HAL
open science

Physicians' Incentives to Adopt Personalized Medicine: Experimental Evidence

David Bardey, Samuel Kembou Nzalé, Bruno Ventelou

To cite this version:

David Bardey, Samuel Kembou Nzalé, Bruno Ventelou. Physicians' Incentives to Adopt Personalized Medicine: Experimental Evidence. 2018. halshs-01928128

HAL Id: halshs-01928128

https://shs.hal.science/halshs-01928128

Preprint submitted on 20 Nov 2018

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

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

amse
école d'économie d'aix-marseille
aix-marseille school of economics

Working Papers / Documents de travail

Physicians' Incentives to Adopt Personalized Medicine: Experimental Evidence

David Bardey
Samuel Kembou Nzalé
Bruno Ventelou

WP 2018 - Nr 23

Aix*Marseille université
cnrs
L'ECOLE DES HAUTES ETUDES EN SCIENCES SOCIALES
Inserm
IRD Institut de recherche pour le développement
Centrale Marseille

Physicians' incentives to adopt personalized medicine: experimental evidence

David Bardey[1], Samuel Kembou Nzalé[2], Bruno Ventelou[2]

November 2018

Abstract

We study physicians' incentives to use personalized medicine techniques, replicating the physician's trade-offs under the option of personalized medicine information. In a laboratory experiment where prospective physicians play a dual-agent real-effort game, we vary both the information structure (free access *versus* paid access to personalized medicine information) and the payment scheme (pay-for-performance (P4P), capitation (CAP) and fee-for-service (FFS)) by applying a within-subject design. Our results are threefold. i) Compared to FFS and CAP, the P4P payment scheme strongly impacts the decision to adopt personalized medicine. ii) Although expected to dominate the other schemes, P4P is not always efficient in transforming free access to personalized medicine into higher quality patient care. iii) When it has to be paid for, personalized medicine is positively associated with quality, suggesting that subjects tend to make better use of information that comes at a cost. We conclude that this last result can be considered a "commitment device". However, quantification of our results suggests that the positive impact of the commitment device observed is not strong enough to justify generalizing paid access to personalized medicine.

Keywords: Personalized medicine, fee-for-service, capitation, pay-for-performance, physician altruism and laboratory experiment.

JEL Classification: C91, I11.

---

[1] Los Andes University (Bogota, CEDE) and visiting fellow at Toulouse School of Economics.
[2] Aix-Marseille Univ., CNRS, EHESS, Centrale Marseille, AMSE

We thank Fabio Galeotti, Ellen Green, Izabela Jelovac, Stephane Robin, Marie-Claire Villeval, Marc Sangnier and Remi Suchon for useful feedbacks during the design of this study. We are grateful to Noémi Berlin, Leontine Goldzahl, Lise Rochaix, Anne Stenger and Jerome Wittwer for valuable feedbacks on the first draft of the paper. Our gratitude also goes to Massimo Finocchiaro Castro, Mathias Kiffman, Sönje Reiche, Ruochen Wu for valuable comments on the latest versions of the paper. Our special thanks to all seminar participants in Strasbourg, Catania, Paris, Bogota and Rennes. Thank you to Marjorie Sweetko for the excellent English editing on the last draft of the paper.

1. Introduction

Personalized medicine involves profiling patients to determine decisions, treatments or medical interventions according to their predicted best response. While the idea dates back to Hippocrates, advances in genomics and epigenetics over the last two decades have helped promote this type of medicine. However, even when personalized medicine technology is available, physicians actually tend to under-use it. Nor does the issue appear to be only one of cost. Health systems, and the populations covered, stand to benefit from the adoption of personalized medicine technologies. Better allocating treatments among patients is a promising way to reduce both health expenditure and adverse consequences of treatments (Nimmesgern *et al.*, 2017). Here, to tackle the issue of effective adoption of personalized medicine, we examine how physicians' payment schemes affect their incentives to use personalized medicine techniques, and the extent to which their patients may benefit from such practice.

To study how physician payment schemes, affect their decisions to use personalized medicine, we designed an experiment to replicate the physician's trade-offs under the option of personalized medicine information. As implemented, subjects were placed in a real-effort task game, as per Green (2014) and, less directly, Lagarde and Blaauw (2017). In our experiment, ninety-five prospective physicians performed a task simulating the option of access to information likely to help them take better care of a patient. First, in order to imitate the relationship between the physician and the patient, our game was similar to that of Green (2014), consisting in proofreading short texts with potential positive benefits for a third party (the "patient") when the texts were corrected well. Second, we designed the task to account for features of personalized medicine, viewed as a set of information that can guide physicians in formulating medical decisions. While a physician without personalized medicine information needs to consider a wide set of symptoms when taking a decision, personalized information allows the physician to focus on a limited subset of symptoms for quicker diagnosis and more effective treatment. To capture this feature, a subset of "priority sentences" was defined in the texts and only actions within this subset was considered to generate potential benefits for the third party. Thus, the proofreader's efficiency crucially depended on the informational input on priority sentences. Precise information on which sentences were "priority" was made available to the subjects, free of charge in some periods of the game but at a cost in others.

Depending on the payment scheme, the incentives for our subjects to make corrections in priority sentences differed, as do motivations for buying personalized medicine. The prospective physicians were subjected to different payment schemes in a mix of within/between design. We explored three pure payment schemes: fee-for-service (FFS), capitation payment (CAP) and payment-for-performance (P4P). CAP was designed as a payment *per* treated patient ("treated" text), while FFS was designed to reward the physician based on the quantity of services provided to the patient (number of words corrected in the text). Our P4P scheme was designed as payment for a minimum number of appropriate corrections in priority sentences. Our empirical strategy involved a two-stage panel least-square estimation, used to compare the behavior of subjects who bought the personalized medicine information with that of subjects who did not buy it, under the different payment schemes.

Our main results, translated from proofreading to medical care, can be summarized as follows. First, as expected, our experiment reveals that P4P –a quality-oriented remuneration scheme– yields stronger incentives to physicians to buy personalized medicine techniques than FFS or CAP. In line with this first result, it seems that our subjects were also sensitive to financial incentives in their patient-care activities: while they treated more patients when paid by CAP, they performed more medical interventions under an FFS payment scheme (already in Green 2014). We also found that CAP and P4P tended to generate similar incentives regarding the number of interventions; however, P4P was less effective in transforming free access to personalized medicine into overall quality care. Interestingly, our results reveal that the impact of the information conveyed by personalized medicine crucially depends on whether access to it is free or has to be bought. When access to personalized medicine comes at a cost, differences due to the informational input are magnified, greatly to the patient's benefit. We interpret this result as a "commitment device". In fact, once the subjects had bought the information, they made much better use of it, compared to the situation where access to personalized medicine information was free to all the physicians.

Finally, using a simple quantification of our experimental results to study whether it is advantageous to generalize paid access to personalized medicine, we find that paid access for all is not recommendable. Thus, since the outcomes from personalized medicine information are better when it comes at a cost, our results convey a strong policy recommendation: instead of providing free access to personalized medicine tools, the regulator should partially subsidize this access, opting for a P4P scheme to enhance the adoption of these tools.

Related literature

Physician payment mechanisms is one of the central topics in health economics (Ellis and McGuire, 1986). As eloquently explained in McGuire, (2000), the incentives generated by different physician payment schemes may depend on institutional features such as the identity of the payer(s), the existence of market competition between physicians, or whether the health system is a gatekeeping one. Our main contribution to this literature is to study physicians' incentives in the context of personalized medicine. To the best of our knowledge, only two articles really tackle this issue. Antoñanzas *et al.*, (2015) study the cost-effectiveness of implementing personalized medicine. Howard *et al.* (2017) report on the interaction between financial incentives and medical decisions when physicians can use personalized medicine tests to choose between conventional radiotherapy and intensity-modulated radiation therapy (IMRT) for US breast cancer Medicare patients, so as to identify patients who are highly responsive to the IRMT option. Physicians work either in free-standing clinics (where their FFS payment plan also includes a monetary reward for treating patients with IRMT), or in hospital-based clinics (where they receive no additional benefits). Howard *et al.* (2017) find that physicians in free-standing clinics tend to under-use personalized medicine tests. Thanks to this original study design enabling comparison between physicians' behavior in two institutional settings, their finding strongly suggests the need to explore the interaction between payment schemes and the adoption of personalized medicine.

Counterfactuals are not always available to evaluate the properties of different payment schemes at work in health systems, and this is especially true of their interaction with personalized medicine, which is relatively new. Our article therefore relies on the literature using experimental methods to study physician payment. Over the last decade, a burgeoning literature has used experimental economics to study physician payment: Hennig-Schmidt *et al.*, (2011); (Brosig-Koch *et al.*, 2017, 2016, 2013); Green, (2014); Godager *et al.*, (2016); Hafner *et al.*, (2017); Lagarde and Blaauw, (2017). The main messages of these articles can be summarized in four points. First, in terms of findings, there seems to be a consensus on the incentives from FFS (over-provision) and CAP (under-provision), as shown in theoretical and other empirical investigations. Second, there is growing evidence on the incentive role of pay-for-performance (P4P). Green (2014) reported, for instance, that P4P combined with FFS gives higher incentives for services than FFS alone, and P4P combined with CAP gives lower incentives for services than CAP alone. Third, recent papers in experimental health economics reveal that it is crucial to take into consideration physicians' altruism toward patients (see Brosig-Koch *et al.*, [2013];

Hafner *et al.*, [2017]). Finally, Ahlert *et al.*, (2012), Hennig-Schmidt and Wiesen (2014) and Hafner *et al.* (2017) warn the research community of the importance of the subject pool. Hennig-Schmidt and Wiesen (2014) find that a medical subject pool behaves differently from a non-medical subject pool, the former tending to be more "patient-oriented" than the latter.

In the light of this literature, we decided to recruit advanced medical students (prospective physicians) for our experiment. From a methodological point of view, we follow Green (2014) and, less directly, Lagarde and Blaauw (2017), who designed a real-effort task experiment rather than the 'declared-effort' used in earlier work on physician payment schemes. Equivalence between real and chosen effort has been proven for altruistic behaviors in gift-exchange games (Brüggen and Strobel, 2007). However, we felt that a real-effort task might be a more appropriate way to elicit subjects' decisions, especially when studying complex tasks like medical interventions. We therefore build on Green (2014)'s task using proofreading of texts. There are, however, three differences between our experiment and Green (2014)'s. (i) Contrary to Green's between-subject design, we use a within-subject design: each prospective physician is observed under two different payment schemes. (ii) We introduce personalized medicine by offering the physician an informational advantage that may benefit the patient (this is the core of our contribution). (iii) We control for heterogeneity in patients' actions by giving them a more passive role.[3] In Green's experiment, the optimal quantity of services depended on the interventions of a first set of subjects (the patients, hereafter subjects-1), while we impose more standardized behavior on the first set of subjects.

2. Data and methodology

Most experimental studies on physicians' behaviors aim at imitating the bilateral relationship between the physician and the patient. To recreate the physician-patient relationship, we first included an experimental session where economics students had to highlight words (Phase 1), thereby furnishing the "patient" base for Phase 2. The second phase addressed our main research question through several sessions with prospective physicians. Our instructions did not involve any contextual indications to physicians that they were acting out a patient/doctor relationship.

[3] As our main focus is the physician's behavior, patients have a "passive" role. In Green's experiment, patients' actions determined the ideal quantity of services from the physicians. Our design seeks rather to control patients' heterogeneity, to better focus on the issue of personalized medicine.

Overall, 48 short texts were used, 36 of primary-school level and 12 of first-year secondary-school level. Each text contained words with errors (spelling, syntax, vocabulary) highlighted in bold. Following Green (2014), we also put in bold 2 correct words, to leave room for over-provision of services by physicians. The words selected were easy to proofread, as the aim of the experiment was not to test the physicians' writing skills. To represent personalized medicine, we randomly underlined certain sentences, referred to as "priority sentences". These alone determined the patients' earnings, standing for their biological characteristics. All the texts contained priority sentences: information on them was not available in period 1 of the game, was given free in period 2, and could be bought in period 3.

2.1 Experimental design: Phase 1 of the game (passive patient role)

For the first step aiming to introduce "real" patients, we recruited 8 subjects from the department of economics of Aix-Marseille University to play the role of patients. Note that the main reason why we needed a first-stage subject (hereafter referred to as "subject-1") was to allow for other-regarding preferences.

Each of the 8 subjects was randomly assigned 2 sets of 6 texts. In one of the 2 sets, some words (both correct and incorrect) were in bold. The task of each of the 8 subjects was to highlight manually, on the unmarked set of texts, words that were in bold on another set. They used a yellow highlighter for this task. They were told that they were participating in a 2-phase game in which they were playing the first phase. To ensure incentive compatibility, we informed them that an additional payment would be generated by other subjects playing phase 2 of the game. Moreover, for this session, each of the 8 subjects was given a fixed endowment of €10. Each "bold-word" missed in the text incurred a penalty of €0.10.

Phase 1 took place in December 2016. All subjects behaved appropriately by "hand-highlighting" in yellow all the words found in bold in the other set of texts. Thereafter, in our instructions for the Phase 2 game with physicians, we made it clear to the physicians that their actions would benefit a real subject elsewhere, called "subject-1". In Figure 1, we present the timeline of the main steps of the experiment.

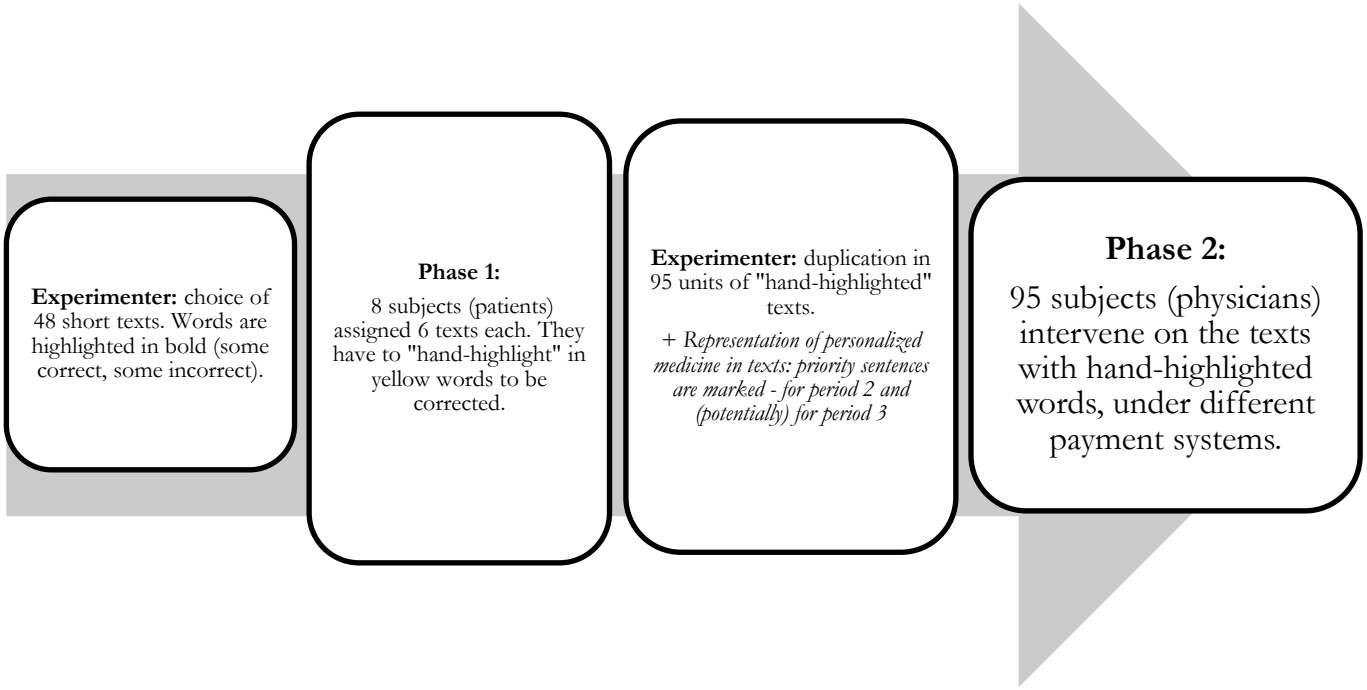


Figure 1: Timeline of the main steps of the experiment

2.2 Experimental design: Phase 2 of the game (physician role)

In phase 2, we ran different experimental sessions with advanced medical students playing physicians. We implemented a within-subject design by "treating" each physician subject with two different payment mechanisms. Such a design enhances statistical inference because each subject is his own control. We introduced a representation of personalized medicine by including access to information on priority sentences (underlined). Each treatment contained three successive periods of proofreading corresponding to three informational contexts:

- **Period 1**: 8 texts were presented without showing priority sentences, corresponding to a situation where personalized medicine is not available.
- **Period 2**: 8 texts were presented with priority sentences underlined, corresponding to a situation where personalized medicine is accessible free of charge.
- **Period 3**: the physician first had 1 minute to choose between a file of 8 texts with priority sentences underlined (personalized medicine) or a file of texts with no information. He then corrected the texts, playing the game either as in period 1 or 2. If he chose to have access to priority sentences, he was charged a fixed 0.50€ per treated text (a text was considered treated if there was at least one correction). He was not charged otherwise.

This corresponds to a situation where the physician chooses whether or not to buy personalized medicine.

Example of a text from period 1 (personalized medicine not available).

Il est inexplicable que nous soyons vivant. Je remonte ma lampe électrique à la main, les traces de l'avion sur le sol. A deux cent cinquante mètres de son point d'arrêt nous retrouvons déjà des ferrailles tordues et des tôles dont, tout le long de son parcours, il a éclaboussé le sable, […]

L'avion sans culbuté, a fait son chemin sur le ventre avec une colère et des mouvements de queue de reptile. A deux cent soixante-dix kilomètre-heure il a rampé. Nous devons sans doute notre vie à ces pierres noires et rondes qui roulent librement sur le sable et qui ont formé plateau à billes.

Example of a text from period 2 (personalized medicine available free).

J'admirais lentement ressortir de mille trous, de milles anfractuosités du roc, tout ce que mon approche avait fait fuir. Tout se mettait à respirer, à palpiter ; le roc même semblait prendre vie et ce qu'on croyait inerte commençait timidement à se mouvoir, des êtres translucides, bizarres, aux allures fantasques, surgissaient d'entre le laxis des algues ; l'eau se peuplait ; le sable clair qui tapissait le fond, par places s'agitait et, tout au bout de tubes ternes qu'on eût pris pour de vieilles tiges de jonc, on voyait une frêle corolle, craintive encore un peu, par petits soubresauts s'épanouir.

Un novice des choses de l'Alpe eût été surpris de constater la légèreté, contrastant avec la lourdeur générale de leur allure, avec laquelle les deux montagnards posaient le pieds sur les cailloux effrités du chemin. Aucune pierre ne roulait et les clous mordaient la terre avec ensemble, donnant l'impression d'une totale adhérence.

Figure 2: Examples of texts given to physicians in periods 1 and 2.

In Figure 2, we show an example of the texts given to physicians. The words in yellow are those to be proofread. Some of these words are correct, while others are not (in the Period 1 example above, only the words "vivant" and "culbuté" contain errors). The main difference between period 1 and period 2 is the fact that priority sentences are underlined in period 2. In period 3, depending on the physician's choice, the texts were either as in period 1 or in period 2.

We gave physicians 5 minutes per period to correct 8 short texts. They were free to allocate their time on the texts as they wished, including not altering some of them. For each treatment (payment mechanism), 24 texts were proposed (8 per period), so physicians could work on up to 48 texts per experimental session. Treatment variables were Capitation payment (CAP),

Pay-For-Performance (P4P) and Fee-For-Service (FFS). Under CAP, the physician was paid €1.75 for each of the 8 texts showing at least one intervention on the highlighted words regardless of whether appropriate. Under FFS, physicians earned €0.30 per intervention, again regardless of appropriacy. Each text had a minimum of 6 highlighted words and a maximum of 12, so that the earnings range per text under FFS was between €1.80 and €3.60. Under P4P, the physician earned €2.50 per text if 80% of words in priority sentences were correctly written at the end of the proofreading, and nothing otherwise. The priority sentences contained between 5 and 9 words, and their positions in the texts varied.

To avoid portfolio strategies, we remunerated 2 periods chosen randomly (one period for the first treatment and one period for the second treatment).

Table 1: Payment mechanism parameters (summary)

CAP	€1.75 per subject-1 treated
FFS	€0.30 per intervention
P4P	€2.50 per subject-1 treated

To implement the within-subject design, the game was presented to physicians as a game in 2 parts. In each part, the physicians played the 3 periods and they were informed of the payment system at the beginning of each part. We randomized the order of P4P.

Table 2: Different sessions of the experiment

Treatment for part 1 – Treatment for part 2	City	Number of physicians	Date of the session
Session 1: P4P – FFS	Marseille	24	January 2017
Session 2: CAP – P4P	Marseille	21	February 2017
Session 3: P4P – CAP	Nice	25	March 2017
Session 4: FFS – P4P	Nice	25	March 2017

The 4 sessions of phase 2 took place between January and March 2017. All our 95 prospective physicians were advanced medical students from Aix-Marseille and Nice Universities. As mentioned in the introduction, medical students were chosen because there is evidence that they provide a better sample for testing healthcare supply behaviors (Hennig-Schmidt and Wiesen, 2014).

Patients' benefits from Phase 2

The experimenter informed physicians that their actions could generate a financial gain for their subject-1 counterparts who had highlighted words in the texts in phase 1 of the dual game. This represents the altruistic part of medical activity.[4] Subject-1 counterparts would receive €5 if 90% of the highlighted words in priority sentences were corrected appropriately, and nothing otherwise.

Recruitment procedure

To ensure anonymity of registrations, students received an email from the student's association informing them about the sessions of the experiment, its expected duration (one hour) and the earnings range (up to €40). A dedicated website was constructed for registration, and all sessions took place in an auditorium. The auditorium was prepared prior to subjects' arrival with all the materials that would be needed during the experiment: pens, sets of texts and instructions. There was enough space between subjects to make it impossible to see other subjects' work. Subjects had 15 minutes to read instructions and complete a comprehension test on them (results available on request). To ensure anonymity with respect to the experimenter and the students' association, subjects' earnings were delivered to them by the university accountants after calculation.

At the end of each experimental session, we collected personal information covering gender, age, other demographic features, attitudes toward risk, and declared altruism. The questionnaire also included a set of other questions capturing attitudes and practices related to the proofreading task that subjects had to perform in the experiment: their perceived writing skills, their performance in secondary school, and their appetite for medical decision technologies (named TECHNO in the econometric analysis). Of this additional set, the first two variables were used as controls for the analysis, while TECHNO was also used as an instrumental variable.

The mean age of the 95 advanced medical students was 22 years old. They were mainly female (57%) and in the later years of their medical studies (58% in their fourth year or higher). The minimum earnings per physician was €6.20 and the maximum €35.40, with a mean and a

[4] There are two main ways to introduce altruistic preferences in experiments on physician payment: (i) experimenters inform the subject playing the role of physician that his game-generated gains for "patients" will benefit charitable organizations, or (ii) experimenters explicitly recruit subjects to play the role of patients, who will receive payments. Both scenarios are incentive-compatible ways of generating altruism used by authors cited in the introduction. However, the first method might be associated with free-riding on other subjects' altruism and therefore lead to overall artificially lower levels of altruism. For this reason, in our experiment we explicitly introduce patients as subjects-1 from the first phase.

median around €20. Out of the 95 subjects playing physicians, 5 did not collect their earnings. All subjects-1 collected their additionally-generated payments.

2.3 External validity of the experiment

Our experiment uses proofreading tasks to simulate situations in which the physician can benefit from personalized medicine. The patient declares his symptoms to the physicians (highlighted words in texts) and the physician intervenes to advise, diagnose and treat the patient (proofreading task). Periods 1 and 2 serve as initialization sessions, with personalized medicine free in period 2. The aim is to familiarize physicians with the game and help them understand the benefits of personalized medicine: the use of external technologies, yielding a more accurate and detailed patient profile. Thus, in period 3 of the experiment, we give the physician access to additional information on the patient's characteristics through the priority sentences, which only generate payment for subject-1. In the context of this experiment, the cost to physicians is monetary and the benefit to patients is a monetary reward.[5] Table 3 summarizes how our experimental settings correspond to real-life medical settings.

Table 3: Correspondence between experimental and real-life personalized medicine settings

	In experimental setting	In real-life setting
Baseline: Period 1 – similar to Green (2014)'s experiment	Crude declaration of wrong words by subject-1 **&** priority sentences not shown	Crude declaration of symptoms by patient **&** subset of relevant symptoms not shown
Period 2: *PM free*	Priority sentence shown, physician can target/focus interventions	Subset of relevant symptoms shown, physician can target/focus interventions
Period 3: *PM bought and voluntary*	Are you willing to buy the information on priority sentences?	Are you willing to buy (/spend time on obtaining) personalized medicine information?
Payment schemes	-% quality of overall text -per intervention -per text	-P4P -FFS -Capitation

[5] This monetarization is, in our view, the main limitation to the external validity of the experiment: in the real word, the reward to the patient is a health benefit and the penalty to the physician a time loss (although this could actually become a monetary loss in many payment systems).

The main attribute of personalized medicine is that it gives physicians the opportunity to focus on the relevant subset of symptoms, thereby achieving more effective selection of medical interventions. Adopting personalized medicine techniques usually has a cost, requiring doctors to leave their office for training in particular, but it increases the efficacy of their patient care. Our experiment aims to capture this fixed cost/variable cost trade-off. Our prospective physicians (in period 3) have to pay a price, intended to capture this opportunity cost of time. Our priority sentences, on the other hand, are intended to capture the potential efficiency gain for physicians from "buying" personalized medicine. Our prospective physicians can allocate this efficiency gain to treat more texts, or to increase the quality of their intervention on each text treated. It is well documented that personalized medicine techniques enable physicians to focus on the subset of symptoms that will allow them to choose the most appropriate therapeutic alternative for their patients' characteristics. Our experimental setting works in a similar way: instead of a badly informed doctor seeing various sentences in the text as alternatives for action, the well-informed doctor uses the information related to the priority sentence to choose the best course of action.

3. Results

We focus on two issues to describe physicians' behaviors. First, we look at their *decision to invest in personalized medicine information* through the decision in period 3 and we describe the main determinants of this choice, mainly in relation to the payment schemes. Second, we look at *the quality of services*. For this second issue, the main variable is having access to the information allowing personalization and its correlation with some key quality indicators; this correlation is also examined in interaction with the payment schemes. There are two sub-questions related to the issue of quality: Do the physicians' qualitative outcomes change when they obtain personalized medicine information free of charge? Do they change when this information is accessible but has to be paid for?

3.1 Result 1: Decision to invest in personalized medicine information

Our first results deal with the decision to acquire information allowing the practice of personalized medicine. In the following table, we report the decision to buy the information on priority sentences by payment mechanism.

Table 4: Decision to buy information and payment mechanisms

Payment system Decisions	P4P	FFS	CAP	Total decisions
Buy	**55**	**9**	**13**	**77**
Not buy	**40**	**40**	**33**	**113**
Total number of subjects	**95**	**49**	**46**	**190**

p-value = 4.236e-06 from a Khi-2 independence test.

Table 4 shows that the number of physicians choosing to buy personalized medicine information, *i.e.* paying for information on which sentences are priority, is greater in the P4P scheme (58% of subjects) than in the CAP (28% of all CAP subjects) and the FFS (19% of FFS subjects). Thus, at first glance the decision to buy personalized medicine information is not independent of the proposed payment scheme (p-value<0.05).

The decision to invest in such information was further investigated using a Probit model. We hypothesized that the decision to purchase information on priority sentences might be influenced not only by the payment scheme but also by a set of other explanatory variables: the physicians' self-declared appetite for innovative technologies (TECHNO variable, as determined from the questionnaire at the end of each session), their declared writing skills, their gender and their secondary school performance.

Our estimation results summarized in Table 5 reveal that there is a positive and statistically significant association between the purchase decision and the appetite of physicians for innovative technologies (Reference for interpretation: very likely). Other variables were used as controls for regressions (coefficients not shown). When it comes to payment methods, the Probit estimation corroborates the descriptive analysis: compared to the P4P, the FFS and CAP are less likely to be associated with personalized medicine purchase.

Table 5: Variables affecting the decision to buy information on priority sentences

		Dependent variable: decision to invest in the information on priority sentences
		Probit model
FFS (Ref: P4P)		-1.072*** (0.254)
CAP (Ref: P4P)		-0.919*** (0.247)
TECHNO	Strongly	0.606** (0.266)
	Weakly	0.838* (0.450)
Controls included?		*Yes*
Constant		-2.526 (1.610)
Observations		190
Log Likelihood		-110.488
Akaike Inf. Crit		238.976
Note:		*p<0.1; **p<0.05; ***p<0.01

The fact that P4P is associated with a higher probability of buying personalized medicine information can be explained by the opening for double motivation under P4P in physicians' preferences: expectation of financial return and altruism. Unlike P4P, buying information on priority sentences under CAP and FFS would stem from altruism alone, since these schemes do not provide physicians with any financial incentive to do so.

To describe quality outcomes, our identification strategy is twofold. We compare physicians' behaviors with and without free personalized medicine information, and we perform the same analysis when such information has to be paid for.

3.2 Result 2: Access to information and physicians' qualitative outcomes

Before describing our results on qualitative outcomes, a natural transition would have been to look at physicians' quantitative outcomes (number of interventions and number of texts). However, since our results are comparable to those in the literature, *i.e.* more interventions (words proofread) in FFS, more patients (texts treated) in CAP, these results are relegated to appendices. Interestingly, it is worth noting that CAP and P4P generated similar outcomes in terms of number of interventions, while FFS and P4P yielded similar outcomes in terms of

number of patients (indicators not statistically different across payment schemes). As our focus here is on the impact of personalized medicine techniques on patients' health, we select the variables involved in quality outcomes, with direct implications for patients' health. We first introduce our results on the setting where access to information was free.

3.2.1 Free access to information and physicians' qualitative outcomes

The design of the experiment allows us to compare results in period 1 with those in period 2, *i.e.* to compare behaviors in a "no information" setting with those in a "free information" setting. As is common, we first report descriptive statistics and complement these by estimating an econometric model to provide further evidence.[6] The econometric model is the following:

$$y_{itT} = c + \alpha_i + \beta_i Pay_{iT} + \gamma_i INFO_{it} + \theta_i Pay_{iT} * INFO_{it} + \Theta_i X_i + \epsilon_{itT} \quad (1).$$

In equation (1):

- y_{itT} is the outcome of $i, (i \in [1-95])$ physician; period t $(t \in \{no\ info, free\ info\})$ in treatment *T*. We will consider three outcome variables: the degree of focus of actions, the number of well-treated (appropriately corrected) texts and the rate of well-treated texts.
- c is the constant and α_i is the individual specific effect.
- Pay_{iT} is the payment mechanism in treatment *T*. This is a categorical variable with three modalities: P4P, FFS and CAP. Our reference is P4P. β is a vector of parameters that identifies the pure effect of the payment method on the outcome.
- $INFO_{it}$ is a binary variable equal to 1 in period 2 ("free information"), and to 0 in period 1 ("no information"). Our reference is "no information" (period 1). γ is a vector that captures the effect of information on the outcome.
- θ captures the interaction effect between free information and payment method. When significant, results are reported.
- X is the fixed set of objective time invariant control variables.
- ϵ is an idiosyncratic error term.

Due to the repetition of observations on the same subject (through our within-subject design), our dataset is a panel. Our three dependent variables are the physician's degree of focus,

[6] In particular, the experiment –voluntarily- created a self-selection process, through the decision to have access to the informational input, as can arise in real life. In some of the studied mechanisms, control of this self-selection bias (by its main correlates) was required.

the number and the rate of well-treated texts. Given the fact that our design uses a task involving specific skills (proofreading of texts), we seek to control the average effects by time-invariant individual characteristics such as performance at secondary school, self-declared writing skills and gender. Estimation results presented below are from a random effect model, applied to control and identify the effect of time-invariant regressors. The implicit assumption is that there are no unobserved individual characteristics influencing our dependent variable. This assumption is valid if the control questions, such as self-declared writing skills and gender, capture a sizable part of the inter-individual heterogeneity. The Hausman test is performed to challenge this intuition. Running a fixed and a random effect model, we do not reject the null hypothesis that the preferred model is the random-effect model.

3.2.1.1 Free access to information and physicians' degree of focus

The focus variable allows us to capture how physicians oriented their intervention with the informational tool at their disposal. We measure the focus variable by looking at the rate of interventions outside priority sentences (number of interventions outside priority sentences divided by the total number of interventions). The degree of focus captures the proportion of actions with no impact on the final benefit to subjects-1 (patients). This criterion is a measure of quality, as it captures how well the physician focuses on the patient's problem. Figure 3 and Table 6 present both descriptive statistics and results of our estimation.

Figure 3: Free access to information and physicians' degree of focus

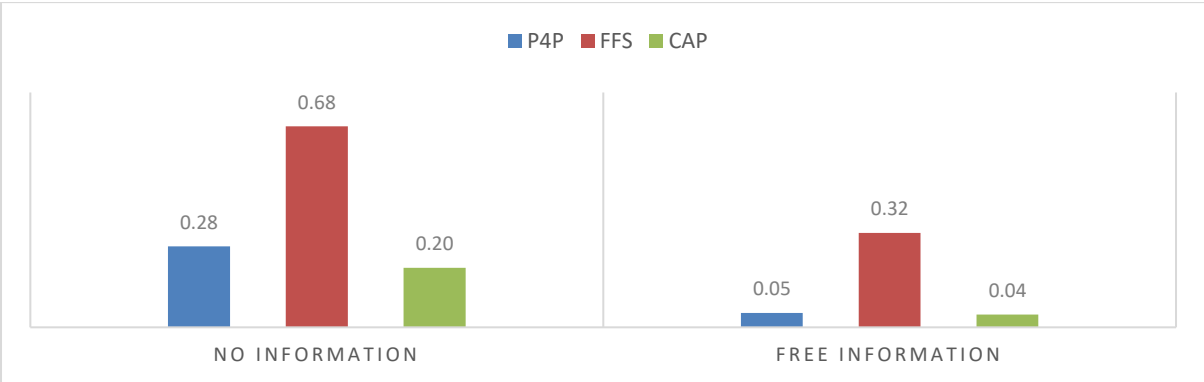


Table 6: Impact of free information and payment mechanisms on degree of focus

	Dependent variable: Focus (rate of interventions outside priority sentences) Model: Random effect panel linear model		
	(1)	(2)	(3)
FFS (Ref: P4P)	0.341***(0.036)	0.303***(0.046)	0.371***(0.056)
CAP (Ref: P4P)	-0.039(0.035)	-0.048(0.036)	-0.081*(0.047)
INFO (Ref: No info)	-0.244***(0.026)	-0.244***(0.026)	-0.227***(0.036)
Controls included?	*No*	*Yes*	*Yes*
INFO in the FFS payment system			-0.136**(0.061)
INFO in the CAP payment system			0.067(0.060)
Constant	0.284***(0.025)	0.511*(0.289)	0.503*(0.290)
Observations	190	190	190
R^2	0.522	0.528	0.550
Adjusted R^2	0.514	0.510	0.527
F Statistic	67.651*** (df = 3; 186)	29.130*** (df = 7; 182)	24.408*** (df = 9; 180)
Note:	*p<0.1; **p<0.05; ***p<0.01		

Remember that correcting words outside priority sentences is not of any benefit to patients.[7] When information is available, the degree of focus as measured by the intervention rate outside priority sentences is lower whatever the payment scheme (columns 1 and 2 of Table 6, variable 'INFO'; and Figure-3-histogram bars in "No information" *vs* bars in "Free information"). From Figure 3, we can see that, when information is made available in period 2, the rate of intervention outside priority sentences is halved in the FFS system, while it becomes five times lower in CAP and P4P. FFS is, in any case, always associated with the highest degree of focus. When access to information is free, P4P and CAP are not significantly different from one another (Table 6 and Figure 3). Thus, we can conclude that physicians really used information to focus their interventions on the priority sentences.

[7]… while it could be costly for the society, depending on the payment scheme.

This stronger impact on focus in FFS can be quantified by the 'INFO * FFS' interaction variable, which is significant, and adds an additional negative effect equal to -0.136 (Table 6, column 3). Despite their financial incentives, when physicians had access to personalized medicine information, they reduced their interventions outside priority sentences even though their income was increasing in the number of interventions. In the context of FFS, this result clearly reveals that our physicians had a more complex objective than mere profit maximization: they were behaving altruistically.[8]

3.2.1.2 Free access to information and physicians' well-treated texts

The number and rate of well-treated texts are other quality indicators that we use to describe physicians' performance. The first variable simply captures whether the physician's actions generated €5 for subject 1, while the second describes the ratio of well-treated to treated texts. The first variable provides insights into how personalized medicine and physicians' payment affect the number of patients effectively treated. The second is a more refined indicator that controls for the number of patients encountered (the denominator), which could differ across payment schemes and/or periods. We summarize our results in Table 7 and Figure 4. Table 7 only presents estimation results for the ratio of well-treated to treated texts. Econometric results on the well-treated texts are very similar.

[8] This finding on altruism is not new. Many previous works have established that physicians can demonstrate altruistic behaviors (most recently, Godager and Weisen, 2013; Green, 2014, among others).

Figure 4: Free information and physicians' number and rate of well-treated texts

As expected, the FFS system shows poor results when this second quality indicator is considered. More surprisingly, CAP and P4P still generate similar incentives, although 'in theory' P4P should be associated with a stronger incentive for quality interventions. The intermediary position of CAP may illustrate the fact that, without stressing the purely quantitative criterion of number of interventions performed (like the FFS), a remuneration scheme rewarding the number of patients treated incites toward a neutral quality/quantity trade-off, and therefore a middling level of quality. However, when we correct for total number of treated texts (as a denominator), CAP actually appears to lead to a lower rate of well-treated texts than P4P. In other words, we easily come back to the expected -and intuitive- result of better quality under P4P, after correcting for the quantitative effect of payment schemes (CAP incites physicians to treat more patients).

Last, as Table 7 shows, we do not find any significant effect of (free) information on quality. We will observe that this result is different when personalized medicine comes at a cost.

Table 7: Impact of free information and payment mechanisms on number of well-treated texts

	Dependent variable: rate of well-treated texts	
	Random effect linear panel model	
	(1)	(2)
FFS (Ref: P4P)	-0.297*** (0.048)	-0.263*** (0.062)
CAP (Ref: P4P)	-0.083 (0.047)	-0.073 (0.048)
INFO (Ref: No info)	0.03 (0.027)	0.003 (0.027)
Controls included?	*No*	*Yes*
Constant	0.536*** (0.031)	0.130 (0.383)
Observations	190	190
R^2	0.174	0.187
Adjusted R^2	0.160	0.156
F Statistic	13.018*** (df = 3; 186)	5.992*** (df = 7; 182)

*p<0.1; **p<0.05; ***p<0.0

Interaction between information and payment system is not included because not significant.

3.2.2 Paid access to information and physicians' qualitative outcomes

Access to information on priority sentences was available but had to be paid for in period 3. Due to the impact on benefits that we introduced, many factors could have played a role in physicians' decisions: expectations of "returns on investment", altruism, perceived writing skills, intrinsic "appetite" for information, and the payment scheme. All these factors are potential sources of endogeneity. We modeled the physician's decision by the binary variable "BUYINFO". Our estimation strategy therefore has to consider the endogenous nature of BUYINFO and propose a consistent method to examine its impact on physicians' behaviors. Having estimated a Probit model for the decision to buy information, we use estimation results in this subsection.

To compare the behaviors of information buyers and non-buyers, we provide estimates of a 2-stage Instrumental Variable Probit model. The Probit model estimated in the "first step" to predict the probability of investing in priority sentences under different payment schemes and

with the set of available individual characteristics is used here (TECHNO is our "instrument"). In this "second step" estimation, we use a two-stage panel least-square estimator, where the predicted value of the first-step model is included as an extra exogenous variable for our regressions. All the results presented in tables below are second-step regression results and corrected for the endogeneity of the decision to buy information on priority sentences.[9]

Formally, we estimate the following set of equations:

$$BUYINFO_{iT} = c + \mu_i Pay_{iT} + \eta_i\, TECHNO_i + \rho_i X_i + v_{iT} \quad (2)$$

$$y_{i,T} = c + \alpha_i + \beta_i Pay_{i,T} + \gamma_i \widehat{BUYINFO}_{iT} + \theta_i Pay_{iT} * \widehat{BUYINFO}_{iT} + \Theta_i X_i + \epsilon_{i,T}. \quad (3)$$

Equation (2) was previously estimated and results provided in Table 4. In equation (3):

- y is the outcome of individual $i, (i \in [1-95])$ in period 3, for treatment *T*. We use the same dependent variables as in the free information case.
- $\widehat{BUYINFO}_{iT}$ is the predicted value of the first-step regression (buyers/non buyers). γ is a vector that captures the <u>pure</u> effect of information on the outcome.
- $TECHNO_i$ captures the physician's appetite for new technologies. This is our IV-variable.
- All other variables are defined as previously.

As usual, to avoid the endogeneity problem, the second-step regression (3) does not include the raw variable BUYINFO, the "choice" made in period 3 *per se*, but rather $\widehat{BUYINFO}_{iT}$, the predicted probability. The TECHNO variable provides a strong instrument for modeling the decision to buy priority sentences, as TECHNO appears independent of all our dependent variables and correlated to the decision to buy personalized medicine. Three independent indicators are used to confirm that the instrument predicts the decision to invest in priority sentences. First, we verify that there is not independence between appetite for innovative technologies and decision to invest in personalized medicine (Fisher test on the contingency table describing the two variables < 10%). Second, comparing our regressions with and without the correction for endogeneity, we reject the null hypothesis that the instrument is weak (p-value <0.05). Third, in the regression analysis that models the decision to purchase personalized medicine, we observe that the appetite for innovative technologies is correlated with the likelihood of buying personalized medicine (See Table 5).

[9] For the 95 subjects, we have a total of 190 decisions observed. Using a panel technology in the second-step estimation preserves the longitudinal dimension of the model.

3.2.2.1 Paid access to information and physicians' degree of focus

We summarize our results on focus in Figure 5 and Table 8.

Figure 5: Degree of focus (paid information)

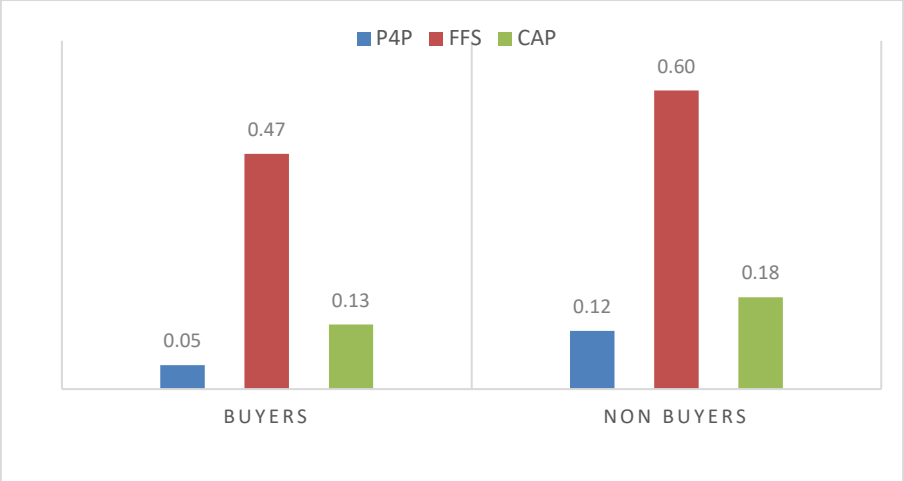


Table 8: Impact of buying information and payment mechanisms on degree of focus

	Dependent variable: focus (rate of interventions outside priority sentences) IV-Probit panel model		
	(1)	(2)	(3)
FFS (Ref: P4P)	0.338*** (0.068)	0.292*** (0.077)	0.653*** (0.123)
CAP (Ref: P4P)	-0.038 (0.057)	-0.078 (0.069)	0.130 (0.144)
BUYINFO (Ref: Non-buyers)	-0.280* (0.143)	-0.403** (0.179)	-0.146 (0.194)
Controls included?	No	Yes	Yes
BUYINFO in the FFS payment system			-1.414** (0.386)
BUYINFO in the CAP payment system			-0.477 (0.376)
Constant	0.269*** (0.086)	0.152 (0.242)	-0.100 (0.245)
Observations	190	190	190
R^2	0.468	0.487	0.523
Adjusted R^2	0.460	0.467	0.500
Residual Std. Error	0.208 (df = 186)	0.207 (df = 182)	0.200 (df = 180)
F Statistic	54.637*** (df = 3; 186)	24.686*** (df = 7; 182)	21.970*** (df = 9; 180)
Note:	*p<0.1; **p<0.05; ***p<0.01		

This table shows the second-step regression. See the first step in Result 1 (Table 5).

When considering the effect of information, the focus (intervention rate outside priority sentences) is much higher in the non-buyers group than in the buyers group. Even though the rate of intervention outside priority sentences is still higher under FFS, Table 8 reveals that the net effect of the information (purged of selection bias) is stronger in FFS than in the other two payment schemes. This result can be interpreted as a *commitment device effect* that appears to operate on physicians deciding to buy personalized medicine information despite being paid by a non-incentivizing scheme like FFS.[10] A commitment device effect is consistent with the fact that the rate of intervention outside priority sentences decreased by 0.14% when access to personalized

[10] This behavioral effect should be observed in CAP but is less visible in our data. In table 8, the coefficient is not significant. We think that our indicator of quality ("degree of focus = rate of interventions outside priority sentences") is not able to detect this subtle change in behavior for buyers (in CAP, 50% of texts bore only one correction, which could have been made randomly both in and outside priority sentences).

medicine was free (column 3 of Table 6), whereas under paid access to information, we observe a decrease of 1.4%. It would appear that our physicians tended to make better use of information when they had to pay for it.

3.2.2.2 Paid access to information and physicians' treatment of texts

Results on the two other indicators of quality are summarized in Figure 6 and Table 9.

Figure 6: Paid access to information and number and rate of well-treated texts

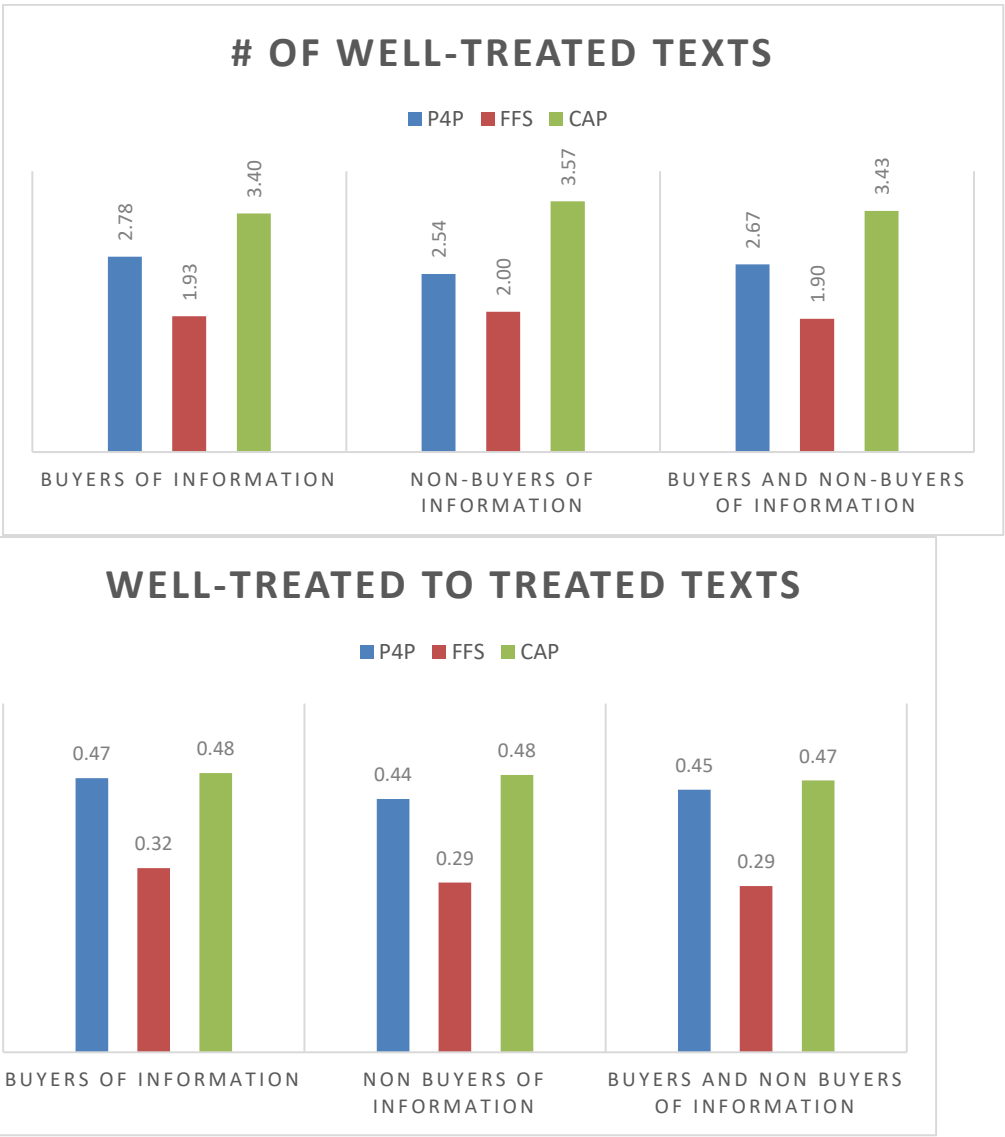


Table 9: Impact of buying information and payment mechanisms on ratio of well-treated to treated texts

	Dependent variable: Ratio of well-treated to treated texts
	Model: IV-Probit panel model
FFS (Ref: P4P)	0.001 (0.089)
CAP (Ref: P4P)	0.121 (0.079)
BUYINFO (Ref: Non-buyers)	0.391* (0.206)
Controls included?	Yes
Constant	-0.187 (0.278)
Observations	190
R^2	0.147
Adjusted R^2	0.114
Residual Std. Error (df = 182)	0.237
F Statistic (df = 7; 182)	4.482^{***}
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

This table is the second-step regression. See the first step in Result 1. Interactions are not shown because not significant.

In Table 9, an interesting difference appears for the *percentage* of well-treated dictations: acquiring information is not only associated with a decrease in the degree of focus, but this time the focus is "effective". It results in a significant effect on the quality criteria (slope equal to +0.395, Table 9). This is probably due to the commitment device already mentioned. When physicians decided to invest in acquiring information, they actually used it, improving their percentage of appropriately corrected texts.

To compare the payment schemes, Table 10 summarizes all our descriptive results on the qualitative variables.

Table 10: payment scheme ranking according to information structure

	No information	With free information	With bought information (comparison of buyers)	With bought information (comparison of non-buyers)
Degree of focus	CAP > P4P >FFS	P4P=CAP > FFS	P4P > CAP > FFS	P4P > CAP > FFS
% of well-treated texts	--	P4P = CAP > FFS	CAP = P4P > FFS	P4P = CAP> FFS

This last table compares the three payment schemes in terms of our two quality outcome variables. We use *t-tests* to compare the different means across payment methods. We consider whether personalized medicine information was accessible, and whether this access was free or had to be bought. This table shows that P4P and CAP generated very similar incentives, except for the focus variable, where P4P did better than CAP for both buyers and non-buyers. In the next section, we provide a quantification framework that enables us to address a potential policy issue: how should access to personalized medicine provided, free of charge or paid for?

4. Policy recommendation: should access to personalized medicine information be free or paid for?

One of our main results is that pricing the information conveyed by personalized medicine could yield a social benefit: physicians better exploit information they had to pay for.[11] However, a thorough policy recommendation should consider both the advantages and the disadvantages of any policy option. Here, charging for access to personalized medicine has the advantage of improving the effectiveness of information (because of the commitment-device effect described earlier), but the drawback of limiting access to information to those physicians not willing to pay for it. We provide a simple framework that highlights this trade-off for the policy-maker. Our key variable will be degree of focus, as it is the only variable for which interactions with payment schemes often appeared significant. We will also limit our reflection to P4P and FFS. Capitation

[11] As we have shown, this is not a self-selection effect, as in principle the selection is controlled for by the IV-method.

does not allow for comparisons, because there is not enough variation between free and paid access to information.

Let us suppose that we have a community composed of *N* physicians (*N = 95* in our case). Depending on the payment scheme (FFS or P4P), physicians jointly produce an outcome H^P (*P* stands for the payment scheme). Aggregate outcome for each payment option is the weighted sum of individuals' performances h^P, realized by B^P buyers and $(N - B^P)$ non-buyers:

$$H^P = B^P.h^P(Info = 1) + (N - B^P).h^P(Info = 0). \quad (1)$$

We are interested in variations of H depending on whether or not there is a price for information. Decomposition of equation 1 and simple differentiation[12] give:

$$\Delta H^P = B^P.\Delta h^P(Info = 1) + \Delta B^P.[h^P(Info = 1) - h^P(Info = 0)]. \quad (2)$$

The complete effect of charging for information is given by equation (2). The quantity Δh^P is the positive effect of the commitment-device (from free to paid info), as measured by the econometric estimation for "focus" as a dependent variable. The quantity ΔB^P is the variation in the number of buyers between period 2 and period 3, for payment scheme P; this number is always negative. In the FFS system, our experiment showed a decrease from 49 (100% of beneficiaries in period 2) to 9, $\Delta B^{FFS} = -40$. In the P4P system, the decrease was from 95 (100% of beneficiaries in period 2) to 55 in period 3, $\Delta B^{P4P} = -40$.

Using the simple calculation framework provided above, we can derive the full impact of charging for access to personalized medicine on the degree of focus of physicians' interventions. The aggregate variation of H in FFS would be given by:

== B^{FFS} x (slopes in Table 8) + ΔB^{FFS} (slope in Table 6 for the var. *Info*)

== 9 x (-1.414) – 40 (-0.227 + -0.136)

= +1.794.

In the same way, the aggregate variation of *H* in P4P would be given by:

---

[12] $\Delta H^P = B^P.\Delta \mathrm{h}^P(Info = 1) + \Delta B^P.h^P(Info = 1) + (\mathrm{N} - B^P).\Delta h^P(Info = 0) - \Delta B^P.h^P(Info = 0)$, or:

$$\Delta H^P = B^P.\Delta h^P(Info = 1) + (N - B^P).\Delta h^P(Info = 0) + \Delta B^P.[h^P(Info = 1) - h^P(Info = 0)]$$

We assume that $\Delta h^P(Info = 0) = 0$ -charging for the info has no impact on those who do not have access. Then, we obtain: $\Delta H^P = \mathrm{B}^P.\Delta h^P(Info = 1) + \Delta B^P.[h^P(Info = 1) - h^P(Info = 0)]$.

== B^{P4P} x (slopes in Table 8) + ΔB^{P4P} (slope in Table 6 for the var. *Info*)

== 55 x (0) – 40 (-0.227)

= +9,08.

A positive value implies that the quantity of useless interventions increases when physicians have access to information. In FFS, the full effect of charging for info is +1.794 more useless interventions. In the experiment, 9 physicians did better (-1.414 useless interventions per physician buying info). But charging a price increased the number of physicians without info to 40, leading to +0.363 (0.227+0.136) useless interventions *per* physician. In the same way, in P4P, the full effect of charging a price for info is +9.08 more interventions outside priority sentences.

Overall, our results show that, despite the existence of a commitment-device effect on the subset of buyers, it is still undesirable to organize paid access to personalized medicine for all physicians. This finding relies on the focus criterion and the set of incentives proposed in this experiment. It would have been interesting to perform the same analysis with other indicators. However, we did not find significant effects for the interaction between access to personalized medicine and the payment mechanism.

5. Concluding remarks

This article reported results from an experiment on physicians' incentives to use personalized medicine techniques. Our experimental design used the same task as Green (2014), where proofreading stood for medical services. Green (2014) and Lagarde and Blaauw (2017) demonstrated the feasibility of mimicking the physician-patient relationship using a real-effort task. We modified Green's experiment to consider the new context of personalized medicine, enriching the assessment of physicians' payment schemes to include physicians' choices on the use of personalized medicine tools, both free and paid for. We thus recreated the fundamental trade-offs of an agent (the physician) deciding on access to an informational technology like personalized medicine. This framework not only allowed us to complement the abundant literature on the incentive properties of physicians' payment schemes, but also to contribute to the economic analysis of a newly-relevant behavior: buying information (/technology) that can enhance the expert's service provision. There may even be room for generalization to other contexts (other types of expertise, like law or education) where the provider has to make an (unobserved) informational procurement effort, enhancing the quality of services.

Two questions have been answered in this article.

What determines the decision to adopt personalized medicine? We found that, compared to capitation and fee-for-service, pay-for-performance is associated with a higher probability of deciding to have access to information on priority sentences. Pay-for-performance was designed to reward the physician based on the number of well-treated patients. Investing in personalized medicine under a P4P scheme can stem both from altruism toward the patient and from expectation of higher financial returns. In CAP and FFS, only the altruistic motive plays a role in the decision, which is probably why these two payments schemes are less likely to incite physicians to pay to adopt personalized medicine.

What is the impact of personalized medicine on the quality of services? We found that information allows physicians to better focus their interventions, regardless of the payment mechanism. This focus effect is greater in FFS (probably because physicians were performing too many interventions in the no-information regime, which left more room for improvement). This result suggests the need to address the use of personalized medicine as related to the current payment mechanisms governing physician activities.

Last, information for personalized medicine, when it is accessible at a cost, is positively associated with the ratio of well-treated patients. Physicians tend to better use the information when they have to pay for it (all things being equal, including the selection bias). We concluded that this was consistent with a "commitment device". Using a simple framework to assess the consequences of paid access to personalized medicine, we found that charging for information is not desirable in P4P and FFS. While the experiment provided evidence that physicians better employ information they have paid for, charging for access will necessarily reduce the adoption of personalized medicine, which could be detrimental for patients. This trade-off must be taken into account when determining the optimal policy.

References

Ahlert, M., Felder, S., Vogt, B., 2012. Which patients do I treat? An experimental study with economists and physicians. Health Econ. Rev. 2, 1.

Antoñanzas, F., Juárez-Castelló, C.A., Rodríguez-Ibeas, R., 2015. Some economics on personalized and predictive medicine. Eur. J. Health Econ. 16, 985–994.

Brüggen, A., Strobel, M., 2007. Real effort versus chosen effort in experiments. Econ. Lett. 96, 232–236.

Ellis, R.P., McGuire, T.G., 1986. Provider behavior under prospective reimbursement: Cost sharing and supply. J. Health Econ. 5, 129–151.

Godager, G., Henning-Schmidt, H., Iversen, T., 2016. Does performance disclosure influence physicians' medical decisions? An experimental analysis. J. Econ. Behav. Organ. Forthcom.

Green, E.P., 2014. Payment systems in the healthcare industry: an experimental study of physician incentives. J. Econ. Behav. Organ. 106, 367–378.

Hafner, L., Reif, S., Seebauer, M., 2017. Physician behavior under prospective payment schemes: Evidence from artefactual field and lab experiments. FAU Discussion Papers in Economics.

Hennig-Schmidt, H., Selten, R., Wiesen, D., 2011. How payment systems affect physicians' provision behaviour—an experimental investigation. J. Health Econ. 30, 637–646.

Hennig-Schmidt, H., Wiesen, D., 2014. Other-regarding behavior and motivation in health care provision: An experiment with medical and non-medical students. Soc. Sci. Med. 108, 156–165.

Howard, D.H., Hockenberry, J., David, G., 2017. Personalized Medicine When Physicians Induce Demand, in: Economic Dimensions of Personalized and Precision Medicine. University of Chicago Press.

Lagarde, M., Blaauw, D., 2017. Physicians' responses to financial and social incentives: A medically framed real effort experiment. Soc. Sci. Med. 179, 147–159.

McGuire, T.G., 2000. Physician agency, in: Handbook of Health Economics. Elsevier, pp. 461–536.

Nimmesgern, E., Norstedt, I., Draghia-Akli, R., 2017. Enabling personalized medicine in Europe by the European Commission's funding activities. Pers. Med. 14, 355–365. https://doi.org/10.2217/pme-2017-0003

Appendix 1: summary of results on quantitative criteria

Figure 1: Free access to information and physicians' number of texts treated

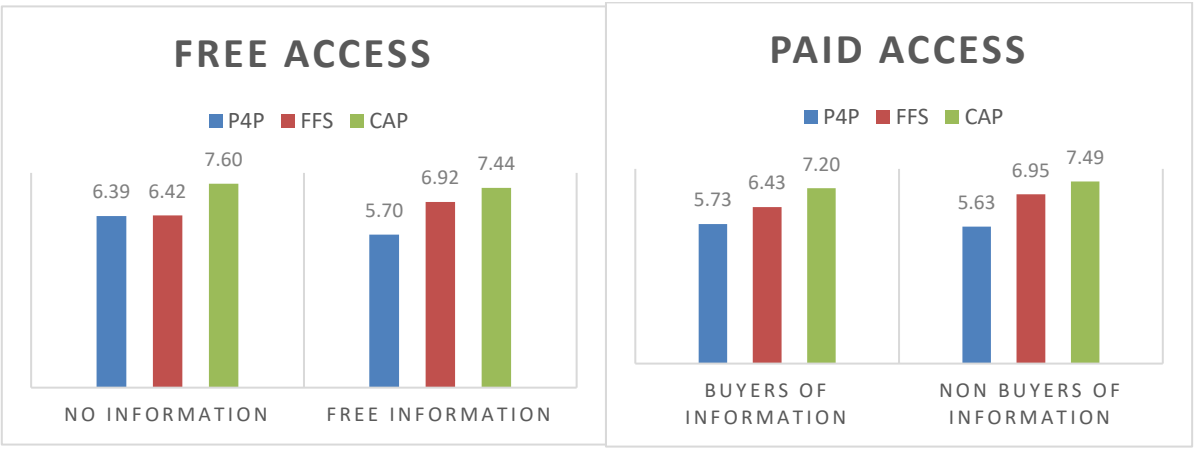
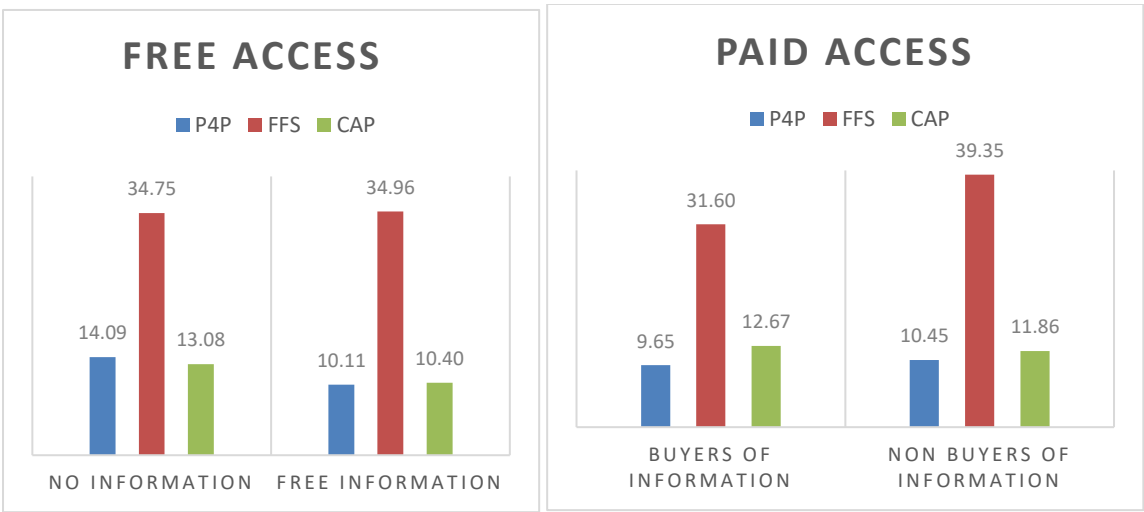


Figure 3: Free access to information and number of services



We summarize all our results dealing with these two quantitative variables in the following table.[13]

	No information	With free information	With paid information (where bought)
# of patients	CAP>FFS=P4P	FFS>CAP>P4P	CAP>FFS>P4P
# of services	FFS>CAP=P4P	FFS>CAP=P4P	FFS>CAP>P4P

Table 1: Payment scheme ranking according to information structure

Appendix 2: Experimental protocol (free translation).

[13] This ranking only uses descriptive statistics.

Note: These instructions were the same across sessions, only differing according to the payment mechanism studied. In the following, we give full instructions with P4P as the remuneration scheme and we provide the specific payment explanation that was used for capitation and fee-for-service.

You are participating in an experiment in economics. During this experiment, you will be paid based on your actions and decisions. Your actions are completely anonymous, and the data generated will be used by researchers. You received an ID when you arrived, and we will soon use it to establish your payment at the end of the experiment (2 – 3 weeks after this session). You will be paid with a voucher worth the amount you earned during this experiment.

You must carry out your actions individually. In other words, you are not allowed to communicate with other participants. Please also put your phone in silent mode and do not use it during the experiment.

This experiment is scheduled to last about 60 minutes and has two parts. You will receive detailed instructions before the start of each part. At the end of the second part, we will ask you to answer a short questionnaire.

<u>If you have any question at any time during the experiment, please raise your hand. This first part is composed of 3 periods.</u>

Preamble

We will ask you to work as an expert on the 24 texts that will be given to you, to correct the mistakes. You will choose the number of texts to correct, as well as the number and nature of the corrections. In the third period, you will be asked to decide whether to invest (via deductions from your earnings) on information that can help you in your task. Your final earnings will result from these choices.

In an experiment prior to this, we asked other subjects to work on the 24 texts that we are going to give you. For each text, they were instructed to highlight (with a yellow highlighter) the words they thought were incorrect. We have reproduced this highlighting in yellow on the computer.

Your role as an expert is to correct mistakes on the words that these subjects-1 rightly or wrongly highlighted. By mistakes we mean lexical mistakes, grammatical errors, misunderstandings and mistakes in conjugation. Your actions will determine your earnings as well as part of the earnings of the first participants (subject-1). A crucial point, which must be borne in mind, is that, for each text, only the corrected errors in certain sentences, which we will now

call "priority sentences", will generate earnings for the subject-1. In certain periods of the game, we will tell you which are the priority sentences, while in other cases you may be asked to decide to have this information (in exchange for a sum of money deducted from your earnings. We will return to this later).

Because of your correction work, the subject-1 will earn €5 for the text if, in the priority sentences (regardless of the total number of errors in the texts), you correct all the incorrect words, with one error allowed each time. Thus, if a text contains, for example, a total of 3 incorrect words in all its priority sentences, you will need to correct 2 out of the 3 to save €5 for the subject-1. The number and location of priority sentences vary from one text to another.

The game is calibrated so that the subject-1 earns between €5 and €30 according to the actions you have performed in this room (this remuneration is added to a small remuneration already granted to the subject-1 for the highlighting).

The 3 periods of the game.

3 game periods will follow. You can work on up to 8 texts for each of the 3 periods, that is 24 in total for this part. The 3 periods are as follows:

(i) During period 1, you will work on 8 texts maximum without any information about the priority sentences. In other words, your texts will only contain the yellow highlights suggested by the subject-1 but you will not know which sentence corrections would ensure him the highest earnings.

(ii) During period 2, we will indicate the priority sentences for the subject-1 by highlighting them. Thus, in period 2, you will know which sentences will potentially be of benefit to the subject-1.

(iii) During period 3, we will let you choose between two possible sets of texts: A and B. Your choice of a given set will be irreversible during this period. In set A, the 8 texts will be presented as in period 2 – with the underlined priority sentences – while in set B, the 8 texts will be presented as in period 1 – without any identification of the priority sentences–. If you opt for set A, you will bear a cost of €0.50 per text processed, pre-deducted from your earnings as the price for access to information. If you choose set B, you will not incur any costs and your texts will be presented as in period 1.

Your earnings (P4P)

For this part, we will remunerate the quality of your intervention on the texts. We have a "quality criterion" which we specify below. For each period, your earnings will be calculated as follows:

- Earnings in period 1 = €2.50* Number of texts for which the quality criterion is satisfied.
- Earnings in period 2 = €2.50* Number of texts for which the quality criterion is satisfied.
- Earnings in period 3 = €2.50* Number of texts for which the quality criterion is satisfied. From this sum we will deduct some expenditures for the period.
    - If you choose set A, the expenditure will be €0.50 * number of texts on which you have worked.
    - If you choose set B, you will have no expenditure.

Quality criterion: Each text has a minimum of 6 words and a maximum of 12 words highlighted in all sentences (priority or not). The quality criterion is exclusively based on the correctness of the priority sentences. The number of words that must be correctly written at the end of your intervention is presented as follows:

Number of words requiring intervention in priority sentences	3	**4**	5	6	7	8	9	10	11	12
Quality criterion: minimum number of words to be written correctly to earn €2.5 per text	2	**3**	4	4	5	6	7	8	8	9

The information in this table is to be read as follows (column in bold): If a text contains 4 words highlighted in the priority sentences (regardless of the number of words outside the priority sentences), at the end of your intervention 3 out of the 4 words must be written correctly in order for you to earn €2.50 for this text.

Note: At the end of the experiment, one of the above three periods will be drawn at random. Your earnings during the period will constitute your payment for that part.

Time allocation

Each period will last 5 minutes, and you are free to allocate your time between the texts as you wish. You can choose not to correct texts (no handwritten intervention on the text),

especially if you want to spend more time on other texts. At the end of each of the 3 periods, the experimenters will retrieve your 8 texts and start the following period with 8 new texts, signaling the kick-off for 5 new minutes. Between periods 2 and 3, you will have a moment dedicated to formulating your choice of one of the two sets of texts (A – priority sentences underlined – or B – priority sentences not underlined –).

End of the general instructions.

The following two paragraphs concern changes to the payment in each period.

Your earnings (FFS)

You receive in this part a "payment for intervention". The "Intervention payment" establishes a payment proportional to the number of corrections made (whether they are appropriate or not).

- Earnings in period 1 = €0.30 * Total number of words you have tried to correct among the words already highlighted.
- Earnings in period 2 = €0.30 * Total number of words you have tried to correct among the words already highlighted.
- Earnings in period 3 = €0.30 * Total number of words you have tried to correct among the words already highlighted. From this sum we will be deduct some expenditures for the period.
    - If you choose text set A, the expenditure will be €0.50 * number of texts on which you have worked.
    - If you choose text set B, you will have no expenditure.

Note: At the end of the experiment, one of the above three periods will be drawn at random. Your earnings during the period will constitute your payment for that part.

Your earnings (Payment per text) – CAPITATION –

You will receive in this part a "text payment". This "Text Payment" is a fixed remuneration per text on which you have made one or more corrections, whether these corrections are appropriate or not.

Earnings in period 1 = €1.75 * Total number of texts on which you have worked.

- Earnings in period 2 = €1.75 * Total number of texts on which you have worked.

- Earnings in period 3 = €1.75 * Total number of texts on which you have worked. From this sum we will be deduct some expenditures for the period.
    - If you choose text set A, the expenditure will be €0.50 * number of dictations on which you have worked.
    - If you choose text set B, you will have no expenditure.

Note: At the end of the experiment, one of the above three periods will be drawn at random. Your earnings during the period will constitute your payment for that part.