



ELSEVIER

Contents lists available at ScienceDirect

Health Policy

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

Impact of a pay for performance programme on French GPs' consultation length

Olivier Saint-Lary^a, Jonathan Sicsic^{b,*}

^a Collège National des Généralistes Enseignants (CNGE), University Versailles Saint-Quentin en Yvelines, Department of Family Medicine, 78180 Montigny le Bretonneux, France

^b CERME3, UMR8211, INSERM U988, Site CNRS, 7, rue Guy Môquet, 94801 Villejuif Cedex, France

ARTICLE INFO

Article history:

Received 24 January 2014

Received in revised form

11 September 2014

Accepted 5 October 2014

Keywords:

Consultation length

General practitioners

Quality of care

Pay for performance

ABSTRACT

Background: In 2009, a voluntary-based pay for performance scheme targeting general practitioners (GPs) was introduced in France through the 'Contract for Improving Individual Practices' (CAPI).

Objective: To study the impact of the CAPI on French GPs' consultation length.

Methods: Univariate analysis, and multilevel regression analyses were performed to disentangle the different sources of the consultation length variability (intra and inter physician). The dependant variable was the logarithm of the consultation length. Independent variables included patient's sociodemographics as well as the characteristics of GPs and their medical activity.

Results: Between November 2011 and April 2012, 128 physicians were recruited throughout France and generated 20,779 consultations timed by residents. The average consultation length in the sample was 16.8 min. After adjusting for patients' characteristics only, the consultation length of CAPI signatories was 14.1% lower than that observed for non signatories ($p < 0.001$). After adjusting for GPs' characteristics and the case mix, the CAPI was no longer a significant predictor of the consultation length. The results did not change significantly from one type of consultation to another.

Conclusion: Although the CAPI was extended to all GPs in 2012, our results provide a cautionary message to regulators about its ability to generate higher quality of care.

© 2014 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

The implementation of pay for performance (P4P) schemes concerning primary care has greatly expanded during the last 10 years [1]. In UK (a pioneer), P4P has been implemented for all GPs since spring 2004 throughout the Quality and Outcomes Framework [2]. Some other countries such as the USA, Australia, New Zealand and Israel have also adopted similar incentive payments methods

[3]. Since 2009, a voluntary-based P4P scheme has been implemented in France towards GPs throughout the *Contract for Improving Individual Practices* (CAPI in French). In France, a fee for service remuneration (FFS) scheme prevails for liberal practice and French patients are allowed to change GPs without any financial penalty. It has been shown that FFS does not encourage physicians to provide preventive services [5]. The main aim of the CAPI was to give GPs an incentive for improving the quality of care assessed by the achievement and/or improvement of 16 indicators covering three main fields: prevention and screening, chronic diseases follow up and prescription optimization (Box 1). The CAPI was initiated by the Public

* Corresponding author. Tel.: +33 149583541; fax: +33 149583438.
E-mail address: jonathan.sicsic@inserm.fr (J. Sicsic).

Box 1: Description of the CAPI indicators

1. Influenza. Patients aged over 65 years vaccinated/Patients over 65 years.
2. Breast cancer. Women from 50 to 74 years old participating in breast cancer screening/Women from 50 to 74 years of age (calculated on 2 years).
3. Vasodilators. Patients over 65 treated/Patients over 65 (Target ¼ decrease).
4. Benzodiazepines long half-life. Patients over 65 treated/Patients over 65 (Target ¼ decrease).

Diabetes

5. Number of diabetic patients with 3 or 4 HbA1C per year/number of diabetic patients.
6. Number of diabetic patients with ophthalmological control in the year/number of diabetic patients.
7. Number of diabetic patients (men over 50, women over 60) treated with statins and antihypertensive/number of diabetic patients (men over 50, women over 60) treated with antihypertensive drugs.
8. Number of diabetic patients (men over 50, women over 60) treated with antihypertensive drugs, statins and aspirin low dose/number of diabetic patients (men over 50, women over 60) treated with antihypertensive and statins.
9. Patients treated with antihypertensive normalized their blood pressure below 140/90 (declarative).

Prescription

10. Proportion of generics for antibiotics.
11. Proportion of generics for proton pump inhibitor.
12. Proportion of generics for statins.
13. Proportion of generics for antihypertensive drugs.
14. Proportion of generics for antidepressants.
15. Prescription of ACE inhibitors/prescription of ACE inhibitors + angiotensin II receptor antagonists.
16. Number of patients treated with low-dose aspirin/number of patients treated with antiplatelet agents

Health Insurance Fund in 2009, against the advice of the unions. GPs had the opportunity to sign individually a CAPI which allowed doctors to receive a maximum annual bonus of €5000 (representing nearly 7% of average turnover) depending on their performance concerning the indicators. No sanctions were planned, and it was possible to depart from the programme at any time upon written request. The number of indicators and the financial reward involved were both rather low. Regarding the implementation of the CAPI programme, the choice of indicators and their levels as well as the formula for bonuses was unilaterally fixed by the Public Health Insurance Fund without formally consulting GPs' scientific groups or unions. Between 2009 and 2011, 38% of eligible GPs subscribed to the contract. At the end of June 2010, more than 70% of signers received a bonus in addition to their fee-for-service remuneration. The average monetary reward was €3000 per year and varied from €1500 for the 10% lowest rewarded doctors, to €4900 for the 10% most rewarded doctors [6]. In 2012, given the CAPI relative success, medical unions and the Public Health Insurance Fund agreed to generalize the P4P device to all GPs unless they explicitly refused (overall only 3% refused).

The P4P program's efficiency is still at the heart of sharp debate, particularly regarding its ability to generate better quality of care [7–12]. Numerous studies have estimated the efficiency of P4P regarding assessed indicators of improvement [13–15] or costs cuts [16]. However, up to now, very few studies have examined the impact of a P4P programme on physicians' consultation length. There

is a debate between clinicians and economists on whether longer consultations are related to better quality of care. One could argue that, especially in the field of chronic disease management, delivering high-quality care tends to involve interventions that may not require more time for the consultation: multidisciplinary care, checklists and reminders, etc. [17]. Moreover, some studies have shown that longer consultations were associated with higher quality according to the patients [18,19]. Also, it has been shown that longer consultations were often associated with the treatment of chronic health problems and health promotion activities [20]. As a consequence, in their guidelines for GPs concerning chronic disease self-management, the Royal Australian College of General Practitioners (RACGP) encouraged doctors to schedule longer appointments to determine patient needs and formulate an appropriate management plan [21]. Results from systematic reviews showed that longer consultations were associated with better results for the patient [22,23] and allowed better care when confronted with the most difficult cases or psychological and social problems [19,24,25]. Furthermore, the Consultation Quality Index (CQI) – which is a holistic marker of quality for GPs – was based on three measures one of which included the consultation length [26].

The appropriateness of considering consultation length as a marker of quality seemed especially relevant in the context of fee-for-service payment which does not encourage long consultation duration. Thus, in line with the empirical literature and considering the French context, our hypothesis was that consultation length could reasonably be considered as a proxy for quality of care.

The analysis of the impact of incentive devices on the supply of health care services has been the subject of both theoretical and empirical literature. According to standard economic theory, the introduction of bonuses targeting some tasks/services improvement may, under certain conditions, increase the level of efforts or time spent on the tasks achievement [27,28]. However, the empirical literature has shown conflicting results. Chanel et al. studied the change of consultation offer (hours of work) following a hypothetical increase of GP fees. They showed that most physicians did not change their labour supply while some decreased it in response to higher compensations [29]. Dumont et al. analyzed physicians response to contractual changes and concluded that the adoption of a less incentivizing contract reduced GPs' consultation length [30]. Coming back to the French context, GP's opinions about the impact of the CAPI on their consultation length were studied by Saint-Lary et al. [31]. According to the interviewed doctors, the CAPI was seen as a way to enhance the time spent on some activities such as preventive practices or the follow up of patients with chronic diseases.

The French P4P device is different from other incentive schemes, as the contract agreement is based on volunteers. As a consequence, a self-selection phenomenon might occur. On the one hand, the CAPI signatories could potentially be more productive regarding the quantity of care provided as shown in Devlin and Sarma [32] and may thus practice shorter consultation lengths. On the other hand, the CAPI signatories might be those who think they produce better care quality and, as a consequence, practice

longer consultation lengths. Our study is a preliminary work aiming at assessing the relationship between the CAPI signature and French GPs' consultation length, in order to generate hypotheses on the impact of P4P devices on the quality of primary care. For this purpose, a specific statistical methodology was used (Section 2), which includes a multi-level model frame enabling to disentangle the different sources of the consultation length's variability. The results are presented in Section 3, and discussed in Section 4.

2. Data and methods

2.1. Data

The data came from the Ecogen database ('Eléments de Consultation en médecine GENérale', in French), made up from 54 residents who worked on the consultations results of 128 representative French GPs between November 1, 2011 and April 30, 2012. The selected GPs were all intern tutors and participated in the project voluntarily. Moreover, all the GPs belonged to a different practice, therefore the sample did not include multiple GPs within a same practice. Data were collected continuously during the 6 months of inclusion: each resident spent at least one day per week at the GP' office and the consultations were transcribed exhaustively. Therefore, the database contained 20,779 consultation-drawn observations. Each observation gave precise information about the context of the consultation (doctor's office or visit, date), the characteristics of the patient (age, gender, occupation, exemption status, whether it is a new patient or not) and of the GP (age, gender, geographical area, total number of consultations in 2011 according to the RIAP¹, sector of activity, practice area, group practice, CAPI signature).

For each observation, the reasons for consultation and the consultation results were collected following the International Classification of Primary Care (ICPC-2, 2nd Edition) along with the precise consultation length timed by the resident.

2.2. Statistical methodology

2.2.1. The multi-level model

We used a multi-level model including 2 levels: the variables concerning the patient or the consultation i were included in level 1²; the variables concerning the physician j were included in level 2. The model enabled to disentangle the components of the consultation variability arising from inter and intra doctor variability.

The general model to be estimated was the following:

$$y_{ij} = \beta_{0j} + \beta X_{ij} + \varepsilon_{ij} \quad (\text{level } 1)$$

$$\beta_{0j} = \gamma_{00} + \delta \text{CAPI}_j + \gamma_1 K_j + \gamma_2 X_j + u_{0j} \quad (\text{level } 2)$$

The interest variable featured the logarithm of the length of consultation i performed by physician j . The vector X_{ij} contained the covariates concerning the patient/consultation i , CAPI_j was a dummy variable indicating whether or not the physician j had signed the CAPI between July 2009 and November 2011. The vector K_j contained the covariates concerning the physician j . According to the Mundlak specification [33], the vector X_j contained the average of all patient/consultation variables in order to control for patients' structures (case mix). At last, two error terms were included into the model: u_{0j} represented the level 2 error term showing the influence on y_{ij} of unobserved characteristics varying from one physician to another and ε_{ij} was the level 1 error term varying from one consultation/patient to another.

2.2.2. The model's assumptions

In this model, several exogeneity assumptions were implicitly made. First, the level 1 variables (X_{ij}) were not correlated to the level 1 error term (hypothesis 1) nor to the level 2 error term (hypothesis 2). A potential violation source of hypothesis 2 could arise from patients' self-selection effects: certain categories of patients (for instance executives) may prefer longer consultations [34]. The introduction of the average of patient characteristics in the explanatory variables was meant to compensate this issue [33]. Secondly, the CAPI signature was not correlated either to the level 1 or level 2 error terms. However, these assumptions were very unlikely satisfied because the CAPI signatories might have selected themselves regarding for instance their age, practice area, number of consultations, as well as patients' characteristics. The introduction of the physicians' characteristics (K_j) and the patients' characteristics (X_j) enabled to control for selection effects on the observables.

2.2.3. The different models tested

Three nested models were estimated, successively adding the patients' characteristics (model 1), the doctors' characteristics (model 2), and the averages of the patients' characteristics (model 3). The three models were written in their developed form as follows:

$$y_{ij} = \gamma_{00} + \beta X_{ij} + \delta \text{CAPI}_j + u_{0j} + \varepsilon_{ij} \quad (\text{model } 1)$$

$$y_{ij} = \gamma_{00} + \beta X_{ij} + \delta \text{CAPI}_j + \gamma_1 K_j + u_{0j} + \varepsilon_{ij} \quad (\text{model } 2)$$

$$y_{ij} = \gamma_{00} + \beta X_{ij} + \delta \text{CAPI}_j + \gamma_1 K_j + \gamma_2 X_j + u_{0j} + \varepsilon_{ij} \quad (\text{model } 3)$$

The number of consultations in 2011, which could be seen as a proxy of the GP's medical activity, was included amongst the doctor characteristics in model 2 and model 3. In the fee-for-service context which prevails in France, the GP is assumed to make a trade-off between the number of consultations and the length of consultation, so the former might be a significant predictor of the latter, or vice versa. The number of consultation per year in 2011 could be endogenous (econometrically speaking) in models 2 and 3, so separate models were estimated including this variable or not. In models 2a and 3a this variable was not included whereas in models 2b and 3b, it was.

¹ The 'Relevé Individuel d'Activité Professionnelle' (Individual Professional Activities Record) is annually provided by the French Public Fund.

² We could not build a three level model distinguishing the patient and the session level because no *id* was collected for the patient.

First of all, the 3 models were estimated from all observations, whatever the reasons for consultation. Next, in order to ascertain that the results did not vary from one type of consultation (for a given reason) to another, model 3, which appeared the most robust, was estimated for different specific consultation reasons targeted by the CAPI indicators: diabetes, high blood pressure, and flu vaccination. Again, separate analyses were performed including or not the number of consultations per year in 2011. All the analyses were computed using PROC MIXED of SAS 9.3.

3. Results

3.1. Univariate and descriptive statistics

The sample's average consultation length was 16.8 min after suppression of extreme values, that is, after 114 (0.5%) observations over 60 min were excluded in order to avoid bias due to potential outliers. Three consultations out of four lasted between 5 and 20 min and more than 90% lasted between 5 and 30 min (see Fig. 1). The distribution of the consultation length is logarithmic, thus justifying that we model the log of the consultation length.

The descriptive statistics regarding the determinants of consultation length are presented in Table 1. The sampled physicians were all eligible to the CAPI. Most of them did not sign the CAPI (57%), were males (66%), over 45 years old (80%), with fixed fees (92%) and in a group practice (79%). In addition, they were very well represented nation-wide except in the south of France where there was a minority of them (6%). Their practice was both rural or semi-rural (48%) and urban (52%). Independence Chi-square tests between CAPI signatory and non CAPI signatory characteristics were performed. There was no significant differences between the two groups for all GPs' characteristics except the number of consultations in 2011, which was higher for the CAPI signatories ($p = 0.0262$). Whether the CAPI signatories were characterized by higher medical activity or whether the CAPI contract entailed higher medical activity could not be disentangled in our study because the total number of consultations was observed only after the signature of the contract.³

The variables regarding the patient such as age, gender or socio-professional category along with the variables regarding the physician had a statistically significant impact on the consultation length in the univariate analyses (Table 1). The average consultation length of the CAPI signatories was 15.6 min vs 17.6 min for the non-CAPI signatories ($p < 0.001$).

3.2. Multivariate models

3.2.1. Results for the all-symptomatic area models

When considering "all-symptoms consultations", over one quarter (28%⁴) of the consultations length variability

was linked to the GPs' characteristics and 72% was linked to those of consultation/patients.

The multivariate models' results are displayed in Table 2. In model 1, the CAPI signatories had a consultation length significantly lower than the non signatories ($-14.2%$, $p < 0.001$). The consultation length was significantly shorter regarding younger patients (less than 44), males, people benefiting from a public insurance (CMU) or the state medical aid (AME⁵), and at the GP's office as opposed to house visits. It was significantly longer for farmers, craftsmen, senior executives or retired people, having a long term disease or another exemption status and for new patients. The patients/consultations' variables explained three percent of the intra-physician variability.

In model 2, the results were adjusted for physician characteristics. The results of the CAPI signature effect on the consultation length were different according to whether or not the number of consultations in 2011 was included into the model. In model 2a, the CAPI signatories had a consultation length significantly shorter than the non signatories ($-13.1%$, $p = 0.005$), whereas in model 2b, the impact of the CAPI was no longer significant ($-7.4%$, $p = 0.102$). The impact of the consultation/patient variables was similar to model 1 whereas the physician variables, being a female, having a low number of consultations per year, and practicing extra fees impacted positively the consultation length. The physician characteristics accounted for 20% of the inter-physician variability in model 2a (for 31% in model 2b).

In model 3, the averages of the patient characteristics were included. The more females and/or farmers among the patient population, the longer the consultation was. The impact of the CAPI variable was similar to model 2 (still negative) but of different magnitude and no longer significant at the 5% level. In model 3a, the CAPI signatories had a consultation length shorter by $-8.7%$ ($p = 0.081$) compared to the non-signatories, and in model 3b, the estimated coefficient was $-3.2%$ ($p = 0.503$). The practice population characteristics additionally accounted for 13% of inter-physician variability in model 3a (14% in model 3b). Thus, the predictors introduced in model 3a explained 34% of inter-physician variability (43% in model 3b).

3.2.2. Multivariate models according to the reason for consultation

The results of the multivariate models according to the reason for consultation are displayed in Table 3. Three reasons for consultation associated with CAPI indicators were selected: diabetes ($N = 2821$), high blood pressure ($N = 3417$), and flu vaccination status ($N = 1356$). Working on more homogeneous consultations decreased the intra-physician variability for diabetes and high blood pressure consultation reasons but not for flu vaccination. The estimated intra-class correlation coefficient was 0.38, 0.37, and

³ The two hypotheses will be discussed in the next section. The total number of consultation in 2011 could not be used as an instrument of the CAPI signature because it was also a significant predictor of the consultation length.

⁴ This figure (the intra-class correlation coefficient) was estimated based on the empty multi-level model, i.e. the model without any predictors.

⁵ In France, private complementary health insurance is mainly purchased to cover re-insurable copayments left by public health insurance. Below an income threshold, individuals can benefit from a free health insurance called the CMU, and foreigners can benefit from a state medical aid called AME ('Aide Médicale d'Etat').

Table 1
Descriptive statistics and univariate analysis.

	Descriptive statistics			Univariate analysis	
	Non-CAPI signatories N (col %)	CAPI signatories N (col %)	Total N (%)	Average observed length (min)	Gross deviation/ ref value
Patient/session variables (N = 20,779)					
Patient's age (mean = 52 years)					
0–16	2111 (17.6)	1392 (15.9)	3503 (16.9)	14.8	***a,b –18%
17–44	3427 (28.5)	2337 (26.6)	5764 (27.7)	15.9	–12%
45–69	4007 (33.4)	3074 (35)	7081 (34.1)	17.6	–3%
70 and over	2461 (20.5)	1969 (22.4)	4430 (21.3)	18.1	ref ***
Patient's gender					
Male	4937 (41.1)	3736 (42.6)	8673 (41.7)	16.3	–5%
Female	7070 (58.9)	5036 (57.4)	12,106 (58.3)	17.1	ref ***
Patient's occupation					
Farmer/craftsman	357 (3)	292 (3.3)	649 (3.1)	17	10%
Senior executive	725 (6)	327 (3.7)	1052 (5.1)	19	23%
Middle category	790 (6.6)	457 (5.2)	1247 (6.0)	17	10%
Employee	2253 (18.8)	1761 (20.1)	4014 (19.3)	16.4	6%
Worker	452 (3.8)	368 (4.2)	820 (4.0)	14.9	–4%
Retired	3821 (31.8)	2979 (34.0)	6800 (32.7)	17.9	15%
No occupation	3604 (30)	2585 (29.5)	6189 (29.8)	15.5	ref ***
Patient's exemption status					
Long term disease (LTD)	2502 (20.8)	2030 (23.2)	4532 (21.8)	18.6	14%
Public insurance (CMU or AME)	428 (3.6)	406 (4.6)	834 (4.0)	14.8	–9%
Other	550 (4.6)	372 (4.2)	922 (4.4)	17	4%
None	8523 (71.0)	5957 (68.0)	14,480 (69.7)	16.3	ref ***
Context					
At office	11,191 (93.2)	8313 (94.8)	19,504 (93.9)	16.6	–14%
At the patient's	816 (6.8)	459 (5.2)	1275 (6.1)	19.2	ref ***
New patient					
Yes	719 (6.0)	427 (4.9)	1146 (5.5)	17.6	5%
No	11,288 (94.0)	8345 (95.1)	19,633 (94.5)	16.7	ref ***
Year					
2011	2825 (23.5)	2142 (24.4)	4933 (23.9)	16.3	–4%
2012	9182 (76.5)	6630 (75.6)	15,732 (76.1)	16.9	ref
Physician variables (N = 128)					
CAPI signature					
Yes	–	–	55 (43.0)	15.6	–11%
No	–	–	73 (57.0)	17.6	ref ns
Age					
30–44	13 (17.8)	12 (21.8)	25 (19.5)	16.1	–8%
45–54	29 (39.7)	21 (38.2)	50 (39.1)	16.2	–7%
55 and over	31 (42.5)	22 (40.0)	53 (41.4)	17.5	ref **
Gender					
Male	47 (64.4)	38 (69.1)	85 (66.4)	16	–12%
Female	26 (35.6)	17 (30.9)	43 (33.6)	18.2	ref ns
Geographical area					
Centre	12 (16.9)	8 (14.8)	20 (16.0)	20.9	25%
East	20 (28.2)	14 (25.9)	34 (27.2)	21.3	28%
Paris area	13 (18.3)	10 (18.5)	23 (18.4)	23.5	41%
North	14 (19.7)	6 (11.1)	20 (16.0)	16	–4%
South	3 (4.2)	5 (9.3)	8 (6.4)	16.5	–1%
West	9 (12.7)	11 (20.4)	20 (16.0)	16.7	ref ***
Number of consultations in 2011					
<4000	26 (35.6)	9 (16.4)	35 (27.3)	18.8	32%
4000–5000	24 (32.9)	15 (27.3)	39 (30.5)	18.2	28%
5000–6000	9 (12.3)	11 (20)	20 (15.6)	14.9	5%
>6000	14 (19.2)	20 (36.4)	34 (26.6)	14.2	ref *
Sector					
Sector 1 (fixed fees)	66 (90.4)	52 (94.5)	118 (92.2)	17.8	–20%
Sector 2 (extra fees)	7 (9.6)	3 (5.5)	10 (7.8)	22.2	ref ns
Practice area					
Rural	14 (19.2)	15 (27.3)	29 (22.7)	16.1	–8%
Semi-rural	22 (30.1)	11 (20)	33 (25.8)	15.8	–10%
Urban	37 (50.7)	29 (52.7)	66 (51.6)	17.5	ref ns
Practice style					
Multi GPs practice	45 (61.6)	34 (61.8)	79 (61.7)	16.6	–2%
Multi disciplinary practice	12 (16.4)	10 (18.2)	22 (17.2)	17.1	1%
Solo practice	16 (21.9)	11 (20)	27 (21.1)	16.9	ref

^a Fisher test: relationship between the independent variable and the consultation length (in log).

^b Significance level: *** = 1%; ** = 5%; * = 10%.

Table 2
Multivariate models of the consultation length (in log).

Models	Model 1	Model 2		Model 3	
		Model 2a	Model 2b	Model 3a	Model 3b
Intercept	2.85***	2.943***	2.831***	2.606***	2.448***
Level 1 variables (X_{ij})					
Age			<i>ref= +70 years</i>		
0–16	–0.088**	–0.087***	–0.087***	–0.069***	–0.069***
17–44	–0.073***	–0.071***	–0.071***	–0.063***	–0.063***
45–69	0.008	0.009	0.009	0.015	0.015
Gender			<i>ref= female</i>		
Male	–0.028**	–0.029**	–0.029**	–0.033***	–0.033***
Occupation			<i>ref= employee</i>		
Farmer/craftsman	0.041**	0.039*	0.039**	0.032	0.032
Senior executive	0.05***	0.05***	0.05***	0.051***	0.051***
Middle category	0.019	0.018	0.018	0.022	0.022
Worker	–0.025	–0.022	–0.022	–0.022	–0.022
Retired	0.021*	0.022*	0.022*	0.025*	0.025*
No occupation	0.001	0.002	0.001	–0.005	–0.005
Exemption status			<i>ref= none</i>		
Long term disease (LTD)	0.087***	0.089***	0.089***	0.094***	0.094***
CMU or AME	–0.046***	–0.044***	–0.044***	–0.041**	–0.041**
Other	0.059***	0.064***	0.064***	0.077***	0.077***
Context			<i>ref= at the patient's</i>		
At office	–0.081***	–0.08***	–0.08***	–0.098***	–0.098***
New patient			<i>ref= no</i>		
Yes	0.096***	0.097***	0.096***	0.096***	0.096***
Year			<i>ref= 2012</i>		
2011	–0.008	–0.008	–0.008	–0.006	–0.007
Level 2 variables					
CAPi signature			<i>ref= no</i>		
Yes	–0.141***	–0.131***	–0.074	–0.087*	–0.032
Physician variables					
Age			<i>ref= +55 years</i>		
30–44		–0.051	–0.070	0.048	0.023
45–54		–0.035	–0.031	–0.052	–0.040
Gender			<i>ref= female</i>		
Male		–0.123**	–0.092*	–0.054	–0.044
Geographical area			<i>ref= west</i>		
Centre		0.063	–0.030	0.009	–0.080
East		0.121	0.046	0.095	0.026
Paris area		0.025	–0.078	0.019	–0.072
North		0.120	0.018	0.075	–0.028
South		–0.022	–0.022	0.068	0.057
Sector			<i>ref= fixed fees</i>		
Extra fees		0.176*	0.18*	0.004	0.099
Practice area			<i>ref= urban</i>		
Rural		–0.060	–0.087	–0.027	–0.101
Semi-rural		–0.009	–0.033	0.046	0.000
Practice style			<i>ref= solo</i>		
Multi GPs practice		–0.044	–0.039	0.015	0.015
Multi disciplinary practice		–0.061	–0.028	0.029	0.067
No. of consultations in 2011			<i>ref= +6000</i>		
Less than 4000			0.227***		0.209***
From 4000 to 5000			0.232***		0.232***
From 5000 to 6000			0.031		0.043
Mundlak approach					
Proportion of females				0.005	0.005
Proportion of patients aged 0–24				–0.004	–0.001
Proportion of patients aged >70				0.000	0.001
Proportion of farmers				0.012	0.017
Proportion of senior executives				0.011*	0.009
Proportion of workers				–0.012	–0.009
Proportion of new patients				–0.005	–0.007
Random effects					
Level 1 (ε_{ij})	0.1833***	0.1911***	0.1911***	0.1914***	0.1914***
Level 2 (u_{0j})	0.0683***	0.0605***	0.0518**	0.0496**	0.0428**

* 10% (significance level).

** 5% (significance level).

*** 1% (significance level).

Table 3
Multivariate models of the consultation length (in log) according to the consultation reason.

Models	Diabetes		Blood pressure		Flu vaccination	
	Model 3a	Model 3b	Model 3a	Model 3b	Model 3a	Model 3b
Intercept	2.276***	2.168***	2.47***	2.332***	2.667***	2.624***
Level 1 variables						
<i>Age</i>				<i>ref = +70 years</i>		
0–16	−0.161***	−0.161***	0.021	0.023	0.056	0.066
17–44	−0.1***	−0.099***	0.063	0.064	−0.127	−0.118
45–69	−0.014	−0.014	0.025	0.025	0.028	0.037
<i>Gender</i>				<i>ref = female</i>		
Male	0.023	0.022	−0.023	−0.024	−0.034	−0.034
<i>Occupation</i>				<i>ref = employee</i>		
Farmer/craftsman	0.049	0.049	0.048	0.047	0.041	0.037
Senior executive	0.075*	0.075*	−0.004	−0.004	0.073	0.067
Middle category	0.036	0.035	0.036	0.036	0.023	0.017
Worker	0.022	0.02	−0.008	−0.01	0.252**	0.247**
Retired	0.002	0.002	0.018	0.018	0.064	0.069
No occupation	0.014	0.014	0.047	0.048	−0.017	−0.021
<i>Exemption status</i>				<i>ref = none</i>		
Long term disease (LTD)	0.055***	0.055***	0.092***	0.092***	0.019	0.024
CMU or AME	−0.129***	−0.129***	0.096	0.096	−0.087	−0.085
Other	0.073*	0.073*	0.057	0.055	0.14	0.124
<i>Context</i>				<i>ref = at the patient's</i>		
At office	−0.043	0.042	−0.038	−0.038	−0.077	−0.076
<i>New patient</i>				<i>ref = no</i>		
Yes	0.197***	0.197***	0.254***	0.252***	0.152***	0.158***
<i>Year</i>				<i>ref = 2012</i>		
2011	−0.002	−0.002	0.007	0.007	−0.026	−0.025
Level 2 variables						
CAPi signature				<i>ref = no</i>		
Yes	−0.093	−0.043	−0.07	−0.026	−0.035	0.004
Physician variables						
<i>Age</i>				<i>ref = +55 years</i>		
30–44	−0.079	−0.107	−0.042	−0.057	0.12	0.084
45–54	−0.088	−0.063	−0.076	−0.056	−0.065	−0.041
<i>Gender</i>				<i>ref = female</i>		
Male	−0.06	−0.058	−0.057	−0.049	−0.057	−0.062
<i>Geographical area</i>				<i>ref = west</i>		
Centre	−0.005	−0.087	−0.086	−0.152	0.084	0.019
East	0.19*	0.124	0.03	−0.019	−0.071	−0.126
Paris area	0.102	0.02	−0.019	−0.087	0.027	−0.047
North	0.158	0.058	0.035	−0.045	−0.059	−0.156
South	0.068	0.059	−0.047	−0.057	0.121	0.111
<i>Sector</i>				<i>ref = fixed fees</i>		
Extra fees	−0.078	0.013	0.014	0.103	0.024	0.123
<i>Practice area</i>				<i>ref = urban</i>		
Rural	−0.001	−0.073	−0.122	−0.179*	0.019	−0.041
Semi-rural	0.035	−0.015	0.047	0.011	0.03	−0.016
<i>Practice style</i>				<i>ref = solo</i>		
Multi GPs practice	−0.009	−0.011	−0.073	−0.072	0.069	0.056
Multi disciplinary practice	0.044	0.078	−0.013	0.012	0.081	0.107
<i>No. of consultations in 2011</i>				<i>ref = +6000</i>		
Less than 4000		0.173*		0.151*		0.133
From 4000 to 5000		0.199**		0.192**		0.183**
From 5000 to 6000		−0.012		0.037		−0.046
Mundlak approach						
Proportion of females	0.008	0.007	0.008	0.008	0.006	0.005
Proportion of patients aged 0–24	0.001	0.004	−0.001	0.001	−0.007	−0.005
Proportion of patients aged >70	0	0.001	0	0.001	−0.003	−0.002
Proportion of farmers	0.022	0.029*	0.004	0.008	0.003	0.009
Proportion of senior executives	0.015	0.012	0.009	0.007	0.014**	0.011*
Proportion of workers	−0.01	−0.007	−0.008	−0.005	−0.021**	−0.019**
Proportion of new patients	0	−0.002	−0.003	−0.004	−0.002	−0.003
Random effects						
Level 1 (ε_{ij})	0.1406***	0.1406***	0.1552***	0.1552***	0.2313***	0.2313***
Level 2 (u_{0j})	0.0694***	0.0646***	0.0698***	0.0677***	0.0407***	0.0338***

* 10% (significance level).

** 5% (significance level).

*** 1% (significance level).

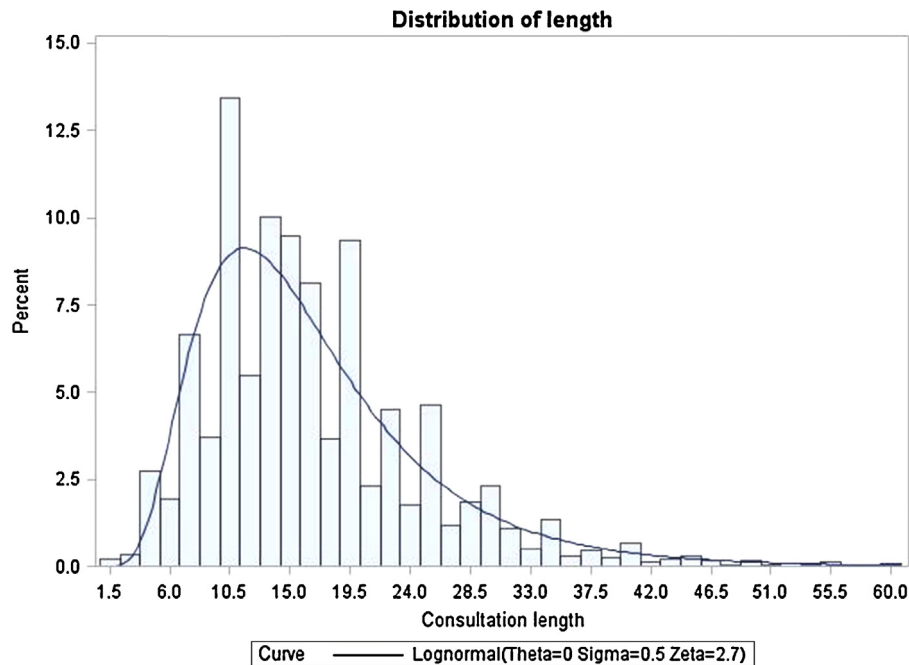


Fig. 1. Distribution of the consultation length ($N=20,665$ consultations).

0.21, respectively, compared to 0.28 for the all symptoms area models.

For the three reasons for consultation, being a new patient increased the consultation length. Younger patients and those benefiting from the CMU or AME had a shorter consultation length for diabetic care. The CAPI signature effect on consultation length was relatively similar from one consultation reason to another: the coefficient being negative and non-significant (except for flu vaccination for model 3b), regardless of whether the number of consultations in 2011 was included in the model or not. However, the coefficient value varied depending on the reason for consultation and whether or not the number of consultations in 2011 was included in the model. In model 3a, the coefficient varied from -9.3% for diabetes to -3.5% for flu vaccination and in model 3b, it varied from -4.3% for diabetes to 0.4% for flu vaccination.

4. Discussion

This study aimed to understand the recently introduced pay for performance device's impact on French GPs' consultation length. The French context was particularly favourable to tackle this issue as France is one of the few countries to have implemented a nationwide P4P system on a voluntary basis within the CAPI structure. A two steps methodology was used. Based on the 20,779 observations extracted from the *Ecogen* database, three nested models were estimated, progressively adding patient characteristics, physician characteristics and practice population characteristics within a multilevel structure. The most relevant model was then estimated according to specific reasons for consultation, subject to one or several CAPI indicators: diabetes, high blood pressure and flu vaccination.

The main concern came from possible self-selection effects leading to biased estimates of the impact of the CAPI on the consultation length. Indeed, CAPI signatories could for instance differentiate from the non signatories in terms of practice style or medical activity. The results of independence tests did not reveal significant differences between the two groups in terms of age, gender, area of practice or group practice. However, the two groups were different in terms of medical activity. The CAPI signatories had a higher number of consultations in 2011 compared to the non signatories, although it was not possible to conclude on the direction of the causality. First of all, one could assume that the CAPI signatories were more productive in terms of medical activity. Yet, the self-selection hypothesis could not be properly tested because the medical activity was not observed before the signature of the contract. Secondly, one could assume that the CAPI entailed an increase in the signatories' medical activity as measured by the number of consultations. Yet, this hypothesis seems less likely for several reasons. The main reason is that the CAPI does not contain per se incentives to increase the number of consultations because it does not count the number of sessions dedicated to a health problem. The only extent to which it could impact the number of consultations is through the patient base effect: the GP could "recruit" more patients to earn more bonuses, thus increasing the number of consultations. However, to the best of our knowledge, the empirical literature that accounts for this effect is scarce or inexistent. In any case, the suspicion of reverse causality lead us to estimate separate models adjusting or not by GPs' medical activity in 2011.

Our main result is that the CAPI did not have a significant impact on the consultation length, after adjusting for all possible confounders. However, the introduction of GPs'

medical activity in 2011 into models 2b and 3b changed the magnitude of the CAPI negative impact, which was lower compared to models 2a and 3a. This result seems to support the hypothesis of GPs self-selection into the CAPI contract partly due to their higher medical activity. The results regarding the consultation specific reasons (diabetes, high blood pressure or flu vaccination) permitted to validate indirectly the result, i.e. a non significant impact of the CAPI onto the French GPs' consultation length. One possible rationale for this result may be that the CAPI did not contain large incentives, i.e. the maximum amount of €5000 (only 7% on average turnover) was small compared to the FFS remuneration of French doctors.

It is to be noted that our models could not take into account all individual physician effects. Even if we controlled for a majority of observable physician variables, some variables remained unobserved such as the practice style, leisure tastes, ethics or intrinsic motivation. A potential hidden variable in the relationship between physician's characteristics, CAPI signature and consultation length might be physician's intrinsic motivation, which has been shown to have a positive impact of the length an individual engages in an activity [35,36]. Assuming that intrinsic and extrinsic motivations are substitutes rather than complements [37–39], one could hypothesize that the CAPI non signatories were more intrinsically motivated, thus explaining their longer consultation length. This assumption is consistent with Saint-Lary et al. [40] study results having shown that the ethical tensions associated to the CAPI contract were decisive factors preventing French GPs from signing it.

4.1. Limitations

A first limitation of the study concerns sample size is about the size of the sample: a population of 128 physicians, although reasonable, can generate some power issues when estimating the physician variables' parameters, which were measured with less precision. The second limitation comes from the way physicians were recruited, all of them being medical student/resident advisors and thus having potentially different practice interests. As a consequence, GPs in group practice and having signed the CAPI were slightly overrepresented (62% vs 54% [41] and 43% vs 38% [6]). However, the recruited GPs were representative of the French GP population with respect to age (52 years in average vs 53 years), and sex ratio (66% vs 67%⁶). A third limitation comes from the fact that during two-thirds of the study period (January, 2012 to April, 2012), the CAPI was generalized to all GPs; thus initially non CAPI signatories were exposed to the new P4P device. Complementary analyses showed that the CAPI generalization did not affect the results of our study. Lastly, no patient IDs were available and so correlation between patients with multiple consultations could not be taken into account in this study.

4.2. Conclusion

In France, the CAPI did not seem to have increased GPs' consultation length even though this device was supposed to promote better quality of care. As P4P is a costly device, these results bring a cautionary message regarding the P4P spread in France and in other countries as well as its capacity to generate high-quality care as measured by the consultation length. As consultation length is not a single measure for quality of care, future research should examine the issue by considering other measurements for quality (e.g. content of the consultation, change in follow up visits, patients' satisfaction) and compare the findings with our preliminary results.

Conflicts of interest

None declared.

Acknowledgements

We would like to thank all the Ecogen team, particularly its coordinator, Pr. Laurent Letrilliart as well as Pr. Alain Mercier and the CNGE who allowed us to use these data. Only the authors of this article are liable to the results. We are grateful to all the residents who collected the data.

We would also like to thank the PROSPERE team and especially Carine Franc for her advice, feedback and comments on the paper and Marc Le Vaillant for his contribution to the methodology assessment. We would also like to thank Anne-Laure Samson for her useful comments on the paper during Paris Dauphine annual doctoral day on October 2013.

References

- [1] Rosenthal MB, Dudley RA. Pay-for-performance. *JAMA The Journal of the American Medical Association* 2007;297:740–4.
- [2] Doran T, Fullwood C, Gravelle H, Reeves D, Kontopantelis E, Hiroeh U, et al. Pay-for-performance programs in family practices in the United Kingdom. *New England Journal of Medicine* 2006;355:375–84.
- [3] Morgan MAJ, Dunbar J. Implementing pay-for-performance in Australian primary care: lessons from the United Kingdom and the United States. *Medical Journal of Australia* 2011;194:213–4.
- [4] Franc C, Lesur R. Systèmes de rémunération des médecins et incitations à la prévention. *Revue Économique* 2004;55:901–22.
- [5] Le contrat d'amélioration des pratiques individuelles (CAPI). Extrait du rapport de la commission des Comptes de Sécurité sociale (septembre 2011): Fiche éclairage "maladie"; 2011. p. 128–31 [http://www.securite-sociale.fr/IMG/pdf/fiche_eclairage_maladie_capi_sept_2011.pdf].
- [6] Mannion R, Davies HTO. Payment for performance in health care. *BMJ* 2008;336:306–8.
- [7] Ryan AM. Has pay-for-performance decreased access for minority patients? *Health Services Research* 2010;45:6–23.
- [8] Woolhandler S, Ariely D, Himmelstein DU. Why pay for performance may be incompatible with quality improvement. *BMJ* 2012;345:e5015.
- [9] Maynard A. The powers and pitfalls of payment for performance. *Health Economics* 2012;21:3–12.
- [10] Jha AK. Time to get serious about pay for performance. *JAMA The Journal of the American Medical Association* 2013;309:347–8.
- [11] Eijkenaar F, Emmert M, Scheppach M, Schöffski O. Effects of pay for performance in health care: a systematic review of systematic reviews. *Health Policy (Amsterdam, Netherlands)* 2013;110:115–30.

⁶ Source: www.ecosante.fr Données: Drees, RPPS (2012).

- [13] Scott A, Sivey P, Ait Ouakrim D, Willenberg L, Naccarella L, Furler J, et al. The effect of financial incentives on the quality of health care provided by primary care physicians. *Cochrane Database of Systematic Reviews Online* 2011;9:CD008451.
- [14] Campbell SM, Reeves D, Kontopantelis E, Sibbald B, Roland M. Effects of pay for performance on the quality of primary care in England. *New England Journal of Medicine* 2009;361:368–78.
- [15] Ryan AM, Doran T. The effect of improving processes of care on patient outcomes: evidence from the United Kingdom's quality and outcomes framework. *Medical Care* 2012;50:191–9.
- [16] Cheng S-H, Lee T-T, Chen C-C. A longitudinal examination of a pay-for-performance program for diabetes care. *Medical Care* 2012;50:109–16.
- [17] Shojania KG, Ranji SR, Shaw LK, Charo LN, Lai JC, Rushakoff RJ, et al. Closing the quality gap: a critical analysis of quality improvement strategies, vol. 2: diabetes care. Rockville (MD): Agency for Healthcare Research and Quality (US); 2004.
- [18] Wilson A. Consultation length in general practice: a review. *British Journal of General Practice* 1991;41:119–22.
- [19] Gude T, Vaglum P, Anvik T, Bærheim A, Grimstad H. A few more minutes make a difference? The relationship between content and length of GP consultations. *Scandinavian Journal of Primary Health Care* 2013;31:31–5.
- [20] Howie JG, Porter AM, Heaney DJ, Hopton JL. Long to short consultation ratio: a proxy measure of quality of care for general practice. *British Journal of General Practice* 1991;41:48–54.
- [21] Royal Australian College of General Practitioners (RACGP). Putting prevention into practice: guidelines for the implementation of prevention in the general practice setting, 2nd ed. Melbourne: RACGP; 2006.
- [22] Wilson A, Childs S. The relationship between consultation length, process and outcomes in general practice: a systematic review. *British Journal of General Practice* 2002;52:1012–20.
- [23] Freeman GK. Evolving general practice consultation in Britain: issues of length and context. *BMJ* 2002;324:880–2.
- [24] Howie JG, Porter AM, Forbes JF. Quality and the use of time in general practice: widening the discussion. *BMJ* 1989;298:1008–10.
- [25] Howie JG, Heaney DJ, Maxwell M. Measuring quality in general practice. Pilot study of a needs, process and outcome measure. *Occasional Paper/Royal College of General Practitioners* 1997:1–32.
- [26] Mercer SW, Howie JG. CQI-2 – a new measure of holistic interpersonal care in primary care consultations. *British Journal of General Practice* 2006;56:262.
- [27] Dewatripont M, Jewitt I, Tirole J. Multitask agency problems: focus and task clustering. *European Economic Review* 2000;44:869–77.
- [28] Eggleston K. Multitasking and mixed systems for provider payment. *Journal of Health Economics* 2005;24:211–23.
- [29] Chanel O, Paraponaris A, Protière C, Ventelou B. Get paid more, work more? Lessons from French physicians' labour supply responses to hypothetical fee increases; 2010.
- [30] Dumont E, Fortin B, Jacquemet N, Shearer B. Physicians' multitasking and incentives: empirical evidence from a natural experiment. *Journal of Health Economics* 2008;27:1436–50.
- [31] Saint-Lary O, Plu I, Naiditch M. Adhérer ou pas au CAPI: de quel clivage des généralistes le paiement à la performance est-il le révélateur? *Revue française des affaires sociales* 2011;2–3:180–209.
- [32] Devlin RA, Sarma S. Do physician remuneration schemes matter? The case of Canadian family physicians. *Journal of Health Economics* 2008;27:1168–81.
- [33] Mundlak Y. On the pooling of time series and cross section data. *Econometrica* 1978;46:69.
- [34] Breuil-Genier P, Goffette C. La durée des séances des médecins généralistes. *Etudes Résultats DREES* 2006.
- [35] Deci EL. The effects of contingent and noncontingent rewards and controls on intrinsic motivation. *Organizational Behavior and Human Performance* 1972;8:217–29.
- [36] Deci EL, Ryan RM. The “What” and “Why” of goal pursuits: human needs and the self-determination of behavior. *Psychological Inquiry* 2000;11:227–68.
- [37] Deci EL. Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology* 1971;18:105–15.
- [38] Frey B. Not just for the money: an economic theory of personal motivation. Cheltenham, UK and Brookfield, US: Edward Elgar; 1997.
- [39] Sicsic J, Le Vaillant M, Franc C. Intrinsic and extrinsic motivations in primary care: an explanatory study among French general practitioners. *Health Policy (Amsterdam, Netherlands)* 2012;108(2–3):140–8.
- [40] Saint-Lary O, Bernard E, Sicsic J, Plu I, François-Pursell I, Franc C. Why did most French GPs choose not to join the voluntary national pay-for-performance program? *PLoS ONE* 2013;8:e72684.
- [41] Gautier A. Baromètre santé médecins généralistes; 2009.