CCHE CCES
Canadian Centre for Health Economics
Centre canadien en économie de la santé

Working Paper Series
Document de travail de la série

HETEROGENEITY IN GENERAL PRACTITIONERS' PREFERENCES FOR QUALITY IMPROVEMENT PROGRAMS: A CHOICE EXPERIMENT AND POLICY SIMULATION IN FRANCE

Mehdi Ammi, Christine Peyron

Working Paper No: 150020

www.canadiancentreforhealtheconomics.ca

December 29, 2015

Canadian Centre for Health Economics
Centre canadien en économie de la santé
155 College Street
Toronto, Ontario

CCHE/CCES Working Paper No. 150020
September 11, 2015

Heterogeneity In General Practitioner's Preferences for Quality Improvement Programs: A Choice Experiment And Policy Simulation in France

Mehdi Ammi[a,*], Christine Peyron[b]

[a] *School of Public Policy and Administration, Carleton University, Ottawa, Ontario, Canada*
[b] *Laboratoire d'Économie de Dijon, UMR 6307 CNRS, U 1200 Inserm and Université de Bourgogne, Dijon, France*

Abstract

Despite increasing popularity, quality improvement programs (QIP) have had modest and variable impacts on enhancing the quality of physician practice. We investigate the heterogeneity of physicians' preferences as a potential explanation of these mixed results in France, where the national voluntary QIP - the CAPI - has been cancelled due to its unpopularity. We rely on a discrete choice experiment to elicit heterogeneity in physicians' preferences for the financial and non-financial components of QIP. Using mixed and latent class logit models, results show that the models should be used in concert to shed light on different aspects of the heterogeneity in preferences. In particular, the mixed logit demonstrates that heterogeneity in preferences is concentrated on the pay-for-performance component of the QIP, while the latent class model shows that physicians can be grouped in four homogenous groups with specific preference patterns. Using policy simulation, we compare the French CAPI with other possible QIPs, and show that the majority of the physician subgroups modelled dislike the CAPI, while favouring a QIP using only non-financial interventions. We underline the importance of modelling preference heterogeneity in designing and implementing QIPs.

JEL Classification: I11; I18; C25

Keywords: quality improvement programs; general practitioners; discrete choice experiment; mixed logit; latent class logit; policy simulation

*Corresponding author. School of Public Policy and Administration, Carleton University, River Building, 1125 Colonel By Drive Ottawa, Ontario, Canada K1S 5B6. Phone: 613-520-2600 x. 3227. Email: mehdi.ammi@carleton.ca

1 INTRODUCTION

Quality improvement programs (QIP) are an increasingly popular approach for enhancing the quality of physician practice in ambulatory care. While heterogeneous, these QIPs systematically target chronic and preventive care, widely used indicators for quality of care (Holmboe et al., 2010). These programs, which seek to change physician practice style, frequently incorporate a financial component, notably pay-for-performance (P4P), and a non-financial one, including clinical guidelines and performance feedback (Cromwell et al., 2011; Gillam et al., 2012; Saint-Lary et al., 2013; Harris et al., 2015).

Despite considerable excitement among policymakers, available evidence suggests that QIPs have modest and variable impacts on quality of care (Mullen et al., 2010; Eijkenaar 2012; Harris et al., 2015). Beyond methodological differences, this observed heterogeneity results from the target and design of the QIPs, as well as from variability in physicians' responsiveness to the programs (Emmert et al., 2012; James 2012; Li et al., 2014; Khoong et al., 2014). Within a single program, differences in physicians' reactions may be explained by differences in contextual constraints, as well as knowledge or attitudes regarding the QIP (Li et al., 2014; Khoong et al., 2014).

Physicians' preferences for QIP are particularly important given that, in many cases, physicians' participation is voluntary and, thus, necessary to ensure the success of the program. From 2009-2011, the French Statutory National Health Insurance implemented a voluntary QIP program (Contract for Improved Individual Practice - CAPI) aimed at general practitioners (GP), which combined P4P and quarterly performance feedback. While the program could only increase their income, only one-third of all French GPs had registered a year and a half after the program's implementation, and the program was subsequently cancelled due to its unpopularity[1]. While GPs' ethical concerns

[1]The CAPI was replaced in 2012 by a P4P program (the ROSP) where physicians are enrolled automatically,

with the program design was one key explanation of the low take-up of the CAPI (Saint-Lary et al., 2013), a QIP better designed to meet physicians' work-related needs may have been more successful.

Health economists have thoroughly studied physicians' preferences regarding their job characteristics (Scott 2001; Rockers et al., 2012), sometimes accounting for preference heterogeneity (Vujicic et al., 2011; Lagarde et al., 2013; Rischatsch and Zweifel 2013). Yet, no studies, to the best of our knowledge, have specifically examined physicians' preferences for QIPs and their components. While recent studies have focused on designs of QIPs that would be effective irrespective of the targeted physicians (Gandjour 2010; Eijkenaar 2012; Kantarevic and Kralj 2013), understanding these physicians' preferences may allow for fine-tuning of the programs and improve acceptance. Moreover, understanding the heterogeneity of physicians' preferences about QIPs may help policymakers tailor and diversify their programs to better match the needs of their targeted population.

This study elicits heterogeneity precisely in physicians' preferences for the components of QIPs. We conduct a discrete choice experiment (DCE) on a sample of French GPs and use mixed and latent class logit models to thoroughly capture preference heterogeneity. With these elements in hand, we estimate by simulation the potential and differential impact on physician welfare of several QIPs, including the French CAPI.

2 DATA AND THE DISCRETE CHOICE EXPERIMENT

2.1 DCE Design

Discrete choice experiments are widely used in the health economics literature to assess preferences (De Bekker-Grop et al., 2012). Our study followed the recommended steps (Amaya-Amaya et al.,

but can request to opt out.

2008) as described below.

The first step of a DCE is to select the attributes of interest and their levels. We selected attributes based on a literature review on QIPs and on two criteria: supposed efficacy suggested by the literature and credibility of application in the French health care context (see Table I). For concreteness, we focused on preventive care, a key quality indicator. Following the same two above criteria, a level for each attribute was defined to reflect the CAPI, which notably also focused on preventive care. The relevance of the list of attributes and of their levels was confirmed in a focus group of ten representative GPs[2]. This led to a final list of eight attributes presented in Table II.

Table I & II about here

The second step is to combine attributes into choice sets. Most of time, the combination relies on experimental plan theory since a full factorial design implies proposing too many choices to respondents (Rose and Bliemer 2008) - 864 scenarios in our case. Using an orthogonal design (Hensher et al., 2005) that resulted in 24 scenarios[3], we achieved the properties of orthogonality and level balance. In order to facilitate respondents' choices, we relied on a common comparator selected from these 24 scenarios, ensuring that this reference scenario is not strictly dominant *a priori* (Scott 2002). Choice sets were constructed by pairs which resulted in 23 choices between pairs of combinations of quality interventions. The 23 choice sets were randomly divided into four blocks so that each respondent made 5 or 6 choices[4]. To limit the risk of reduction of the sample and the subsequent loss of statistical efficiency due to non-response, we did not include an opt-out possibility. An example of choice set is provided in appendix A.

[2]The number of considered attributes should not be so high as to allow respondent to make trade-offs. If there are no clear recommendations on the maximum number, the DCE health economics literature generally uses at most eight attributes (Kjaer 2005).

[3]The experimental plan was run with the JMP7 software.

[4]Caussade et al. (2005) showed that setting between 6 and 13 choice situations minimises the error variance of the estimates.

Finally, the DCE was pilot tested with a focus group of self-employed GPs to ensure the exercise was understandable by physicians, and certain levels of attributes were reformulated as a result. A pre-test on a subsample of 100 GPs was conducted to verify that the reference scenario was not strictly dominant as *a priori* expected. This was confirmed and the questionnaire was kept in its pre-test form.

2.2 Data

The DCE questionnaire is composed of three parts. In the first part, questions regarding the GP's opinion about health care reforms in general practice and the public health role of GPs are used as a warm-up. The second part is the choice experiment. The third part collects sociodemographic and professional information about each GP. The questionnaire is self-administered during the summer of 2009 in a postal survey with one repeated attempt for non-response.

The population under study consists of all the GPs in active practice in one French geographic region[5] (N=1,368). After the pre-test, the questionnaires were sent to the 1,268 remaining physicians. 303 questionnaires were returned completed, resulting in a response rate of 22%. This response rate is consistent with other DCE studies (Gerard et al., 2003; Mentzakis et al., 2011, Scott et al., 2013) and with self-administered postal surveys to French general practitioners (Le Fur et al., 2009).

Descriptive statistics are presented in Table III with regional and national values. The comparison of the different values shows that GPs working in a rural setting are slightly overrepresented in our sample. The responding GPs are also more active, with the weekly number of acts being significantly

[5]We restricted ourselves to the region of Bourgogne because of prior relationships with the regional health professional's union who facilitated the constitution of the focus-group and offered logistic support for the survey. The restriction to one region is also due to monetary and time constraints. The fund obtained from the *Conseil Regional de Bourgogne* did not allow for a survey of more than one region.

higher than the national mean[6]. With these exceptions, our sample compares well with the reference population. Of course, our methodology does not allow for national representativeness.

Table III about here

With the exception of the level of remuneration, all attributes of the DCE are coded using "effects coding" (Bech and Gyrd-Hansen 2005). We constructed the questionnaire in order to test the symmetry (Kjaer et al., 2006), the completeness and the continuity axioms (Ryan et al., 2009)[7] and found that the axioms are largely respected: totally for the first, and respectively by 82% and 65% of the respondents for the two other axioms. Following current practice, we kept all the responses for the analysis (San Miguel et al., 2005; Lancsar and Louviere 2006; Ryan et al., 2009).

3 ECONOMETRIC FRAMEWORK

3.1 Modelling heterogeneity

The analysis of DCE data relies on classical choice models and random utility theory (RUM) (McFadden 1974). When applying the DCE approach, the utility of an individual n choosing alternative i at the t choice situation can be written as

$$U_{nit} = V_{nit} + \epsilon_{nit}$$

Where $V_{nit} = \sum_{k=1}^{K} \beta_k x'_{nitk}$ is the deterministic part of the utility (with k attributes), observable to the researcher and sometimes referred to as the indirect utility, and ϵ_{nit} is the unobservable,

[6]This point is, however, not particularly concerning as the regional and national values are derived from an administrative database (*système national d'information inter-régimes* - SNIIR) known to underestimate physicians' activity. The SNIIR includes the very low activity physicians, pulling down the average number of acts.

[7]Beyond the choice exercise, supplementary choices and follow-up questions were introduced in the DCE in order to test the internal validity of the data collected. More information on the test procedures used is available upon request.

stochastic part and is treated as random[8]. The individual will choose the alternative yielding the highest utility. Assuming that the stochastic part is distributed independently and identically (IID), extreme values give rise to the McFadden (1974) conditional logit (CL). The probability of an individual n choosing i among J alternatives is then

$$P_{ni} = \frac{exp(V_{ni})}{\sum_{j=1}^{J} exp(V_{nj})}$$

The CL is the most commonly used method to analyse DCE data, but relies on restrictive assumptions on the stochastic terms (Hensher et al., 2005), fails to incorporate the panel structure of most DCE data and does not account for preference heterogeneity. The two principal models that circumvent these limitations are the mixed logit (MXL) (McFadden and Train 2000; Hensher and Greene 2003) and the latent class model (LCM) (Greene and Hensher 2003).

The choice between these two models critically depends on expectations about the variation of preferences (Hole 2008): if we expect preferences to vary greatly between individuals, the MXL is preferred; the LCM is preferred if we suspect individuals to be grouped in homogeneous latent groups. Rather than choosing one model before estimation, and as the selection of the specification is an empirical question overall, we run both MXL and LCM. By doing this, we avoid errors in model selection stemming from the researcher's belief about the distribution of preferences. Moreover, the information the models provide is complementary. MXL provides information about how heterogeneity is distributed relative to each attribute while LCM informs on the heterogeneity among subgroups of physicians (in latent groups).

The unconditional probability of a mixed model that allows for individual-specific variation in tastes

[8]This random part is precisely why RUM can deal with axiomatic violations. The "errors" may come from this stochastic part, which is unexplained by the researcher.

and accounts for the panel dimension of choices is as follows (Train 2009):

$$P_{nI}(\theta) = \int S_{nI} f(\beta|\theta) d\beta$$

Where $S_{nI}(\beta) = \prod_{t=1}^{T} \frac{exp(\beta' x_{nit})}{\sum_{j=1}^{J} exp(\beta' x_{nit})}$ is the conditional probability that the individual n realises a choice sequence $I = \{i_1, ..., i_t\}$, $f(\beta|\theta)$ is a density function of the individual-specific β with distribution parameters θ (see Train (2009) for more on the family of mixed models).

Preference heterogeneity is reflected in the density function, $f(\beta|\theta)$, and its choice is therefore critical. The distribution of β can be either continuous or discrete, implying MXL or LCM, respectively. This highlights the similarity of these two models, while at the same time their major difference. Train (2009) points out that LCM may be considered a specific case of MXL, where the distribution function is degenerated at specific points.

The other major difference between the models is the estimation method. Each model relies on log-likelihood maximization, with the log-likelihood given by $LL(\theta) = \sum_{n=1}^{N} lnP_n(\theta)$. Unlike the LCM, this expression cannot be solved analytically in MXL and simulation methods are used for approximation (Greene and Hensher 2003; Train 2009).

3.2 Simulating policy

The goal of the policy simulation is to evaluate the effects of changes in the three main components of a QIP (financial, non-financial and organizational), and we use the compensating variation (CV) method to measure the relative impact on GPs' welfare of such change (Lancsar et al., 2007; Ryan et al., 2008). This supposes that a QIP exists before the change, which is the case with the CAPI.

The CV is calculated using the utility estimates computed after the regressions in the following

expression (Lancsar et al., 2007)

$$CV = -\frac{1}{\beta_w}\left[ln\sum_{j=1}^{J} exp(V_j^0) - ln\sum_{j=1}^{J} exp(V_j^1)\right]$$

Where β_w is the marginal utility of income, V_j^0 is the indirect utility for each option j before the policy change and V_j^1 the same after the policy change. In our case, we consider only two policy options at a time, the CAPI versus something else. The formula is then simplified to (Amaya-Amaya et al., 2008)

$$CV = -\frac{1}{\beta_w}[V_j^0 - V_j^1]$$

The question of heterogeneity is evaluated by estimating CV for each latent group of physicians with LCM. While the MXL model does not convey information on subcategories of GPs, it does indicate if there is substantial heterogeneity of preferences for different attributes. It will be relevant to compute and compare CV for GPs differing significantly in respect to the specific attributes of the QIP (e.g. those obtaining positive versus negative marginal utility from the attribute).

3.3 Model specification

We include an intercept in all models. This alternative-specific constant (ASC) is necessary since choices are made relative to a fixed comparator (the constant scenario) (Bech and Gyrd-Hansen 2005; Ryan et al., 2008). In our case, this ASC has no natural interpretation since it would indicate a preference for the common comparator over the other alternative net of the influence of the attributes. The ASC is expected to be statistically insignificant and can be used to test for specification error (Scott 2001).

When specifying a mixed logit it is critical to choose which parameters are allowed to vary and which distribution these latter will follow. The normal and log-normal distributions are the most commonly used for the random coefficients (Kjaer and Gyrd-Hansen 2008; Hole 2008; Train 2009). The log-normal distribution is however criticised for its long right tail, which may cause unrealistic estimates (Hensher and Greene 2003; Sillano and Ortïé¡zar 2005). We thus choose the normal distribution[9].

The possibility to specify the coefficients as random is one of the great strengths of the MXL. The ASC is fixed since it has no reason to vary between the respondents. Fixing the monetary attribute (the remuneration) has several advantages (Revelt and Train 1998). In our case, the main one is the capacity to calculate CV. The possibility of significant preference heterogeneity in terms of remuneration cannot be ruled out and should be considered in order to fully understand physicians' preferences. GPs valuing less payment can indeed be explained in an intrinsic motivation framework, among others. We therefore run two MXL: one with all coefficients normally distributed except the constant and the amount of remuneration coefficient (MN1) and the other with only the constant term fixed (MN2).

An advantage of the LCM over the MXL is that the choice of the random parameters and their distribution is not an issue, but difficulty still remains in choosing the number of latent classes. Without an intuitive way to choose the number of classes, the decision is often made on the basis of goodness-of-fit measures (Hole 2008; Mentzakis et al., 2011). The selection of the number of classes is made on the basis of the Akaike (AIC), Bayesian (BIC) and consistent Akaike (CAIC) information criteria.

[9]Because of the qualitative nature of the majority of our attributes, there is no reason to think that one level should be preferred to another *a priori*. It is therefore difficult to select the sign of the distribution. MXL with log-normal distribution are run for sensitivity analyses and do not exhibit large differences in the fit. Results are available from the authors.

The results for the selection of the number of classes are presented in Table IV. The BIC and CAIC show that the best fit is obtained with four classes, a number we retain for the following analyses[10].

Table IV about here

4 RESULTS

4.1 Heterogeneity in GPs' preferences

The estimation results for the mixed logit are presented in Table V, where model MN1 has all coefficients normally distributed except the constant term and the amount of remuneration coefficient and model MN2 also has the remuneration term normally distributed.

Table V about here

The sign, significance and magnitude of the mean coefficients are very stable between the two models, underlining the robustness of the results. The ASC is not significant, indicating that respondents have made their choice only on the basis of the attributes in the list (so the model is correctly specified). The estimates reveal the existence of preference heterogeneity among GPs that is quite concentrated around some attributes.

The standard deviations are significant for the pay-for-performance and the assistance by NPP in model MN1. In MN2, this is also the case for the application of guidelines, the type of practice, and the level of remuneration. The heterogeneity in preferences for pay-for-performance is particularly

[10]The simplest way to account for heterogeneity of preferences is to incorporate the personal characteristics of the respondents in the models. It can be done with interaction terms in the MXL and to explain class membership probability in the LCM. We argue these individual characteristics have to considerably improve the fit of the models in order to be worth keeping for final analysis, which is not the case with our data (results available upon request). Most of the personal characteristics are found to be insignificant in the MXL, the integration of these characteristics does not drastically improve the fit to the data in MXL, and even worsens the fit of the LCM. As a result and following Hole (2008), we focus only on the more parsimonious models in our analyses.

relevant. This remuneration scheme is a source of marginal disutility at the mean but is positively valued by 22% and 24% of physicians (in MN1 and MN2, respectively). These figures are consistent with the proportion of French GPs having chosen to adhere to the CAPI (around 30%, Saint-Lary et al., 2013). It is also worth noting that the indifference to the assistance by NPP at the mean masked a strong heterogeneity. Indeed, 60 to 62% would like to benefit from this kind of assistance. Finally, even the amount of remuneration is marked by heterogeneity, with 14% of physicians not valuing an increase in income for the targeted activities (MN2).

Table VI about here

The latent class model estimates are presented in Table VI. Over all the classes, the ASCs are insignificant. For the first class, the only significant attributes are continuing education and assistance by NPP. Continuing education has a positive effect on indirect utility while assistance by NPP has a negative one. In the second class, the significance of the attributes is slightly different. While continuing education remains significant, this time it has a negative effect. GPs in this class prefer higher payment and to be paid more often, as the sign and significance of the frequency attribute attests. They dislike the *forfait* but they are indifferent to pay-for-performance. They also prefer solo practice. All attributes are significant for classes 3 and 4, however distinct behaviour is observed. The doctors in these two latent classes place negative value on alternative payment relative to FFS while preferring more frequent payment. They also prefer to work in groups. They differ in respect to all the other attributes. In contrast to the third class, an increase in remuneration has a negative effect on indirect utility in the fourth class. Class 3 physicians disvalue all types of clinical guidelines but positively value continuing education and information feedback, contrary to class 4. Physicians in the fourth class value assistance by NPP while those in the third class do not. With the preference for group practice in both classes, this result suggests a preference for

physician groups only in class 3 while multidisciplinary teams are preferred in class 4.

At this point it is worth comparing the results of the two kinds of models. One of the major conclusions, holding in both MXL and LCM, is the negative impact on indirect utility of an increase in remuneration observed for some GPs. It shows that this result is not only a matter of statistical artefact resulting from the use of a normal distribution in the MXL (Hole 2008). The MXL underlined heterogeneity of preferences for P4P. This heterogeneity is also found in the LCM, with the third and fourth classes disliking this payment while the coefficient is positive in the second class (but significant only at 10%). The strong difference in preferences for assistance by NPP found in MXL is also seen in LCM. The negative coefficients in classes 1 and 3 are contrasted by a strong positive preference in class 4. All in all, this suggests a stability of the main conclusions made from the different models, with preference heterogeneity remaining among classes.

Finally, we compare the goodness of fit of the models using the log-likelihood, the Akaike and Schwarz information criteria in order to see how the heterogeneity of preferences is distributed among GPs. Results in Table VII indicate very little advantage to LCM while MXL (MN2) has better BIC. The minimal difference between the best fitting models suggests that each provides relevant information on the heterogeneity of GPs' preferences.

Table VII about here

4.2 Simulating alternative quality improvement programs

The policy simulation study relies on the calculation of compensating variation. The goal is here to evaluate the relative impact on physicians' welfare of alternative QIPs to the CAPI. These alternatives were chosen to be consistent with, and believable in, the context of French general

practice.

The DCE attributes are used to depict five QIPs - the CAPI and four alternative policies[11]. The first is close to the emerging organizational model in French primary care (*maisons pluridisciplinaires et pôles de santé*) implemented to foster quality of care, and also known in the literature as "integrated" primary care model (Romanow 2002). The second introduces a mixed remuneration scheme that can better balance quantity and quality in physicians' activity (Dumont et al., 2008). In order to measure only the effect of the payment scheme, we assume an increase in income similar to the CAPI. The third QIP is composed of only non-financial mechanisms that do not require a sharp transformation in physicians' organization (i.e. no multidisciplinary team). The fourth is designed as a maximal satisfaction policy and is used as a benchmark[12]. Even if the maximum satisfaction of GPs is not necessarily an objective *per se*, comparing it to the CAPI gives a sense of the distance separating this QIP from the most desirable one. The details of each policy are presented in Table VIII.

Table VIII about here

The indirect utilities and the corresponding CV are first computed for all GPs on the basis of MN1 estimates. With mixed logit models, we concentrate on the attributes which are consistently heterogeneous in the two models (MN1 and MN2): P4P and assistance by NPP. For each, we identify "inclined" who obtain positive marginal utility from these attributes and "adverse" who obtain negative marginal utility. The LCM provides natural subgroups for the estimation of CV, which are computed in the four latent classes. It should be noted that only the significant coefficients enter in the computation of CV for each subgroup of interest. As GPs are indifferent to insignificant

[11] For more details, the reader may refer to appendix B.

[12] The maximal satisfaction policy is designed for all GPs. Our goal is to compare different nationally uniform policies such as the CAPI.

attributes, using their estimate values would distort the welfare estimates. Results are presented in Table IX.

Table IX about here

Even if there is some heterogeneity, the first striking result is that CAPI is a source of indirect disutility in the majority of the subgroups considered (5 out of 8).

The compensating variation indicates the annual benefits for GPs of choosing an alternative QIP rather than the CAPI. P4P "inclined" have a positive indirect utility from the CAPI of course. However, with the exception of the mixed remuneration program, all other alternative policies still give a greater benefit than the CAPI[13]. P4P "adverse" would prefer each of the alternative policies to the CAPI, if they were proposed. The non-financial policy has the greatest CV, but the gap with integrated primary care is reduced. Whether they are "inclined" or "adverse" to assistance by NPP, GPs disvalue the CAPI and prefer all alternatives. We expected the NPP "inclined" to have a greater benefit from P1 because of the multidisciplinary team but P3 is more valued, though the difference is relatively small.

The NPP "adverse" have their lowest (though still positive) CV for P1 and their preferred alternative is the non-financial program P3.

The patterns are very different between latent classes. Classes 1 and 4 obtain negative and extremely negative indirect utility from the CAPI, respectively, while the sign is positive in classes 2 and 3. Compared to the other subgroups, CV is very high in class 1[14]. The benefit of having the non-financial policy rather than the CAPI is equivalent to 93,705€, almost the same amount as for

[13] In this paragraph, when talking about greater CV, we do not take the maximum satisfaction into account and only concentrate on the reliably implantable policies.

[14] This is partly because only two attributes are found significant for this latent class and enter the calculation of the CV. Their presence or absence therefore has a disproportionate impact on the CV for each policy.

the maximum satisfaction program. There is no benefit from shifting from the CAPI to the mixed remuneration scheme. This last result holds for class 2. This class is very specific since it is the only subgroup where other policies result in losses. It is even the case for P4, designed to be the most desirable for GPs in the whole, underlining again the particularity of this latent group. For class 3, mixed remuneration has the highest CV, with a relative benefit of 18,474€. With the exception of P1, alternative policies still dominate the CAPI. For class 4, integrated primary care offers the highest relative benefit (53,925€) while the CV for the non-financial policy remains important (47,148€).

5 CONCLUSION AND DISCUSSION

Using a discrete choice experiment, we elicited French GPs preferences for the different components of QIPs. We showed the strength of heterogeneity in their preferences and demonstrated how this heterogeneity leads physicians to evaluate very differently the same interventions aimed at improving the quality of care. The heterogeneity in preferences is concentrated on some components, especially P4P and assistance by a NPP. There is also variation in preferences by latent groups of GPs, with some physicians valuing some components of QIP (continuing education and assistance by NPP in group 1), while other physicians value the same components differently (group 3 versus 4). Given this heterogeneity, the crucial policy lesson is that QIPs could be adapted to meet physicians' preferences by offering a menu of programs and allowing GPs to self-select. If policymakers were to choose only one QIP, CV indicates that they should implement a program using only non-financial interventions.

Some limitations should be noted. First, the study population is not representative of French GPs,

but compares well with the region population. Moreover, most of GPs' personal characteristics were not significant when introduced in the models, especially the case of rurality - a key difference between the study region and the national level. Second, the limited response rate, though consistent with the DCE literature, may have led to selection bias in the sample. While we do not have information on the non-responders, the opinions expressed in the first part of the questionnaire are reassuring in the sense that they are quite close to those expressed in other French studies (Buttet and Fournier 2003; Levasseur et al., 2004; Aulagnier et al., 2007b). Third, the use of a forced choice design might have biased the estimates if physicians wished to choose neither of the two proposed QIP. However, physicians who were not willing to choose one of the two options in a given choice set actually did not respond at the specific choice occasion, the forced choice is still used in health professional DCE studies (Lagarde et al., 2013), and this "forced choice" strategy is consistent with the new orientation of the French national QIP program (the ROSP is mandatory). Finally, we choose to use a common comparator when we constructed the choice set, which does not necessarily maximize the statistical efficiency of the experimental design (Rose and Bliemer 2008). Yet, fixed comparator facilitates the heuristic decision making process and help increases the "respondent efficiency", which can be defined as the capacity of a respondent to express his preferences in the context of the DCE as closely as possible to his "real" preferences (Louviere 2001). Given that private practice physicians are heavily time-constrained, particularly in the French fee-for-service context, we believe this trade-off between statistical and respondent efficiency has allowed us to obtain a satisfactory response rate and better quality and completeness of responses relative to other designs.

Despite these limitations, this study adds to the broader literature on the heterogeneity of health professionals' preferences (Vujicic et al., 2011; Rockers et al., 2012; Lagarde et al., 2013; Scott et

al., 2013) and for the first time, combines LCM and MXL approaches. Each model contributes a better understanding of physicians' preferences and using such an approach can help policymakers to better design their QIP.

Acknowledgments

We are grateful for comments from Sophie Bïé¡jean and Erin Strumpf on an earlier version of this paper. We thank Arne Risa Hole for his advices on his *mixlogit* Stata command and we thank the *Union régionale des professionnels de santé* for their help with the focus-group meetings and their logistic support for the survey. We are of course in great debt to the GPs who took time to answer the survey. Financial support is acknowledged from the Regional Council of Bourgogne (CPER 2009 and PARI 2010). Usual disclaimers apply.

References

Amaya-Amaya M, Gerard K, Ryan M. 2008. Discrete Choice Experiments in a Nutshell. in Ryan M, Gerard K, Amaya-Amaya M (Eds). *Using Discrete Choice Experiments to Value Health and Health Care*. Springer.

Aulagnier M, Obadia Y, Paraponaris A, Saliba-Serre B, Ventelou B, Verger P, Guillaumat-Tailliet F. 2007a. L'exercice de la médecine générale libérale. Premiers résultats d'un panel dans cinq régions françaises. *Etudes et résultats* 610.

Aulagnier M, Videau Y, Combes J-B, Sebbah R, Paraponaris A, Verger P, Ventelou B. 2007b. Pratiques des médecins généralistes en matière de prévention : les enseignements d'un panel de médecins généralistes en Provence-Alpes-Côte d'Azur. *Pratiques et Organisation des Soins* 38: 259-268.

Bras P-L, Duhamel G. 2008. *Rémunérer les médecins selon leurs performances : les enseignements des expériences étrangères*. Inspection Générale des Affaires Sociales. Paris.

Bournot M-C, Goupil M-C, Truffeau F. 2008. Les médecins généralistes : un réseau professionnel étendu et varié. *Etudes et résultats* 649.

Buttet P, Fournier C. 2003. Prévention et éducation pour la santé. Quels rôles pour les médecins et les pharmaciens? in: Gautier A (Ed) *Baromètre santé médecins/pharmaciens* 2003. INPES: Saint-Denis.

Caussade S, Ortúzar J de D, Rizzi L I, Hensher D A. 2005. Assessing the influence of design dimensions on stated choice experiment estimates. *Transportation research part B: Methodological* 39(7): 621-640.

Chung S, Palaniappan L, Wong E, Rubin H, Luft H. 2010. Does the Frequency of Pay-for-Performance Payment Matter? - Experience from a Randomized Trial. *Health Services Research* 45(2): 553-564.

Cromwell J, Trisolini M G, Pope G C, Mitchell J B, Greenwald L M. (Eds.) 2011. *Pay for Performance in Health Care: Methods and Approaches*. RTI Press publication.

De Bekker-Grob EW, Ryan M, Gerard K. 2012. Discrete Choice Experiments in Health Economics: A Review of the Literature. *Health Economics* 21(2): 145-172.

Dexheimer JW, Talbot TR, Sanders DL, Rosenbloom ST, Aronsky D. 2008. Prompting Clinicians about Preventive Care Measures: A Systematic Review of Randomized Controlled Trials. *Journal of the American Medical Informatics Association* 15(3): 311-320.

Dumont E, Fortin B, Jacquemet N, Shearer B. 2008. Physicians' Multitasking and Incentives: Empirical Evidence from a Natural Experiment. *Journal of Health Economics* 27(6): 1436-1450.

Eco-Santé France, Régions & Départements. 2015. `http://www.ecosante.fr/`. [Last access April 28, 2015]

Eijkenaar E. 2012. Pay for performance in Health Care: an international overview of initiatives. *Medical Care Research and Review* 69(3): 251-276.

Emmert M, Eijkenaar E, Kemter H; Esslinger AS, Schöffski O. 2012. Economic evaluation of pay for performance in health care: a systematic review. *European Journal of Health Economics* 13(6): 755-767.

Farmer A, Légaré F, Turcot L, Grimshaw J, Harvey E, McGowan J, Wolf F. 2008. Printed educational materials: effects on professional practice and health care outcomes. *Cochrane Database of*

Systematic Reviews Jul 16;(3):CD004398.

Forsetlund L, Bjørndal A, Rashidian A, Jamtvedt G, O'Brien MA, Wolf F, Davis D, Odgaard-Jensen J, Oxman AD. 2009. Continuing education meetings and workshops: effects on professional practice and health care outcomes. *Cochrane Database of Systematic Reviews* Apr 15;(2):CD003030.

Fréchou H, Guillaumat-Tailliet F. 2008. Les revenus libéraux des médecins en 2005 et 2006. *Etudes et résultats* 643.

Gandjour A. 2010. Investment in quality improvement: how to maximise the return. *Health Economics* 19(1): 31-42.

Gerard K, Shanahan M, Louviere J. 2003. Using stated preference discrete choice modelling to inform health care decision-making: a pilot study of breast screening participation. *Applied Economics* 35(9): 1073-1085.

Gerard K, Shanahan M, Louviere J. 2008. Using discrete choice modelling to investigate breast screening participation. in: Ryan M, Gerard K, Amaya-Amaya M (Eds) *Using Discrete Choice Experiments to Value Health and Health Care*. Springer.

Gillam S, Niroshan Siriwardena A, Steel N. 2012. Pay-for-Performance in the United Kingdom: Impact of the Quality and Outcomes Framework - A Systematic Review. *Annals of Family Medicine* (10)5: 461-468.

Greene WH, Hensher DA. 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological* 37(8): 681-698.

Grol R. 2001. Successes and Failures in the Implementation of Evidence-Based Guidelines for Clinical Practice. *Medical Care* 39(8 Suppl 2): 46-54.

Harris, S.B., Green, M.E., Brown, J.B., Roberts, S., Russell, G., Fournie, M., Webster-Bogaert, S., Paquette-Warren, J., Kotecha, J., Han, H., others 2015. Impact of a quality improvement program on primary healthcare in Canada: A mixed-method evaluation. *Health Policy* 119(4): 405-416

Hensher DA, Greene WH. 2003. The Mixed Logit model: The state of practice. *Transportation* 30: 133-176.

Hensher DA, Rose JM, Greene WH. 2005. *Applied choice analysis: a primers.* Cambridge University Press: Cambridge.

Hole AR. 2008. Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment. *Journal of Health Economics* 27(4): 1078-1094.

Holmboe E, Weng W, Arnold G,Kaplan S, Normand S, Greenfield S, Hood S, Lipner R. 2010. The comprehensive care project: measuring physician performance in ambulatory practice. Health Services Research 45(6 Part II):1912-1933.

James J. 2012. Health policy brief: Pay-for-performance. *Health Affairs* october 11.

Kantarevic J, Kralj B. 2013. Link between pay for performance incentives and physician payment mechanisms: evidence from the diabetes management incentive in Ontario. *Health Economics* 22(12): 1417-1439.

KjærT. 2005. A review of the discrete choice experiment - with emphasis on its application in health care. *Health Economics Papers* University of Southern Denmark 2005/1; 139.

Kjær T, Bech M, Gyrd-Hansen D, Hart-Hansen K. 2006. Ordering effect and price sensitivity in discrete choice experiments: need we worry? *Health Economics* 15(11): 1217-1228.

Kjær T, Gyrd-Hansen D. 2008. Preference heterogeneity and choice of cardiac rehabilitation pro-

gram: Results from a discrete choice experiment. *Health Policy* 85(1): 124-132.

Khoong E, Gibbert W, Garbutt J, Sumner W, Brownson R. 2014. Rural, Suburban, and Urban Differences in Factors That Impact Physician Adherence to Clinical Preventive Service Guidelines. *The Journal of Rural Health* 30(1):7-16.

Lagarde M, Pagaiya N, Tangcharoensathian V, Blaauw D. 2013. One size does not fit all: investigating doctors' stated preference heterogeneity for job incentives to inform policy in Thailand. *Health Economics*. 22(12):1452-69.

Lancsar E, Louviere J. 2006. Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? *Health Economics* 15(8): 797-811.

Lancsar E, Louviere J, Flynn T. 2007. Several methods to investigate relative attribute impact in stated preference experiments. *Social Science & Medicine* 64(8): 1738-1753.

Le Fur P, Bourgueil Y, Cases C. 2009. Le temps de travail des médecins généralistes. Une synthése des données disponibles. *Questions d'économie de la santé* 144.

Levasseur G, Bataillon R, Samzun J. 2004. *Le schéma régional d'organisation sanitaire. Baromètre des pratiques en médecine générale. Panel Breton.* Rapport d'enquête de l'URML Bretagne.

Li J, Hurley J, Decicca P, Buckley G. 2014. Physician response to pay-for-performance: Evidence from a natural experiment. *Health Economics* 23(8): 962-978.

López-de-Munain J, Torcal J, López V, Garay J. 2001. Prevention in Routine General Practice: Activity Patterns and Potential Promoting Factors. *Preventive Medicine* 32(1): 13-22.

Louviere J. 2001. Choice experiments: an overview of concepts and issues. in: Bennett J, Blamey R (Eds), *The Choice Modelling Approach to Environmental Valuation.* Edward Elgar: Northhampton.

McFadden D, Train K. 2000. Mixed MNL models for discrete response. *Journal of Applied Econometrics* 15(5): 447-470.

Mentzakis E, Ryan M, McNamee P. 2011. Using discrete choice experiments to value informal care tasks: exploring preference heterogeneity. *Health Economics* 20(8): 930-944.

Mousquès J, Bourgueil Y, Le Fur P, Yilmaz E. 2010. Effect of a French experiment of team work between general practitioners and nurses on efficacy and cost of type 2 Diabetes Patients Care. *Health policy* 98(2-3): 131-143.

Mullen K, Franck R, Rosenthal M. 2010. Can you get what you pay for? Pay-for-performance and the quality of healthcare providers. *Rand Journal of Economics*. 41(1): 64-91.

ORS Bourgogne. 2010. Soins et prévention - Généralistes. *Faits marquants*. Fiche 5.7.

Pham H, Schrag D, Hargraves J, Bach P. 2005. Delivery of preventive services to older adults by primary care physicians. *The Journal of the American Medical Association* 294(4): 473-481.

Revelt D, Train K. 1998. Mixed logit with repeated choices: households' choices of appliance efficiency level. *Review of Economics and Statistics* 80(4): 647-657.

Rischatsch M, Zweifel P. 2013. What do physicians dislike about managed care? Evidence from a choice experiment. *European Journal of Health Economics* 14(4): 601-613.

Rockers PC, Jaskiewicz W, Wurts L, Kruk ME, Mgomella GS, Ntalazi F, Tulenko K. 2012. Preferences for working in rural clinics among trainee health professionals in Uganda: a discrete choice experiment. *BMC Health Services Research* 12:212.

Rose J, Bliemer M. 2008. Stated preference experimental design strategies. in Hensher DA, Button KJ (Eds), *Handbook of Transport Modelling*. Elsevier Oxford.

Romanow RJ. 2002. *Building on Values: The Future of Health Care in Canada.* Commission of the Future of Health Care in Canada.

Ryan M, Gerard K, Amaya-Amaya M. 2008. *Using Discrete Choice Experiments to Value Health and Health Care.* Springer.

Ryan M, Watson V, Entwistle V. 2009. Rationalising the 'irrational': a think aloud study of discrete choice experiment responses. *Health Economics* 18(3): 321-336.

Saint-Lary O, Bernard E, Sicsic J, Plu I, François-Purcell I, Frans C. 2013. Why Did Most French GPs Choose Not to Join the Voluntary National Pay-for-Performance Program? *PLoS ONE* 8(9):e72684.

San Miguel F, Ryan M, Amaya-Amaya M. 2005. 'Irrational' stated preferences: a quantitative and qualitative investigation. *Health Economics* 14(3): 307-322.

Scott A. 2001. Eliciting GPs' preferences for pecuniary and non-pecuniary job characteristics. *Journal of Health Economics* 20(3): 329-347.

Scott A. 2002. Identifying and analysing dominant preferences in discrete choice experiments: An application in health care. *Journal of Economic Psychology* 23(3): 383-398.

Scott A, Witt J, Humphreys J, Joyce C, Kalb G, Jeon S, McGrail M. 2013. Getting doctors into the bush: General Practitioners' preferences for rural location. *Social Science & Medicine* 96: 33-44.

Sicart D. 2009. *Les médecins. Estimations au 1er janvier 2009.* Drees Série statistiques Document de travail 138.

Sillano M, Ortuzar J. 2005. Willingness-to-pay estimation with mixed logit models: some new evidence. *Environment and Planning A* 37(3): 525-550.

Town RJ, Kane RL, Johnson PE, Butler M. 2005. Economic Incentives and Physician's Delivery of Preventive Care. A Systematic Review. *American Journal of Preventive Medicine* 28(2): 234-240.

Train K. 2009. *Discrete choice methods with simulations.* Cambridge University Press: Cambridge.

Vujicic M, Shengelia B, Alfano M, Thu HB. 2011. Physician shortages in rural Vietnam: using a labor market approach to inform policy. *Social Science & Medicine* 73(7): 970-977.

Tables

Table I: Interventions used in quality improvement programs for GPs

Component of the QIP	Justification
Financial component	
Amount of payment	The literature suggests a threshold of 5% of the doctors' income as a minimum for the incentive to be effective (Bras and Duhamel 2008).
Method of remuneration	Financial incentives can improve the quality of care, but depend on the method and frequency of payment (Town et al., 2005; Eijkenaar 2012)[15]. The three remuneration methods used in France are pay-for-performance (P4P), fee-for-service (FFS) and a kind of partial capitation known as a *forfait*[16].
Non-financial component	
Clinical guidelines	The efficacy of clinical guidelines is ascertained (Farmer et al., 2008). However, the kind of guideline used matters, and guidelines to which individual clinicians have contributed may be more effective in changing their behavior (Grol 2001).
Feedback on activity	Performance feedback, where physicians get quantitative feedback relate to their practice, increases quality of care (Dexheimer et al., 2008).
Continuing education	Participation in continuing education increases adherence to clinical recommendations (Forsetlund et al., 2009).
Organisational component	
Type of practice	There is an association between group practice and better quality of care (Lïé¡pez-de-Munain et al., 2001; Pham et al., 2005).
Non-physician provider	Quality of care is improved by cooperation of GPs with non-physician providers such as nurses (Mousquïé¡s et al., 2010).

[15] This point is subject to debate. A recent study finds no effect of the frequency of P4P (Chung et al., 2010). However, representative GPs in the focus group cited the importance of this attribute.

[16]The French *forfaits* are a partial capitation payment that represents a small part of GPs income (6% of income (Fréchou and Guillaumat-Tailliet 2008)) for certain patients (chronically ill) or for the coordination and continuity of care. They complement the FFS but are absolutely not designed as a major payment. For example, the GP receives 40 euros a year for following each patient classified by the health insurance plan as chronically ill (*forfait pour affection de longue durée*). In comparison, sector 1 GPs are paid 23 euros for each consultation at the physician's office.

Table II: List of attributes and levels

Attributes	Levels
Level of remuneration[17] (annual increase)	100 Euros
	6,100 Euros
	12,100 Euros
Method of remuneration	Lump sum (*forfait*[18])
	Lump sum and fee-for-service
	Lump sum and pay-for-performance
Frequency of remuneration	Monthly
	Annually
Prevention clinical guidelines	None
	Participatory guidelines (participation in their definition and application)
	Pre-established guidelines (evidence-based application)
Feedback on preventive practices	Yes
	No
Continuing education in prevention	Yes
	No
Type of practice	Group of GPs
	Solo practice
Assistance by non-physician providers during preventive work	Yes
	No

[17]We retain three levels: 0, 5 and 10%. It was not possible to propose a truly null amount, so an amount very close to zero was proposed. French physicians are not accustomed to thinking about their income in percentage terms, thus the payment attribute was proposed in raw of the average income (in euros) rather in relative terms (in percentage).

[18]The French *forfaits* are a partial capitation payment that represents a small part of GPs income (6% of income (Frïé¡chou and Guillaumat-Tailliet 2008)) for certain patients (chronically ill) or for the coordination and continuity of care. They complement the FFS but are absolutely not designed as a major payment. For example, the GP receives 40 euros a year for following each patient classified by the health insurance plan as chronically ill (*forfait pour affection de longue durée (ALD)*). In comparison, sector 1 GPs are paid 23 euros for each consultation at the physician's office.

Table III: Descriptive statistics

Variables	Sample	Mean value in Bourgogne	Difference sample and regional (p-value)	Mean value in France	Difference sample and national (p-value)
Age (mean)	51.5	51.2[(1)]	0.451 (n.s)	51.3[(1)]	0.588 (n.s.)
Gender (% of women)	27%	30%[(3)]	0.479 (n.s.)	31.2%[(2)]	0.277 (n.s.)
Sector of activity (% in sector 1)	93.1%	87.3 %[(1)]	0.485 (n.s.)	89.3 %[(1)]	0.623 (n.s.)
Rural practice (%)	44.5%	33%[(4)]	0	15.7 %[(2)]	0
Group practice (%)	47.5%	39.6%[(4)]	0.118 (n.s.)	44.5%[(2)]	0.567 (n.s.)
Health network membership (%)	41.9%	39%[(5)]	0.496 (n.s.)	Between 27 and 44%[(5)]	Not determined
Weekly acts (mean)	119	102.8[(1)]	0	French region[(5)] 102.4[(1)]	0

In the absence of exhaustive and homogeneous data source on private practice self-employed GPs, the regional and national values are derived from different sources:
(1) All private practice GPs - 2008 data - SNIIR - source: Eco-Santé France, Régions & Départements 2015 - IRDES (for the weekly activity, the number of annual acts has been divided by 46 weeks).
(2) All private practice GPs - 2009 data - ADELI - source: Sicart (2009)
(3) All private practice GPs - 2009 data - SNIIR - source: ORS Bourgogne (2010)
(4) Survey panel of five regions (*panel de médecins généralistes libéraux* DREES, URML, FNORS) - 2007 data - source: Aulagnier et al. (2007a)
(5) Survey panel of five regions (*panel de médecins généralistes libéraux* DREES, URML, FNORS) - 2007 data - source: Bournot et al. (2008)

Table IV: Selection of the number of classes for the LCM

	AIC	BIC	CAIC	Log Likelihood
2 classes	1885.706	1971.1219	1994.1219	-919.8531
3 classes	1859.6501	1989.6307	2024.6307	-894.8251
4 classes	1783.5122	1958.0576	2005.0576	-844.7562
5 classes	1780.0136	1999.1239	2058.1239	-831.0069
6 classes	1787.6742	2051.3492	2122.3492	-822.83712

Table V: Estimation of the mixed logit models

		MN1		MN2	
		Coefficient	t-Stat	Coefficient	t-Stat
Level of remuneration	Mean	0.0002***	(9.03)	0.0002***	(6.59)
	SD	-	-	0.0003***	(7.21)
Forfait	Mean	-0.4706*	(-2.41)	-0.6635*	(-2.57)
	SD	0.1203	(0.41)	0.1227	(0.34)
Pay-for-performance	Mean	-0.5085*	(-2.36)	-0.6608*	(-2.38)
	SD	0.9771***	(5.06)	1.2575***	(6.13)
Frequency	Mean	0.2652	(1.66)	0.3264	(1.69)
	SD	0.0782	(0.40)	0.2098	(1.00)
Definition of guidelines	Mean	0.4966*	(2.35)	0.6776*	(2.55)
	SD	0.2992	(0.97)	0.0796	(0.36)
Application of guidelines	Mean	0.2563	(1.24)	0.3396	(1.27)
	SD	0.106	(0.33)	0.5811*	(2.06)
Continuing education	Mean	0.6580***	(3.89)	0.8654***	(4.39)
	SD	0.371	(1.11)	0.0312	(0.11)
Information feedback	Mean	0.4070*	(2.07)	0.4801*	(2.06)
	SD	0.4751	(1.78)	0.1112	(0.30)
Solo practice	Mean	0.3476*	(2.19)	0.4902**	(2.61)
	SD	0.2119	(0.82)	0.4721*	(2.23)
Assistance by NPP	Mean	0.1057	(0.61)	0.1641	(0.78)
	SD	0.9063***	(5.72)	1.2831***	(7.45)
ASC	Mean	1.3462	(1.40)	1.737	(1.51)
Number of observations		3390		3390	
Number of respondents		303		303	
Log Likelihood		-908.4154		-879.5045	
AIC		1856.8309		1801.0089	
BIC		1979.4026		1929.7092	

*Significant at 5%; **significant at 1%; ***significant at 0.1%

Table VI: Estimation of the latent class logit model - 4 classes

	Class 1		Class 2		Class 3		Class 4	
	Coefficient	t-Stat	Coefficient	t-Stat	Coefficient	t-Stat	Coefficient	t-Stat
Level of remuneration	-0.0001	(-1.56)	0.0002***	(7.35)	0.0023***	(19.04)	-0.0030***	(-18.85)
Forfait	-0.2202	(-0.35)	-0.8085*	(-2.07)	-9.6873***	(-14.26)	-26.8455***	(-18.89)
Pay-for-performance	-1.5179	(-1.94)	0.6209	(1.71)	-24.0380***	(-29.94)	-20.2301***	(-15.56)
Frequency	-0.5197	(-0.84)	0.9612**	(2.70)	1.2295*	(2.31)	34.5608***	(34.54)
Definition of guidelines	1.8732	(1.76)	-0.1382	(-0.41)	-3.8807***	(-6.19)	46.0073***	(27.24)
Application of guidelines	2.0941	(1.79)	0.5921	(1.70)	-14.4822***	(-22.98)	17.9852***	(19.47)
Continuing education	3.6665***	(4.51)	-1.0573**	(-3.13)	11.0212***	(14.31)	-6.5797***	(-5.23)
Information feedback	-0.6791	(-1.29)	0.1495	(0.37)	7.4359***	(8.38)	-4.5607***	(-4.39)
Solo practice	-1.2745	(-1.63)	1.0318***	(3.44)	-3.9727***	(-7.85)	-8.1784***	(-9.51)
Assistance by NPP	-1.3672*	(-2.32)	0.4714	(1.88)	-11.7878***	(-17.33)	40.3411***	(36.88)
ASC	2.8629	(0.54)	2.7353	(1.00)	-67.6568	(.)	72.4956	(.)
Average class share	0.136		0.317		0.231		0.316	
Number of observations	3390							
Number of respondents	303							
Log Likelihood	-844.7561							
AIC	1779.5122							
BIC	2055.2985							

*Significant at 5%; **significant at 1%; ***significant at 0.1%

Table VII: Goodness-of-fit measures of the different specifications

	AIC	BIC	Log Likelihood
MN1	1854.961	1977.532	-907.4804
MN2	1794.699	1923.4	-876.3496
LC(4)	1783.5122	1958.0576	-844.75616

Table VIII: CAPI and alternative QIPs

	CAPI	Integrated primary care model (P1)	Mixed remuneration (P2)	Non-financial interventions (P3)	Maximum satisfaction (P4)
Level of remuneration	4200	4200	4200	0	4200
Method of remuneration	*Forfait* and P4P	*Forfait*	*Forfait* and FFS	No	*Forfait* and FFS
Frequency of remuneration	Annual	Annual	Annual	No	NA
Prevention clinical guidelines	No	Pre-established	No	Participatory	Participatory
Continuing education in prevention	No	Yes	No	Yes	Yes
Feedback on preventive practices	Yes	Yes	No	Yes	Yes
Group practice	No	Yes	No	No	No
Assistance by non-physician providers	No	Yes	No	No	NA

Note: In the last column, the frequency of remuneration and assistance by NPP are not considered because GPs are indifferent to it at the mean. The maximum satisfaction is defined for all GPs. The French *forfait* are paid annually per patient (P1). FFS means a payment at each visit and cannot be "monthly" or "annual", but mixed remuneration here includes a *forfait*, so we select the annual frequency for P2.

Table IX: Policy simulation: compensating variation (Euro per year)

		CAPI	Integrated primary care model	Mixed remuneration	Non-financial interventions	Maximum satisfaction
All GPs	Indirect utility	-0.5143	0.8974	0.1594	1.9093	3.5391
	CV	x	9113	4349	15646	26167
P4P "inclined"	Indirect utility	0.1824	1.0987	-0.7463	2.1993	3.1624
	CV	x	5915	-5995	13020	19238
P4P "adverse"	Indirect utility	-0,7121	0,8542	0,4087	18,295	36,504
	CV	x	10112	7236	16408	28163
Assistance by NPP "inclined"	Indirect utility	-0.7463	1.4502	-0.6057	1.6088	3.0452
	CV	x	14180	908	15204	24477
Assistance by NPP "adverse"	Indirect utility	-0.3814	0.2714	1.142	2.1565	4.103
	CV	x	4214	9835	16384	28951
Class 1	Indirect utility	-1.9848	2.6139	-1.9848	5.0337	5.3483
	CV	x	61397	0	93705	97905
Class 2	Indirect utility	2.0676	-2.9193	2.0676	-0.0255	0.914
	CV	x	-22293	0	-9357	-5157
Class 3	Indirect utility	70,764	-50,059	499,679	223,915	658,679
	CV	x	-5204	18474	6596	25323
Class 4	Indirect utility	-152.8812	6.3605	-76.4541	-13.6526	45.8256
	CV	x	53925	25881	47148	67290

Appendices

Appendix A: Example of choice set

	Option A	**Option B**
Income increase per year	**6100€**	**12100€**
Method of remuneration	***Forfait* + Fee-for-service**	***Forfait* + Pay-for-performance**
Frequency of remuneration	**Annually**	**Annually**
Work in group of general practitioners	**Yes**	**No**
Prevention clinical guidelines	**None**	You participate in their **definition and application**
Continuing education in prevention	**No**	**Yes**
Feedback on preventive practices	**No**	**No**
Assistance by non-physician providers during preventive work	**No**	**Yes**
	I prefer A	**I prefer B**
Tick one box	☐	☐

Note: translated from France

Appendix B: Construction of the CAPI scenario

Attributes	Level	Justification
Level of remuneration	4200	The maximum bonus a GP can earn is 7000€ a year, from which only 60% is imputable to preventive services. We select this maximum in order to evaluate the highest benefit that can be expected from the CAPI.
Method of remuneration	Forfait and Pay-for-performance	The CAPI introduced P4P in France. A forfait per patient is adjusted depending on the attainment of the clinical practice targets.
Frequency of remuneration	Annual	The payment is made at each anniversary of the signed contract.
Prevention clinical guidelines	No	Even though various guidelines exist, they are not linked with the CAPI.
Feedback on preventive practices	Yes	Information is fed back to the doctor each trimester as part of the CAPI.
Continuing education in prevention	No	Continuing education is only on a voluntary basis and is not linked to the CAPI.
Group practice	No	No incentive for GPs working in teams is included in the CAPI.
Assistance by NPP	No	Assistance by NPP is not provided or supported under the CAPI.