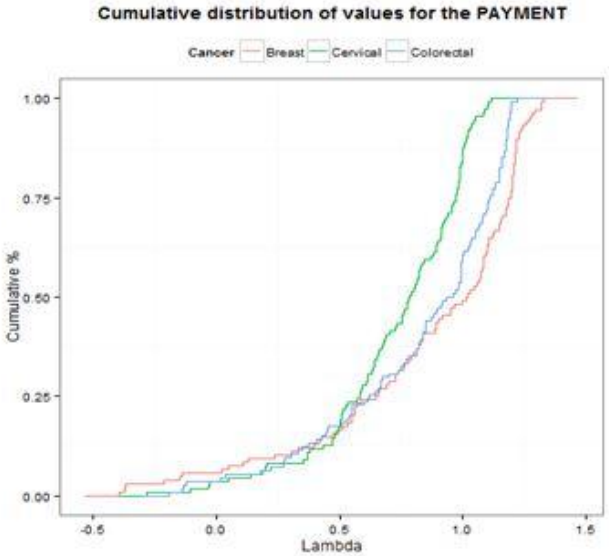
Figure 2. Cumulative density functions (CDFs) of individual MRS estimates across cancer screening contexts







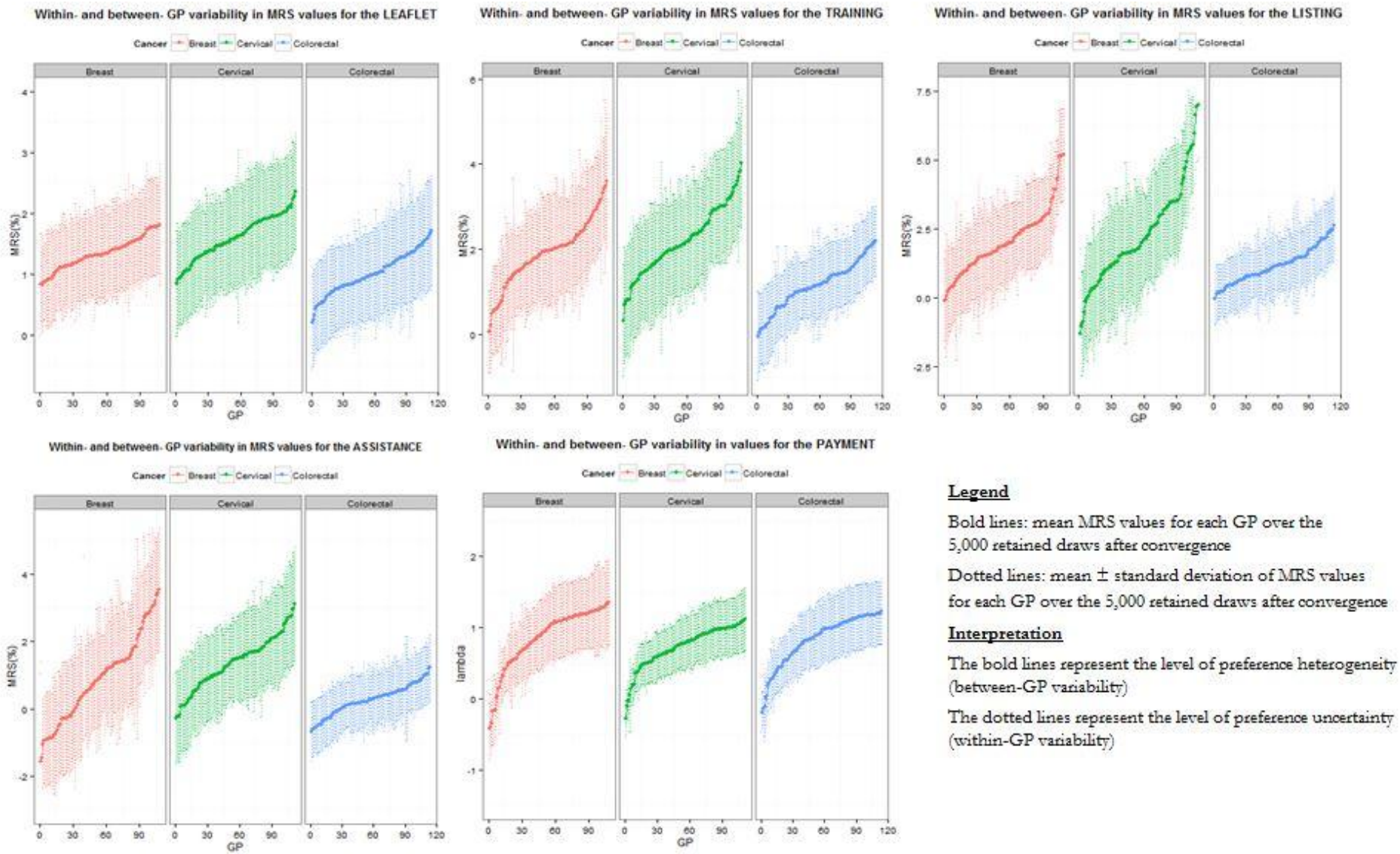




Equality of distribution (Kolmogorov-Smirnov test : ks)

LEAFLET	ks(breast *vs* cervical) : p-value <0.0001 ks(breast *vs* colorectal) : p-value <0.0001 ks(cervical *vs* colorectal) : p-value <0.0001
TRAINING	ks(breast *vs* cervical) : p-value 0.0743 ks(breast *vs* colorectal) : p-value <0.0001 ks(cervical *vs* colorectal) : p-value <0.0001
LISTING	ks(breast *vs* cervical) : p-value 0.0454 ks(breast *vs* colorectal) : p-value <0.0001 ks(cervical *vs* colorectal) : p-value <0.0001
ASSISTANCE	ks(breast *vs* cervical) : p-value = 0.0012 ks(breast *vs* colorectal) : p-value <0.0001 ks(cervical *vs* colorectal) : p-value <0.0001
PAYMENT	ks(breast *vs* cervical) : p-value <0.0001 ks(breast *vs* colorectal) : p-value = 0.0050 ks(cervical *vs* colorectal) : p-value <0.0001

Figure 3. Within- and between- GP variability in individual MRS values across cancer screening contexts



Legend

Bold lines: mean MRS values for each GP over the 5,000 retained draws after convergence

Dotted lines: mean ± standard deviation of MRS values for each GP over the 5,000 retained draws after convergence

Interpretation

The bold lines represent the level of preference heterogeneity (between-GP variability)

The dotted lines represent the level of preference uncertainty (within-GP variability)

Appendix

Appendix 1. Details on the MCMC procedure to estimate the HB-MXL models (equation 3)

The MCMC algorithm was used to estimate the joint posterior distribution of the coefficients given the data (Allenby and Rossi, 1998). The Markov chain was initialized using non-informative priors: a centred normal distribution with large variance for the random parameters' mean, and an inverted Wishart distribution for the random parameters' standard deviation (Train, 2003). The random parameters were assumed to follow a multivariate normal distribution with vector of mean parameters and matrix of variance-covariance parameters to be estimated. We used a full covariance matrix (6x6=36 elements) to model variability in GPs preferences for the screening attributes and how these preferences relate with each other. This is another advantage of the bayesian approach, which allows easily exploring parameters space with many dimensions.

The main issue when using MCMC is to determine when the Markov chain has converged. In order to check that convergence was attained, we inspected the traceplot to see if the draws traversed the posterior (Train, 2003) as well as the autocorrelation over the draws. We also performed the Heidelberger-Welch test that assesses the convergence of the Markov chain by testing the hypothesis that the chain comes from a stationary process (Heidelberger and Welch, 1983). These diagnostics were used to define the length of the 'burn-in period' (i.e. the number of iterations occurring prior to convergence that were discarded), the number of iterations that were used to conduct inference, and the proportion of iterations that were retained after convergence to reduce correlation across the Markov chain (Kass et al., 1998).

In the cervical and colorectal cancer subsamples, 50,000 iterations were used as 'burn-in' after which every 10-th draw was retained from 50,000 additional iterations, providing a total of 5,000 draws from the posterior distribution of the parameters. In the breast cancer subsample, 50,000 iterations were used as 'burn-in', but autocorrelation was more important and convergence could only be achieved by retaining every 20-th draw from 100,000 additional iterations, providing a total of 5,000 draws to conduct inference (see Appendix 2 for a summary of convergence diagnostics).

Appendix 2. Summary of the MCMC convergence diagnostics

	Heidleberger test		*Burn-in*	*Iter. after*	*Iter. kept after*
	Result	*p-value*	*period*	*convergence*	*convergence*
Breast cancer screening					
Intercept	passed	0.806	50,000	100,000	5,000
LEAF	passed	0.318	50,000	100,000	5,000
TRAIN	passed	0.602	50,000	100,000	5,000
LIST	passed	0.312	50,000	100,000	5,000
ASSIST	passed	0.409	50,000	100,000	5,000
PAY	passed	0.529	50,000	100,000	5,000
Cervical cancer screening					
Intercept	passed	0.281	50,000	50,000	5,000
LEAF	passed	0.124	50,000	50,000	5,000
TRAIN	passed	0.293	50,000	50,000	5,000
LIST	passed	0.181	50,000	50,000	5,000
ASSIST	passed	0.128	50,000	50,000	5,000
PAY	passed	0.341	50,000	50,000	5,000
Colorectal cancer screening					
Intercept	passed	0.471	50,000	50,000	5,000
LEAF	passed	0.637	50,000	50,000	5,000
TRAIN	passed	0.606	50,000	50,000	5,000
LIST	passed	0.606	50,000	50,000	5,000
ASSIST	passed	0.052	50,000	50,000	5,000
PAY	passed	0.606	50,000	50,000	5,000

Appendix 3. Comparison of the characteristics of traders and serial non-traders

	Non-traders		Traders		Overall		Chi-square test
	N	*col %*	*N*	*col %*	*N*	*%*	*P-value*
Gender							**0.542**
Female	20	31.3*	93	27.5	113	28.1	
Male	44	68.8	245	72.5	289	71.9	
Age							**0.521**
<45 years	13	20.3	68	20.1	81	20.2	
45-54 years	32	50.0	146	43.2	178	44.3	
>54 years							
Practice location							**0.085**
Paris neighborhood	12	18.8	60	17.8	72	17.9	
North	5	7.8	22	6.5	27	6.7	
East	21	32.8	79	23.4	100	24.9	
West	16	25.0	66	19.5	82	20.4	
South							
Type of location							**0.592**
Rural (<=10,000 inhabitants)	24	37.5	115	34.0	139	34.6	
Urban (>10,000 inhabitants)	40	62.5	223	65.9	263	65.4	
Type of practice							**0.213**
Solo	38	59.4	172	50.9	210	52.2	
Group	26	40.6	166	49.1	192	47.8	
Number of hours worked /week							**0.965**
<= 50 hours	32	50.0	170	50.3	202	50.3	
>50 hours	32	50.0	168	49.7	200	49.8	
Screening practices							**0.602**
Systematic	28	43.8	121	35.8	149	37.1	
Very often	21	32.8	121	35.8	142	35.3	
Often	11	17.2	77	22.8	88	21.9	
Sometines/never	4	6.3	19	5.6	23	5.7	
Trading behaviour							
Always 'yes'	37	57.8	-	-	-	-	
Alwayes 'no'	27	42.2	-	-	-	-	
Total	64	15.9	338	84.1	402	100.0	

Lecture (*): among non-traders, 31.3% were females, compared to 28.1% in the overall population. Among non-traders, 57.8% systematically responded 'yes' to the choice questions, and 42.2% systematically responded 'no'.

The significance of bold represents the Chi-square independence test for the association between the individual characteristics and the trading behaviour. A p-value<0.05 indicate a significant relationship.

Supplementary files

Supplementary file A. Full description of attributes in each cancer screening context.

BREAST CANCER	
Attributes	***Definition***
Leaflet	Provision of cancer screening leaflets for your patients, containing information about breast cancer and mammography
Training	Compensated training including the following items: - Analysis of difficulties in screening uptake, strategies of conviction/application, coordination with other specialists - Benefits and harms balance of mammography, degree of treatment and quality of life according to prognosis
Listing	Bi-annual listing of patients aged 50-74 years having performed a mamogram in the past two years
Assistance	Qualified staff assistance to monitor the screening of your patients, including either: - Presence of taff for the reception/advice of patients, twice a month - Networking with a call centre to contact or revive patients not to date with their screening
Payment	Additional payment conditional or reaching (or drawing closer to) 80% of female patients between 50 and 74 years of age having performed a mammogram in the past two years

CERIVICAL CANCER	
Attributes	***Definition***
Leaflet	Provision of cancer screening leaflets for your patients, containing information about cervical cancer and Pap smears
Training	Compensated training including the following items: - Analysis of difficulties in screening uptake, strategies of conviction/application, coordination with other specialists - Realisation of Pap smears, degree of treatment and quality of life according to prognosis
Listing	Bi-annual listing of patients aged 25-65 years having performed a Pap smears in the past three years
Assistance	Qualified staff assistance to monitor the screening of your patients, including either: - Presence of taff for the reception/advice of patients, twice a month - Networking with a call centre to contact or revive patients not to date with their screening
Payment	Additional payment conditional or reaching (or drawing closer to) 80% of female patients between 25 and 65 years of age having performed a Pap smears in the past three years

COLORECTAL CANCER	
Attributes	***Definition***
Leaflet	Provision of cancer screening leaflets for your patients, containing information about colorectal cancer and Hemoccult tests
Training	Compensated training including the following items: - Analysis of difficulties in screening uptake, strategies of conviction/application, coordination with other specialists - Benefits of early detection, degree of treatment and quality of life according to prognosis
Listing	Bi-annual listing of patients aged 50-74 years having performed an hemoccult test in the past two years
Assistance	Qualified staff assistance to monitor the screening of your patients, including either: - Presence of taff for the reception/advice of patients, twice a month - Networking with a call centre to contact or revive patients not to date with their screening
Payment	Additional payment conditional or reaching (or drawing closer to) 50% of patients (males and females) between 50 and 74 years of age having performed an Hemoccult test in the past two years

Supplementary file B. Results of the binary logit model among the general population (including non-traders)

	Breast cancer		Cervical cancer		Colorectal cancer	
	Main effects (1)	Interactions (2)	Main effects (1)	Interactions (2)	Main effects (1)	Interactions (2)
	Coef. [95% CI]	*Coef. [95% CI]*	*Coef. [95% CI]*	*Coef. [95% CI]*	*Coef. [95% CI]*	*Coef. [95% CI]*
1. Preferences estimates						
Intercept	-1.53 [-1.90; -1.17]	-1.41 [-1.81; -1.03]	-1.50 [-1.87; -1.15]	-1.47 [-1.85; -1.10]	-1.50 [-1.88; -1.13]	-1.32 [-1.71; -0.94]
LEAF	0.33 [0.16; 0.49]	0.32 [-0.07; 0.57]	0.39 [0.20; 0.58]	0.50 [0.25; 0.75]	0.25 [0.06; 0.45]	0.16 [-0.10; 0.42]
TRAIN	0.59 [0.38; 0.79]	0.55 [0.35; 0.75]	0.63 [0.40; 0.87]	0.63 [0.40; 0.86]	0.44 [0.22; 0.66]	0.39 [0.16; 0.60]
LIST	0.70 [0.47; 0.94]	0.55 [0.26; 0.84]	0.76 [0.50; 1.02]	0.65 [0.34; 0.96]	0.47 [0.24; 0.70]	0.28 [0.00; 0.56]
ASSIST	0.08 [-0.15; 0.31]	-0.09 [-0.39; 0.22]	0.36 [0.12; 0.60]	0.24 [-0.08; 0.57]	0.00 [-0.21; 0.22]	-0.19 [-0.49; 0.10]
PAY	0.29 [0.22; 0.35]	0.29 [0.22; 0.36]	0.26 [0.20; 0.33]	0.28 [0.21; 0.35]	0.38 [0.31; 0.45]	0.37 [0.30; 0.44]
LEAF×PAY	-	-0.02 [-0.11; 0.07]	-	-0.07 [-0.15; 0.02]	-	0.01 [-0.07; 0.76]
LIST×ASSIST	-	0.30 [-0.02; 0.63]	-	0.17 [-0.16; 0.51]	-	0.40 [0.03; 0.76]
2. MRS estimates (in %)						
LEAF	1.15 [0.61; 1.70]	-	1.48 [0.77; 2.19]	-	0.67 [0.16; 1.18]	-
TRAIN	2.05 [1.23; 2.86]	-	2.41 [1.4; 3.42]	-	1.17 [0.58; 1.76]	-
LIST	2.45 [1.53; 3.36]	-	2.89 [1.65; 4.12]	-	1.24 [0.58; 1.90]	-
ASSIST	0.27 [-0.55; 1.01]	-	1.37 [0.40; 2.34]	-	0.01 [-0.56; 0.58]	-
3. Model statistics						
Number of GPs	132	132	128	128	131	131
Number of observations	1440	1440	1423	1423	1448	1448
Log-Likelihood	-911.42	-910.74	-895.49	-894.97	-899.25	-897.9
Pseudo R2	0.087	0.087	0.089	0.090	0.103	0.105

Supplementary file C. Number (%) of 'I don't know' responses across cancer screening contexts

	Breast cancer		**Cervical cancer**		**Colorectal cancer**	
	N	*%*	*N*	*%*	*N*	*%*
Scenario 1	10	7.4	13	9.8	9	6.7
Scenario 2	17	12.6	14	10.5	7	5.2
Scenario 3	21	15.6	19	14.3	16	11.9
Scenario 4	15	11.1	15	11.3	10	7.5
Scenario 5	13	9.6	15	11.3	22	16.4
Scenario 6	13	9.6	11	8.3	15	11.2
Scenario 7	13	9.6	19	14.3	15	11.2
Scenario 8	21	15.6	12	9.0	11	8.2
Scenario 9	14	10.4	14	10.5	16	11.9
Scenario 10	12	8.9	14	10.5	17	12.7
Scenario 11	12	8.9	11	8.3	5	3.7
Scenario 12	5	3.7	2	1.5	7	5.2
Mean	14	10.2	13	10.0	13	9.3