Aix-Marseille Université
École Doctorale de Sciences Économiques et de Gestion (E372)
Faculté d'Économie et de Gestion
Aix-Marseille School of Economics (AMSE)

Thèse de Doctorat ès Sciences Économiques

Présentée et soutenue publiquement par

Samuel KEMBOU NZALE

en vue de l'obtention du grade de docteur d'Aix-Marseille Université

Essays on healthcare providers' incentives and motivations

Essais sur les incitations et les motivations des médecins

Soutenue publiquement le 11 Octobre 2019 devant le jury composé de :

Alain TRANNOY	Aix-Marseille School of Economics, *Examinateur*
Paul BELLEFLAMME	Université Catholique de Louvain, *Examinateur*
Mylène LAGARDE	London School of Economics, *Rapporteur*
Jérôme WITTWER	Université de Bordeaux, *Rapporteur*
Izabela JELOVAC	Université de Lyon, *Directrice de thèse*
Bruno VENTELOU	Aix-Marseille School of Economics, *Directeur de thèse*

L'université d'Aix-Marseille n'entend ni approuver, ni désapprouver les opinions particulières du candidat :

Ces opinions doivent être considérées comme propres à leur auteur.

Dedicace:

To my wife Crescence, my mother Julienne and my father Jean-Guy.

Acknowledgments

I Can Do All Things Through Christ Who Strengthens Me - Philippians 4 :13

This PhD work has been a challenging but very exciting journey. I could have never gone this far without the precious help of many people to whom I would like to express my gratitude.

I would first of all like to be thankful to God Almighty, Father of our Lord and Savior Jesus Christ. Great and marvelous are His works. I am so grateful for all what He has done in my personal and academic life. I could have never gone this far without Him. I am also deeply grateful to my lovely wife, Arsène Crescence Kembou. Her humility, ease understanding and passion for living are constant source of motivation. I am always impressed by the confidence that she has me. I want to apologize for having had to make tough decisions for time allocation during all these years. My darling, please, see in this work, the results of your everlasting encouragements.

My gratitude also goes to my PhD advisors, Professor Izabela Jelovac and Professor Bruno Ventelou. They believed in me four years in the past by supporting my application to the PhD program of Aix-Marseille School of Economics. They have been working with me during these years, providing guidance and advise at different stages of my PhD life. They have also shown patience towards me, and are such a source of encouragement. This work might not be at his optimal potential, given all the effort they invested. I want to take this opportunity to apologize in advance. Dear Bruno and Izabela, I am sure that our ways will continue to cross and you will have other opportunities to appreciate the results of your efforts and investments.

Besides Bruno and Izabela, I would also like to thank all the PhD committee. Special thanks to Professor Mylène Lagarde and Professor Jérôme Wittwer, the two referees. Thank you so much for making all those insightful comments, suggestions and critics. I want to hope that this work reflects the effort that you invested. Thanks also go to Professor Paul Belleflamme for his strategic advise on how I could make the best out of my different research ideas. Thank you for accepting to be part of this adventure since my second year of PhD. I also express my deep gratitude to Professor Alain Trannoy for being so spontaneous when he was invited to be part of this PhD committee. I am grateful for all his comments and advise on this work.

My thanks also go to my former University of Lyon's professors who shared their passion for research : Marie-Claire Villeval has been such a source of inspiration for me. She has provided guidance during my first steps in experimental economics and she invited me for a seminar in Lyon where I presented the experimental protocol of one the studies, part of this work. Stephane Riou, Philippe Solal, Antoinette Baujard have inspired me to do some theoretical work. Their rigor and their discipline have motivated me. Lastly, Sylvie Demurger, Philippe Polomé, Julien Salanié, Lionel Perier have reinforced my appetite for doing applied work. I am grateful.

Special thanks also to the Aix-Marseille School of Economics' community, all the colleagues, faculty and support functions : Marc Sangnier, Raghul S Venkatesh, Nicolas Berman, Roberta Ziparo, Renaud Bourlès, Thimotee Demont, Mohammad Abu-Zaineh, Sebastian Bervoets, Gilles Dufrénot, Sylvie Boyer, Mathieu Faure, Alain Paraponaris, Anne-Gaelle Le Corroller, Sameera Awawda, Tanguy Le Fur, Edward Levavasseur, Jordan Loper, Armel Ngami, Ulrich Nguemdjo-Kamguem, Adrien Pacifico, Vera Danilina, Anushka Chawla, Edwin Fourrier-Nicolaï, Morgan Raux, Rémi Vivès, Lara Vivian, Mathilde Valero, Anna Zaytseva, Bernadette Vouriot, Corinne Michaud, Agnes Chaussonaud, Elisabeth Barthelemy, Aziza Sikar and Marine Bolea.

My Y15 office colleagues at AMSE, Alberto Prati, Pavel Molchanov, Victor Champonnois, Khalid Maman Waziri, Barnejee Anwesha, Suzanna Khalifa and Julieta Peveri have been source of encouragements. I will miss our fun times. I wish you the

best for all your future endeavors. Thank you to the people I met in Cambridge and Dartmouth college. Thank you to Professor Soenje Reiche for giving me the opportunity to spend these 5-inspirational months in Cambridge. It was very interesting to be connected to that environment. I still remember the theory workshop where I presented the first chapter of this PhD and all the insightful interactions that we had. Thank you to Emile and Ruochen for their friendship during my stay there. I am also very grateful to Bill Weeks who invited me to Dartmouth college in January 2018 and who displayed an outstanding commitment to help me maximize the gains of this stay.

I will also express my gratitude to all my colleagues at Innovations for Poverty Actions where I am Country Representative and Research Manager for Ivory Coast. My thanks also goes to all the researchers with whom we collaborate daily. Thank you for making this experience so beneficial for me and my career.

My friends and family are the last but not the least. Thank you to the following people who played a parental or mentor role in my life : Julienne Nzalé, Jean Guy Mboudjeké Nzalé, Raymond Mopoho, Jacqueline Nzalé, Madeleine Tchouandom, Honorine Kamdem, Odile Bella, Jeanne Nguemtchou, Fénélon Kamtié, Emmanuel Bekono, André Ambendet, Joseph Kabange, Valery Youdom, Eugene & Joëlle Sobngwi.

I am also grateful to my siblings, all the Nzalés : Hugues, Serge, Majolie, Francine, Eliane, Jacques, Ben, Josiane, Eric, Guy, Raymond, Margot.

Million thanks to (i) my *friends and family who were always supportive at all the stages of this PhD life* : Arnault Kamdem, Alain N'Ghauran, Steve Sighe, Jean Téné, Bill Tchamba, Martial Tchio, Leticia Lompo, Vincent Négou, Hyacinthe & Edith Kankeu, Carine Meyimdjui, Jaurès Zangue. (ii) All my gratitude to the *international community of friends and colleagues who has also played a great role in my personal life* : In St-Etienne : Hervé & Cendrine Mogto, Saïd & Christel Kelani, Florentin Kamga, Sandrine Kamga, Stephanie Akoundje, Ruben Fotso, Mark Zingbagba ; In Yaoundé : Serge & Golda Gouetna, Aubin Kacyem, Yves Wasnyo, Eric Hubert Ngoko, Franck Mouafo ; In Marseille : Aline, Adeline, Prince, Alain & Béatrice, Jasmine, Jean Hubert & Samira, Ida, Herbert & Elisabeth, Mamy & Ony, Abdias, Aïcha, Anne, Arnaud, Cherloune, Chris, Elorisse, Nelly, Francisca, Peggy, Imanoeala, Emma, Jeannine, Jetty, Nicolas & Marie-Cécile, Michèle, Muriel, Pierre, Prune, Robert, William, Nadège, Yvan, Gratien, Grâce ; everywhere else in the world : Fulbert & Sophie Kamguia, Liliane Nzeusseu, Dany Axel Kom.

May you all listed above find in these acknowledgment words the expression of my profound gratitude.

Contents

List of Tables

List of Figures

General Introduction

Health systems around the world often face major challenges: increase in disease prevalence, lack of coordination of health stakeholders, inefficient drug management systems, inadequate human resource management, skewed funding affecting medical equipment and staffing. All these challenges lead to poor access and overall inefficient health systems (World Health Organization, 2007). In many contexts, one of the first proposed solutions to mitigate these challenges is human resources management improvement. Health professionals are in fact one of the key actors determining the health system efficiency. In this work, we propose three essays around incentives and motivations of healthcare providers.

There are three main factors that can likely favor how we maximize the contribution of healthcare providers to build strong health systems. First, the remuneration systems that are used; second, how we factor ethical aspects in the design of health policies (healthcare professionalism, concern for reputation, altruism, etc.) and third, all effort enhancing technologies that can be provided to healthcare providers (medical equipment, training opportunities, administrative support, etc.). The importance of each of these aspects appears in both the economic literature and anecdotally in the political and social debates:

- Starting in 1986 with Ellis and McGuire, incentives of healthcare providers have remained a central topic in the health economic literature. Healthcare reforms touching healthcare providers' payment systems are recognized to be difficult to implement and generally lead to intense social debates.

- Since the seminal experimental work on physicians' payments (Hennig-Schmidt et al., 2011), there has been a growing literature trying to capture the relevance and the distribution of healthcare providers' soft-motivational elements such as altruism, professionalism, reputation concern, etc., using experimental economics. These elements are most of the time prominent in social debates on healthcare providers' behaviors.

- It is obvious that there have been improvements in all kinds of medical technologies. One of the most important aspect of these improvements is probably the introduction of personalized medicine which involves profiling patients to determine decisions, treatments or medical interventions according to their predicted best response. Antoñanzas et al. (2015) are the first that studied the economics of personalized medicine. Challenges and barriers to personalized medicine development appear in public communications around its development. This includes creatively thinking about better patient involvement; the question of data ownership, security and privacy; the quality of the medical technologies used and the incentives of medical professionals to adopt such tools.

This PhD work uses different methods to study incentives of different payment systems factoring the other two above-described contextual elements.

In **Chapter 1**, we propose a theoretical principal-agent framework to analyze optimal contracts in a setting that characterizes health and education sectors. The main ingredients of our theoretical modeling are: (i) limited liability constraints, (ii) information asymmetry (moral hazard on the agent's effort and adverse selection on the level of altruism), (iii) possibility to access for free an effort enhancing equipment that is divisible and contractible. We use a salary system as the payment method, with the regulator overseeing contract design.

Two aspects are worth mentioning to highlight our contribution to the strand of relevant studies. First, in our framework, the regulator and the agent jointly produce a non-contractible health or education outcome. The agent brings in her classical personal and professional effort, while the regulator contributes through providing access to effort enhancing technologies (an example being personalized medicine in the health sector). Second, the regulator faces a heterogeneous population of altruistic and selfish agents. Altruistic benefits of agents do not enter their participation constraints to rule out the fact that altruistic satisfaction can only be obtained inside a principal-agent relationship. We opt for a stronger version of the participation constraint, that is limited liability constraint which grants the agent a payment that is not lower than her cost of effort. We find that, the optimal menu of contracts that maximizes welfare, should specify higher transfers for altruistic agents and higher access to effort enhancing technologies to selfish providers.

This chapter increases our comprehension of the inter-play between the three ingredients described above: payment systems, the importance of ethics and the role of effort enhancing technologies. The assumption that the effort enhancing technology is divisible and contractible makes our proposed mechanism hard to implement. As mentioned above, an example of an effort enhancing technology is access to personalized medicine in the health sector. Its adoption generally depends on many other factors: whether its access is free or costly; the payment structures that are in place; how accurate personalized medical tests are; the role played by patients; data privacy and security; etc. In Chapter 2, we study alternative ways of adopting personalized medicine technologies (free *vs.* non-free access) and we also study the role of physician payment systems to favor this adoption.

Chapter 2 reports on a laboratory experiment that seeks to study physician incentives to adopt personalized medicine techniques. We use 95 prospective physicians in a real-effort game where their contribution to patient diagnosis and treatment is mimicked through a proofreading assignment. Personalized medicine access is introduced by allowing the game-player to know the relevant areas where his proofreading effort would likely maximize patient benefit. This research work increases our understanding of the main drivers of adoption of personalized medicine techniques. Both a free and a costly access to personalized medicine technologies are envisioned:

- In a free access setting, all the costs implied when using personalized medicine techniques during routine care are incurred by the regulator. France currently uses this setting. The health authority funds 28 genetic platforms geographically distributed around the country and physicians can rely on these platforms in their routine care at no cost for them.

- In a paid access setting, we aim to model a system where the healthcare professional bears all the costs of using personalized medicine techniques. Training and administrative costs are two examples of costs that should be accounted for when designing payment systems.

We study incentive properties of three payment systems: Fee-For-Service (FFS), Capitation (CAP) and Pay-For-Performance (P4P). We find that under P4P systems, paid access to personalized medicine techniques is higher compared to CAP and FFS. The inter-play between altruism and expectation of return in a P4P paid access framework, likely drives this decision. Regardless of the payment system however, we find that healthcare providers tend to better use personalized medicine techniques when they acquire it at a cost. We have called this a “commitment-device” phenomenon. Our results therefore convey an important policy message: access to personalized medicine techniques is not optimal when it is provided for free to all healthcare providers; it could rather be made costly, and P4P initiatives promoted to enhance adoption of personalized medicine techniques. This pivot role of the P4P is an interesting result. There is in fact existing research assessing P4P initiatives and finding modest (if any) incentive effect of P4P approach.

In **Chapter 3**, we study the persistence of P4P effects. There is a literature documenting the fact that P4P effects last for longer periods. We take advantage of our experimental design where P4P incentives are decided to be either anterior or posterior to more classic payment systems (FFS and CAP payment systems in our case). Our experimental setting allows us to compare the behavior of "exposed" (those who were "treated" with a P4P system prior to having CAP or FFS), to "non-exposed" (those who were treated with FFS and CAP prior to having P4P) providers. Looking at a subset of overall behaviors, we find that there is a persistent but ambiguous effect of P4P incentives. In FFS and CAP, exposed medical professionals increase their focus on what is relevant for the patient: they make higher efforts in areas that are the most relevant to generate patient benefits. At the same time however, in CAP, exposed healthcare providers reach overall lower levels of quality, being very likely an indication that P4P exposure destroys intrinsic motivation.

Overall, we have used both a normative and an experimental framework to study incentives and motivations of healthcare providers. The three main features that has received attention are: (i) remuneration systems of healthcare providers; (ii) consideration of ethical aspects describing healthcare providers (focus on altruism); (iii) study of effort enhancing technologies (focus on personalized medicine). The two first points have received considerable attention in earlier work, except for the study of adverse selection on altruism, where there is a relatively limited number of theoretical papers. The last point on effort enhancing technologies, particularly the focus on personalized medicine has not received much attention in the literature. Our "spelling check-task experiment" has allowed us to analyze the effect of different payment schemes for physicians in the presence of personalized medicine, *i.e*, whether the way physicians are paid impacts on their willingness to acquire information on their patients' medical needs. This research increases our understanding of how personalized medicine technologies can be provided.

Before moving to the main body of the thesis, we propose a foreword to describe the access to personalized medicine technologies in France. The aim of this presentation is to give more context to the notion of personalized medicine. We study the determinants of patient access to personalized medicine technologies in France, where it is currently free of charge for patients and physicians who treat them. This preliminary analysis would be of interest for a policy maker when studying access to personalized medicine technologies. Relation between personalized medicine and physician incentives will be studied more in-depth in the chapters 2 and 3, with chapter 1 proposing a more general framework for our entire work.

References

- Antoñanzas, F., Juárez-Castelló, C. A., & Rodríguez-Ibeas, R. (2015). Some economics on personalized and predictive medicine. The European journal of health economics, 16(9), 985-994.
- Ellis, R. P., & McGuire, T. G. (1986). Provider behavior under prospective reimbursement: Cost sharing and supply. Journal of health economics, 5(2), 129-151.
- Hennig-Schmidt, H., Selten, R., & Wiesen, D. (2011). How payment systems affect physicians' provision behaviour—an experimental investigation. Journal of Health Economics, 30(4), 637-646.
- World Health Organization. (2007). Everybody's business–strengthening health systems to improve health outcomes: WHO's framework for action.

Introduction générale (en français)

Les systèmes de santé du monde entier sont souvent confrontés à des défis majeurs : augmentation de la prévalence des maladies, faible coordination entre les différents acteurs du système, inefficacité des systèmes de gestion des médicaments, gestion inadéquate des ressources humaines, financement asymétrique affectant les ressources disponibles. Tous ces challenges conduisent à un accès qui reste globalement limité pour les patients, mais aussi à des systèmes de santé qui sont globalement inefficaces (World Health Organization, 2007). Dans la plupart des cas, l'une des premières solutions proposées pour atténuer ces problèmes est l'amélioration de la gestion des ressources humaines. Les professionnels de la santé sont en effet au cœur de nos systèmes de santé. Dans cette thèse, nous proposons trois essais sur les incitations et les motivations des médecins.

Trois principaux facteurs sont susceptibles d'améliorer la contribution des médecins et ainsi favoriser le processus de construction de systèmes de santé solides. Premièrement, les systèmes de rémunération qui sont utilisés ; deuxièmement, la manière dont nous prenons en compte les aspects éthiques dans le conception des schémas de paiement (la prise en compte des éléments tels que le professionnalisme, le souci de réputation et l'altruisme des médecins) ; et troisièmement, tout effort venant du régulateur et ayant pour objectif d'améliorer la qualité des ressources disponibles pour les médecins (la fourniture d'équipements médicaux, les opportunités de formation disponible, les soutiens administratifs offerts ; etc.). L'importance de chacun de ces aspects semble apparaître que ce soit dans la littérature économique ou alors de manière anecdotique dans les débats sociaux :

- Depuis les travaux de Ellis et McGuire en 1986, les incitations des médecins sont restées un sujet central dans la littérature en économie de la santé. Par ailleurs, les réformes du secteur de la santé – en particulier celles visant les systèmes de paiement –, sont reconnues comme difficiles à mettre en œuvre et conduisent généralement à des débats houleux.

- Depuis la contribution de Hennig-Schmidt et al., (2011) sur l'analyse des incitations des médecins en utilisant des méthodes expérimentales, de plus en plus de chercheurs s'intéressent à la distribution des facteurs de motivation des médecins. Les éléments tels que l'altruisme, le professionnalisme et la réputation, etc., sont ainsi de plus en plus étudiés à l'aide des méthodes issues de l'économie expérimentale. Ces éléments sont la plupart du temps mis en évidence dans les débats sociaux sur les comportements des médecins.

- Les améliorations en termes de technologies médicales sont de plus en plus évidentes. L'un des aspects les plus marquants de ces améliorations est probablement l'introduction de la médecine personnalisée. Cette dernière implique que l'on établisse le profil des patients pour déterminer les décisions, traitements ou interventions médicales en fonction de la meilleure réponse prévue. Antoñanzas et al. (2015) sont les premiers auteurs à avoir étudié l'économie de la médecine personnalisée. Les défis et les obstacles au développement de la médecine personnalisée apparaissent dans les communications publiques relatives à son développement. Cela inclut une réflexion approfondie sur une meilleure implication des patients ; toutes les questions de propriété et de sécurité des données ; la question centrale sur la qualité des technologies médicales utilisées et enfin le rôle que devrait jouer les incitations des professionnels de la santé dans l'adoption de tels outils.

Ce travail mobilise différentes méthodes pour étudier les incitations de différents systèmes de paiement en tenant compte des éléments contextuels décrits ci-dessus.

Dans le **Chapitre 1**, nous proposons une modélisation de type principal-agent pour étudier les propriétés des contrats optimaux à l'aide d'un cadre conceptuel qui caractérise les secteurs de l'éducation et de la santé. Les principaux "ingrédients" de notre modélisation théorique sont les suivants : (i) contraintes de responsabilité limitée, (ii) asymétrie de l'information (risque moral sur l'effort de l'agent et sélection adverse sur son niveau d'altruisme), (iii) possibilité d'accéder gratuitement à des technologies d'amélioration de l'effort, considérées ici comme étant divisible et explicitement formulable dans un contrat. Nous utilisons un système salarial comme méthode de paiement, le régulateur étant en charge de la conception du contrat.

Deux aspects méritent d'être mentionnés pour souligner notre contribution : Premièrement, dans notre analyse, le régulateur et l'agent produisent conjointement un résultat non "contractible" qu'est la santé ou l'éducation. L'agent apporte dans cette relation son "classique" effort personnel et professionnel, tandis que le régulateur contribue en fournissant un accès à des équipements pouvant améliorer la qualité de l'effort de l'agent (les techniques de médecine personnalisée sont un exemple dans le secteur de la santé). Deuxièmement, le régulateur fait face à une population hétérogène d'agents, des altruistes et des égoïstes. Les avantages altruistes des agents ne rentrent pas en compte dans leurs contraintes de participation, ceci afin d'exclure le fait que la satisfaction altruiste ne peut être obtenue que dans le cadre d'une relation principal-agent. Nous optons pour une version plus stricte de la contrainte de participation, appelée contrainte de responsabilité limitée, laquelle octroie à l'agent un paiement qui n'est pas inférieur au coût de ses efforts. Notre analyse suggère que le menu optimal de contrats qui maximise le bien-être devrait spécifier des salaires plus élevés pour les agents altruistes et un meilleur accès aux technologies d'amélioration de l'effort pour les agents égoïstes.

Ce chapitre améliore notre compréhension de l'interaction entre les trois facteurs décrits ci-dessus : les systèmes de paiement, l'importance des aspects éthiques dans la conception des contrats et le rôle des technologies d'amélioration de la qualité de l'effort. Le fait que nous modélisions la technologie d'amélioration de l'effort comme un équipement divisible et pouvant être formulé explicitement dans un contrat, rend le mécanisme que nous proposons un peu difficile à mettre en œuvre. En fait, l'adoption des technologies d'amélioration de l'effort, – un exemple étant la médecine personnalisée – dépend de nombreux autres facteurs, tel qu'indiqué plus haut : la modalité d'accès (gratuit ou payant) et les systèmes de paiement en place certes, mais aussi le niveau de précisions des tests médicaux personnalisés ; le rôle joué par les patients ; les conditions de confidentialité et de sécurité des données ; etc. Dans le chapitre 2, nous étudions différentes manières d'accéder aux technologies de la médecine personnalisée (accès gratuit *vs.* non-gratuit) et nous abordons également le rôle des systèmes de paiement des médecins dans l'adoption de ces technologies.

Le **Chapter 2** présente une expérience de laboratoire visant à étudier les incitations des médecins à adopter des techniques de médecine personnalisée. 95 futurs médecins constituent notre échantillon et nous les soumettons à un jeu à efforts réels. Dans ce jeu, leur contribution au diagnostic et au traitement du patient est simulée par une tâche de relecture sur des textes. L'accès à la médecine personnalisée est introduit en permettant au joueur de connaître les zones pertinentes où ses efforts de relecture ont plus de chance de porter au maximum les bénéfices que le patient pourrait en tirer. L'accès peut être gratuit ou payant dans notre jeu :

- Dans un contexte d'accès gratuit, tous les coûts liés à l'utilisation des techniques de médecine personnalisée lors de soins de routine sont à la charge de l'organisme de réglementation. La France utilise actuellement un tel système. L'autorité de santé finance 28 plateformes génétiques réparties géographiquement dans tout le pays et les médecins peuvent les utiliser pour leurs soins de routine sans aucuns frais pour eux.

- Dans la modalité d'accès payant, nous faisons l'hypothèse que le médecin supporte les coûts d'accès à la médecine personnalisée. Les coûts administratifs et de formation sont des exemples de ces coûts et devraient être pris en compte dans la conception des schémas de paiement.

Les propriétés incitatives de trois mécanismes de paiement sont étudiées : le paiement à l'acte (PA), le paiement par capitation (CAP) et le paiement à la performance (P4P). Nos analyses suggèrent que le système de paiement à la performance (P4P) (en comparaison au PA et au CAP) est associé à un taux élevé de choix de la médecine personnalisée payante. L'interaction entre l'altruisme et le retour sur investissement en P4P explique potentiellement cette différence. Cependant, quel que soit le système de paiement utilisé, nous remarquons que les médecins ont tendance à mieux utiliser les techniques de médecine personnalisée lorsqu'ils les acquièrent moyennant un coût. Nous avons qualifié ce phénomène de "dispositif d'engagement". Nos résultats portent donc le message de politique publique suivant : l'accès aux techniques de médecine personnalisée n'est pas optimal lorsqu'il est fourni gratuitement à tous les médecins ; On pourrait envisager plutôt un accès coûteux en accompagnant cela des initiatives de paiement à la performance pour booster l'adoption de techniques de médecine personnalisée. Ce rôle pivot du P4P est un résultat intéressant. Il existe en effet une littérature abondante évaluant les initiatives P4P et trouvant un effet incitatif modeste (lorsqu'il existe).

Dans le **Chapitre 3**, nous étudions la persistance des effets du système P4P. Il existe une littérature documentant le fait que les effets du P4P persistent dans la durée. Nous utilisons la même expérimentation du chapitre 2 dans laquelle le P4P est proposé avant ou après les systèmes de paiement plus classiques (PA et CAP dans notre cas). Notre cadre expérimental nous permet de comparer les médecins "traités" (ceux qui sont passés par un système P4P avant d'avoir un CAP ou un PA) à des médecins "contrôles" (ceux traitées avec un PA et un CAP avant d'avoir un P4P). En examinant un sous-ensemble de comportements, nous trouvons que les incitations du P4P ont un effet persistant mais plutôt ambigu. Dans le PA et le CAP, les médecins traités se concentrent davantage sur ce qui est pertinent pour le patient : ils font des efforts plus importants dans les zones les plus pertinentes pour traiter le patient. Dans le même temps toutefois, sous un CAP, les médecins traités atteignent des niveaux de qualité généralement inférieurs, ce qui est très probablement un signe que l'exposition au P4P détruit la motivation intrinsèque.

Au total, nous avons utilisé un cadre théorique et expérimental pour étudier les incitations et les motivations des médecins. Les trois principaux ingrédients retenus ont été : (i) les systèmes de rémunération des médecins ; (ii) la prise en compte des aspects éthiques décrivant les médecins (accent mis sur l'altruisme) ; (iii) l'étude de l'adoption des technologies d'amélioration de l'effort (accent mis sur la médecine personnalisée). Les deux premiers points ont fait l'objet d'une attention considérable dans des travaux antérieurs, à l'exception de l'étude de la sélection adverse sur l'altruisme, où le nombre d'articles théoriques est relativement limité. Le dernier point sur les technologies d'amélioration de l'effort, en particulier la médecine personnalisée, n'a pas fait l'objet de beaucoup d'attention dans les travaux antérieurs. Notre expérience de correction orthographique nous a permis d'analyser l'effet de différents modes de rémunération pour les médecins en présence de la médecine personnalisée. Nous avons cherché à comprendre si le mode de rémunération des médecins influe sur leur volonté d'acquérir des informations sur les besoins médicaux de leurs patients. Notre travail de doctorat nous permet de mieux comprendre comment les technologies de médecine personnalisée peuvent être fournies.

Avant de passer à nos différents chapitres, nous proposons un avant-propos qui décrit l'accès aux technologies de la médecine personnalisée en France. Le but de cette présentation est de donner plus de contexte au point (iii) sur la médecine personnalisée. Dans cette brève analyse, nous étudions les déterminants de l'accès des patients aux technologies de médecine personnalisée en France. Ce sera probablement l'analyse préliminaire à laquelle pourrait s'intéresser un régulateur. La relation entre la médecine personnalisée et les incitations offertes aux médecins sera étudiée plus en détail dans les chapitres 2 et 3, tandis que le Chapitre 1 proposera un cadre théorique beaucoup plus général sur les contrats optimaux applicable à ce contexte.

References

- Antoñanzas, F., Juárez-Castelló, C. A., & Rodríguez-Ibeas, R. (2015). Some economics on personalized and predictive medicine. The European journal of health economics, 16(9), 985-994.
- Ellis, R. P., & McGuire, T. G. (1986). Provider behavior under prospective reimbursement: Cost sharing and supply. Journal of health economics, 5(2), 129-151.
- Hennig-Schmidt, H., Selten, R., & Wiesen, D. (2011). How payment systems affect physicians' provision behaviour—an experimental investigation. Journal of Health Economics, 30(4), 637-646.
- World Health Organization. (2007). Everybody's business–strengthening health systems to improve health outcomes: WHO's framework for action.

Foreword: What determines access to personalized medicine in France.

Before thinking about payment systems and their incentives for personalized medicine use, a more straightforward question is whether patients benefit from it when it is available. France offers a unique context to study this question. In fact, in 2006, the French National Cancer Institute (INCa) funded 28 regional genetic centers designed to facilitate access to molecular profiling of cancer patients (Institut National du Cancer, 2014a). In this section, we describe the main determinants of French patient access to personalized medicine technologies.

Personalized medicine represents an opportunity to improve patients' outcomes by allowing physicians to use technological tools that determine whether patients are likely to benefit from particular treatments (Whitcomb, 2012). A potential barrier to personalized treatment relies on the access to genetic testing that must inform that treatment. In an effort to improve care outcomes, France undertook to make genetic testing routinely available to patients and physicians who treat them. In general, molecular profiling is particularly important for lung cancer patients because of the very high rates of genetic alterations in lung compared to other cancers (Alexandrov et al., 2013). Such profiling is generally the decision of the treating physician. In France, at least one molecular alteration was found in 43.2% of current or previous smokers' lung cancers and 74.8% of non-smokers' lung cancers. Guidelines were developed to ensure routine use of molecular profiling among lung cancer patients (Barlesi et al., 2016).

The INCa and the French Cooperative Thoracic Intergroup (IFCT) collected data from 28 regional centers to determine what kind of genetic mutations patients with advanced non-small-cell lung cancer (NSCLC)[1] had and what their clinical outcomes were (Barlesi et al., 2016).[2] We use that study to determine whether uptake of this technology varies according to different ecological factors that might influence local

[1] A cancer for which molecular profiling is recommended

[2] That study concluded that routine nationwide profiling is feasible and offers patients a clinical benefit, albeit at a "non-negligible financial cost".

use of genetic testing, such as socioeconomic status, the local supply of genetic testing centers, or the local supply of physicians.

To examine these relationships, we conducted an analysis of geographic variation in the rates of the French department-level use of genetic profiling for NSCLC and explored associations between those rates and department-specific ecological variables that might explain differences in utilization rates, with an eye toward understanding inequity of access.

Our data came from the following sources:

- From the *Biomarkers France study*, we collect data on patients diagnosed with advanced NSCLC who were referred by their physician for genetic testing between April 2012 and April 2013. This study sought to calculate the incidence and consequences of molecular alterations among patients with advanced NSCLC (Institut National du Cancer, 2014b). Hypothetically, all advanced NSCLC patients should have been identified because genetic profiling is recommended for their evaluation during routine care. During that time period, data from 15,814 unique patients with NSCLC patients are collected (Barlesi et al., 2016). Those data include a unique prescribing physician identifier that indicates the department in which the physician who ordered the genetic test worked.[3] In France, patients are not restricted to using healthcare services in the department in which they live. We therefore estimated the number of tests provided to patients living in a given department. To do that, we assumed that patients who obtained these tests did so using the same in- and out-of-department patterns that patients who had been admitted for lung cancer did.

- From *Agence technique de l'information sur l'hospitalisation (ATIH)* , we obtained data on admissions that had a primary diagnosis for lung cancer (defined as ICD 10 codes C34) , during the same period; these data include both the department in which the patient lives and the department in which the patient is admitted. For each department, we determine where unique patients living in that department are admitted for lung cancer. We use the Dartmouth Atlas Project's indirect method and the department-level number of lung cancer admissions of males and females aged 20-99 in 2012-2013 to reallocate healthcare utilization in the department of residence of the patient (See the note below Figure 1 for an example of our calculations). We were then able to generate sex-adjusted rates of patients who received genetic testing per 100 lung cancer admissions for each department, with department-level reallocated tests utilization in the

[3] mainland France is divided into 94 administrative units called 'départements'; these administrative units are the basis for organization of most social services.

numerator and the department-level sex-specific population of lung cancer patients in the denominator. We exclude *Somme* (department 80) because there appear to be an error in data collection on the number of patients who had genetic tests done there. Therefore, for 93 departments in mainland France, we use established methods to calculate 4 common measures of geographic variation in the per-capita use of genetic testing: (1) the extreme ratio, (2) the inter-quartile ratio, (3) the coefficient of variation and (4) the systematic component of variation (SCV) (Weeks et al., 2016, 2014; Wennberg and Gittelsohn, 1973).

- From *ATIH, Institut National de la Statistique et des Etudes Economics (INSEE)* and *Système National d'Information Inter-Régimes de l'Assurance Maladie (SNIIRAM)*, we obtained 2 types of ecological variables that might influence the use of molecular testing. First, we hypothesized that the per-capita department-level overall use of the healthcare system or supply of healthcare resources that might be consumed in the diagnosis and treatment of NSCLC could influence testing utilization rates. Therefore, we obtained the overall per-capita hospitalization rate and the per-capita number of general practitioners, surgeons, oncologists, pathologists, and radiotherapists from national databases and included them in the modeling. We included radiotherapists because their supply might be an indicator of higher technology available within a particular department. We also included dummy variables to account for the presence of a referral cancer hospital and the presence of a genetic testing center in each department. Second, because several studies found that the socio-economic status of the patient is a prominent determinant of high quality cancer care (Lejeune et al., 2010; Woods et al., 2005) and type of care received by non-small cell lung cancer patients (Greenwald et al., 1998; Pollock and Vickers, 1998; Yorio et al., 2012; Jiang et al., 2017) , from the same sources we obtained department-level measures of local economic distress: the poverty rate (a dummy was created for departments with poverty rates superior to 15%), and the proportion of people receiving "Couverture Maladie Universelle Complémentaire" (CMU-C), a supplemental health insurance that is given only to those whose income is below a particular level. We provide results for patients aged 18-99 and for the specific group of patients aged 60 and older. The 60 and older had the large majority of lung cancer admissions (72.2%) and genetic tests (65.9%).

We used two methods to determine whether these ecological variables explained geographic differences in department-level sex-adjusted per-capita genetic testing utilization rates. First, we used Ordinary Least Square (OLS) regression analysis to model the relationship between sex- adjusted rates of genetic testing for NSCLC and the ecological factors that we considered. Second, we tested for spatial auto-correlation by calculating the Moran's I statistic. Since auto-correlation was evident (i.e., Moran's I was < 0.001), we used a spatial error-lag regression model (weighting departmental results using a Rook criterion for the contingency matrix). We modeled per-capita use of genetic testing as the dependent variable for all patients, and we performed a sensitivity analysis using only patients aged 60 and older (who had the large majority of lung cancer admissions (72.2%) and genetic tests (65.9%)). For each sample, a parsimonious version of the regression is given -with 10% as a criterion for the variables selection. We show our results that account for the correction of spatial auto-correlation.

1. In mainland France, between April 2012 and April 2013, for every 100 lung cancer admissions, 46.87 patients aged 20-99 (and 42.82 patients aged 60-99) obtained genetic testing for NSCLC (see Table 1). Rates of genetic testing per 100 lung cancer admissions ranged over 3-fold for both age groups: from 23.75 to 77.32 for patients aged 20-99 (and from 21.68 to 74.68 for older patients). Nièvre (department 58) had the lowest rates and Côtes-d'Amor (department 22) had the highest rates for both age groups. Extreme and inter-quartile ratios were similar for both age groups as were the coefficient of variation and systematic component of variation (which, being greater than 5, indicated a high degree of geographic variation) (McPherson et al., 1982).

2. Figure 1 provides a map showing quintiles of rates of genetic test rates use for NSCLC among those aged 20-99 (left) and those aged 60-99 (right). For both age groups, rates were generally lower for department in the Champagne-Ardenne-Lorraine and Languedoc-Roussillon regions and in central France.

3. Our spatial regression models indicated that the per-capita supply of surgeons, general practitioners and radiotherapists were most strongly (the former negatively so) associated with use of genetic testing (Table 2). We also found that neither the dummy "living in a department with a genetic testing center" nor the dummy "living in a department possessing a referral cancer hospital" was associated with departmental use of genetic testing. We also found that the local poverty rate was negatively associated with utilization rates: For the 20-99 population of patients, deprived departments are associated with a 10% lower proportion of use of genetic testing technologies over the period (this proportion is 8% for the 60-99).

Table 1 – Measures of geographic variation in rates (per 100 lung cancer admission aged 20-99 or 60-99) of use of molecular profiling in France, April 2012 - April 2013.

		Age 20-99	**Age 60-99**
National rate		46.87	42.82
	Minimum rate	23.75	21.68
	Maximum rate	77.32	74.68
Extreme ratio		3.25	3.43
Inter-quartile ratio		1.40	1.44
Standard deviation		12.08	11.89
Coefficient of variation		0.25	0.27
Systematic component of variation x 10		5.40	6.02

Table 2 – Results of the regression analyses.

	Spatial regressions			
Ages of population included and models	**20-99**	**20-99 (parsimonious model)**	**60-99**	**60-99 (parsimonious model)**
Poverty rate(dummy w. ref = rate >15%)	-7.54** (3.68)	-9.91*** (3.09)	-6.86* (3.58)	-8.64*** (3.03)
Per capita supply of...				
General practitioner	0.11** (0.05)	0.08** (0.04)	0.11** (0.05)	0.08** (0.04)
Surgeons	-1.75* (1.01)	-1.96** (0.91)	-1.84* (0.98)	-2.24** (0.93)
Radiotherapists	6.47* (3.93)	6.59* (3.65)	7.75** (3.82)	8.12** (3.55)
Pathologists	-3.40 (2.48)		-3.23 (2.41)	
Oncologists	0.95 (4.08)		0.77 (3.97)	
Beds	0.10 (0.11)		0.19 (0.10)	
Per-capita admission rate	-1.13 (0.90)		-1.56* (0.87)	-0.80 (0.65)
Presence of a genetic testing center	2.20 (4.14)		1.52 (4.03)	
Presence of a referral cancer hospital	-2.35 (4.19)		-2.40 (4.08)	
Proportion receiving CMUC	-0.48 (0.49)		-0.38 (0.49)	
Constant	60.97*** (14.95)	45.19*** (2.85)	62.16*** (14.52)	54.24*** (11.08)
Observations	93	93	93	93
Log Likelihood	-346.38	-348.23	-34.71	-345.27
sigma2	96.11	100.36	90.77	94.38
Akaike Inf. Crit.	720.77	710.46	715.42	706.53
Wald Test (df = 1)	13.07***	11.78***	12.90***	10.94***
LR Test (df = 1)	7.41***	8.14***	7.32***	7.21***

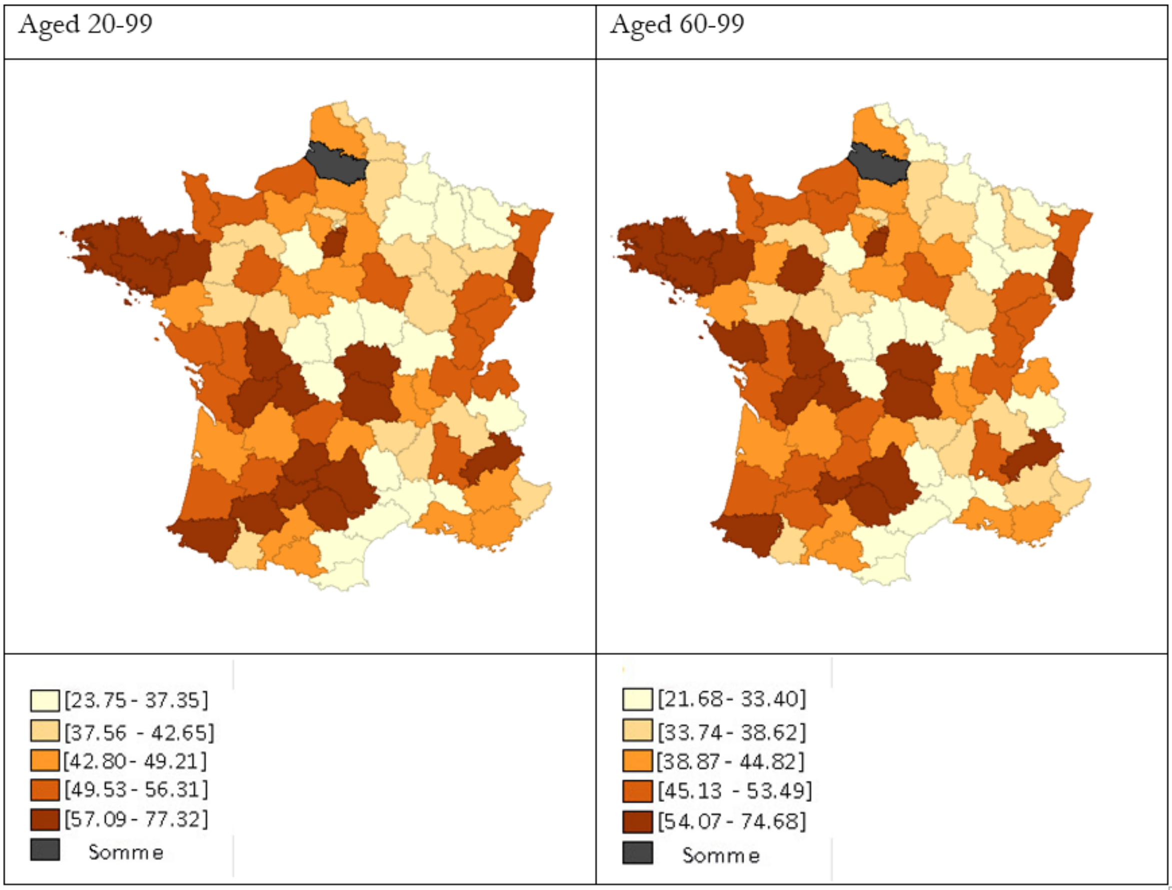


Figure 1 – Department-level quintiles of rates of genetic testing for NSCLC in France among inhabitants aged 20-99 (left) and those aged 60-99 (right), April 2012 – April 2013. Legends indicate the range of rates in each quintile.

Note: More details on the Dartmouth indirect method (Dartmouth Institute, 2018)

For each department, we know where unique patients living in that department were admitted for lung cancer. Using that information, we calculated the department-specific proportion of hospital stays (for males and females, separately) that were provided to patients who lived in that department and in any other department. For instance, during the study's period, among males, there were 68 lung cancer admissions in Loir-et-Cher (department 41): 96% of those admissions were for patients who lived in Loir-et-Cher, but 2.5% were for patients who lived in Indre-et-Loire (department 36) and 1.5% were for patients who lived in Loiret (department 45). To estimate the number of genetic tests done on patients who lived in a particular department, we then allocated tests obtained in a department according to how patients had been admitted for lung cancer. Therefore, continuing our example, we allocated the 30 genetic tests that were ordered on males by physicians working in Loir-et-Cher accordingly: 28.78 (96%) to Loir-et-Cher, 0.77 (2.5%) to Indre-et-Loire, and 0.44 (1.5%) to Loiret. We then added all allocated tests expected to have been received by males and females, separately, who lived in each department.

We find substantial variations across departments and several correlates with ecological variables. Rates of use of personalized medicine technologies were affected by the supply of health professional as well as the deprivation of the living area of the patient. We were initially surprised to discover an inverse relationship between the per-capita supply of surgeons and the use of genetic testing; however, it is possible that surgeons influence the therapeutic choice in favor of a rapid surgical intervention and then use genetic testing less frequently. A higher per-capita supply of radiotherapists was perhaps reflecting a greater overall supply of advanced cancer healthcare services in the local setting. However, the fact that the presence of a genetic testing center or a referral cancer hospital in the department was not a statistically significant predictor of genetic testing rates provides an interesting result. It actually tends to validate the territorial grid of the genetic centers and reference cancer hospitals across France and their effective communication with the decentralized hospitals.

We also find that patients living in high poverty departments were less likely to receive genetic testing after correcting for other explanatory factors. This inequality of access observed is an issue for the French healthcare system which claims to provide free and equitable access to care for all cancer patients. There are recent US studies that have documented the link between NSCLC patients' place of residence and their access to treatments: (Yorio et al., 2012; Jiang et al. 2017). Yorio et al., (2012) showed in a study done within a single academic medical center in Texas that socioeconomically disadvantaged patients with stage I-III NSCLC were less likely to receive "standard" therapy; while Jiang et al., (2017) showed that Nebraska NSCLC patients residing in high poverty neighborhoods were twice less likely to receive surgery than those in low poverty neighborhoods. In our study, we complement this work by giving evidence that access to personalized medicine for NSCLC patients is influenced by the social gradient of the department in which the patient lives. Although French authorities determined that routine nationwide genetic profiling is feasible, our findings suggest that it is currently inequitable and that a focus on departments with high poverty levels would reduce that inequity.

To conclude, we recognize that this analysis has several limitations. First, through the reallocation process, we used administrative data for lung cancer admissions from 2012-2013 to estimate where patients who obtained genetic testing lived. Patients might use different healthcare utilization patterns for genetic testing and hospitalization for lung cancer, and future studies should collect data on patients' residence to more accurately evaluate their access to genetic testing. Second, we were not able to observe the precise proportion of advanced non-small cell lung cancer among the total lung cancer in each department, which would be a better denominator for utilization rates (but we believe that the expected differences across departments in this proportion is weak and cannot explain such high variations in utilization rates). Finally, use of genetic testing for advanced NSCLC in 2012-2013 might not reflect current utilization patterns; there is hope that the equality in access has improved in the recent years (Nay et al., 2016).

Nonetheless, this study suggests that department economic distress might negatively impact routine use of genetic testing. French policymakers should target deprived areas to provide equal access to personalized medicine for advanced NSCLC patients. While we found that genetic testing was done in all departments, future work should explore ways to reduce inequities in the use of genetic tests and seek to better explain the geographic variation in rates that we found.

References

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.-L., 2013. Signatures of mutational processes in human cancer. Nature 500, 415–421.

ATIH, 2012. Agence technique de l'information sur l'hospitalisation `http://www.atih.sante.fr/` (accessed 1.22.18).

Barlesi, F., Mazieres, J., Merlio, J.-P., Debieuvre, D., Mosser, J., Lena, H., Ouafik, L., Besse, B., Rouquette, I., Westeel, V., 2016. Routine molecular profiling of patients with advanced non-small-cell lung cancer: results of a 1-year nationwide programme of the French Cooperative Thoracic Intergroup (IFCT). The Lancet 387, 1415–1426.

Dartmouth Institute, 2018. An Example of Indirect Adjustment. Indirect Adjust. Dartm. Atlas Health Care. `http://www.dartmouthatlas.org/downloads/methods/indirect_adjustment.pdf` (accessed 1.22.18).

Greenwald, H.P., Polissar, N.L., Borgatta, E.F., McCorkle, R., Goodman, G., 1998. Social factors, treatment, and survival in early-stage non-small cell lung cancer. Am. J. Public Health 88, 1681–1684.

ICD-10 CM Codes, 2015. 2018 ICD-10-CM Codes C34*: Malignant neoplasm of bronchus and lung [WWW Document]. URL `http://www.icd10data.com/ICD10CM/Codes/C00-D49/C30-C39/C34` (accessed 1.22.18).

Institut National du Cancer, 2014a. Plateformes hospitalières de génétique moléculaire des cancers: faits marquants et synthèse d'activité 2013. (Ouvrage collectif édité par l'INCa), Collection Bilans d'activité et d'évaluation. Collection Bilans d'activité et d'évaluation, Boulogne-Billancourt,.

Institut National du Cancer, 2014b. Synthèse de l'activité des plateformes hospitalières de génétique moléculaire des cancers en 2012, en vue d'optimiser leur évolution (Collection Bilans d'activité et d'évaluation), ouvrage collectif édité par l'INCa. Institut National du Cancer, Boulogne-Billancourt.

Jiang, X., Lin, G., Islam, K.M.M., 2017. Socioeconomic factors related to surgical treatment for localized, non-small cell lung cancer. Soc. Sci. Med. 175, 52–57. https://doi.org/10.1016/j.socscimed.2016.12.042

Lejeune, C., Sassi, F., Ellis, L., Godward, S., Mak, V., Day, M., Rachet, B., 2010. Socio-economic disparities in access to treatment and their impact on colorectal cancer survival. Int. J. Epidemiol. 39, 710–717.

McPherson, K., Wennberg, J.E., Hovind, O.B., Clifford, P., 1982. Small-area variations in the use of common surgical procedures: an international comparison of New England, England, and Norway. N. Engl. J. Med. 307, 1310–1314.

Nay, O., Béjean, S., Benamouzig, D., Bergeron, H., Castel, P., Ventelou, B., 2016. Achieving universal health coverage in France: policy reforms and the challenge of inequalities. Lancet Lond. Engl. 387, 2236–2249. https://doi.org/10.1016/S0140-6736(16)00580-8

Pollock, A.M., Vickers, N., 1998. Deprivation and emergency admissions for cancers of colorectum, lung, and breast in south east England: ecological study. BMJ 317, 245–252. https://doi.org/10.1136/bmj.317.7153.245

Populations légales 2013 – Populations légales 2013 | Insee, 2013. https://www.insee.fr/fr/statistiques/2119468?sommaire=2119504 (accessed 1.22.18).

SNIIRAM, 2016. - Sniiram. Système Natl. D'Information Interrégimes L'Assurance Mal.https://www.ameli.fr/l-assurance-maladie/statistiques-et-publications/sniiram/finalites-du-sniiram.php (accessed 2.6.18).

Weeks, W.B., Jardin, M., Dufour, J.-C., Paraponaris, A., Ventelou, B., 2014. Geographic variation in admissions for knee replacement, hip replacement, and hip fracture in France: evidence of supplier-induced demand in for-profit and not-for-profit hospitals. Med. Care 52, 909–917.

Weeks, W.B., Ventelou, B., Paraponaris, A., 2016. Rates of admission for ambulatory care sensitive conditions in France in 2009–2010: trends, geographic variation, costs, and an international comparison. Eur. J. Health Econ. 17, 453–470.

Wennberg, J., Gittelsohn, A., 1973. Small area variations in health care delivery: a population-based health information system can guide planning and regulatory decision-making. Science 182, 1102–1108.

Whitcomb, D.C., 2012. What is personalized medicine and what should it replace? Nat. Rev. Gastroenterol. Hepatol. 9, 418–424. https://doi.org/10.1038/nrgastro.2012.100

Woods, L.M., Rachet, B., Coleman, M.P., 2005. Origins of socio-economic inequalities in cancer survival: a review. Ann. Oncol. 17, 5–19.

Yorio, J.T., Yan, J., Xie, Y., Gerber, D.E., 2012. Socioeconomic Disparities in Lung Cancer Treatment and Outcomes Persist Within a Single Academic Medical Center. Clin. Lung Cancer 13, 448–457. https://doi.org/10.1016/j.cllc.2012.03.002

Chapter 1

Regulation and altruism

This chapter is based on a joint research with Izabela Jelovac. The first draft of this work was my master dissertation that I wrote under Izabela's supervision in winter and spring 2015. I proposed the thematic and received guidance from Izabela. When I got funding for the PhD, we continued working together on this paper until it was published in March 2019 in the Journal of Public Economic Theory. The current version is the latest that was published with very minor revisions added in italic to reflect the comments made during the Pre-defense.

We study optimal contracts in a regulator-agent setting with joint production, altruistic and selfish agents, limited liability and uneasy outcome measurement. Such a setting represents sectors of activities such as education and health care provision. The agents and the regulator jointly produce an outcome for which they all care to some extent that is varying from agent to agent. Some agents, the altruistic ones, care more than the regulator does while others, the selfish agents, care less. Moral hazard is present due to both the agent's effort and the joint outcome that are not contractible. Adverse selection is present too since the regulator cannot a priori distinguish between altruistic and selfish agents. Contracts consist of a simple transfer from the regulator to the agents together with the regulator's input in the joint production. We show that, under the conditions of our setting and when we face both moral hazard and adverse selection, the regulator maximizes welfare with a menu of contracts, which specify higher transfers for the altruistic agents and higher regulator's inputs for the selfish agents.

Keywords: altruism, moral hazard, adverse selection, regulator-agent joint production.

JEL Classification: D64, D86.

1.1 Introduction

In sectors such as education and health care, the measurement of results is uneasy and thus results hardly influence payments or rewards. Moreover, the outcomes in such sectors often depend on the contributions of both the agent and the regulator. The agent can be a teacher or a health care provider and she contributes with her professional effort. The regulator represents the collectivity and he provides inputs such as computers, classrooms, universities, hospitals, medical technologies,[1] etc. Also, the agent and the regulator happen to share the same objectives, at least partially. A teacher cares for the quality of the education and a physician cares for the quality of health care. The agent and the regulator care for these outcomes to some extent, which is varying from one agent to another. If an agent cares more than the regulator does, we call her altruistic. Instead, if she cares less than the regulator does, we call her selfish. This heterogeneity in agent's altruism is in line with the empirical evidence reported by Watt *et al.* (2014) for teachers and by Lagarde and Blaauw (2014) and Brosig-Koch *et al.* (2017) for physicians. In our setting, the concepts of altruism, mission oriented behavior, professional ethics or public service motivation are actually equivalent.

In this paper, we analyze optimal contracts in a principal-agent setting with limited liability that reflects the characteristics of such sectors. The principal (he) is a regulator who cares for some outcome jointly produced with the agent (she). The agent's effort and the outcome are not contractible. Moreover, the agent has private information on her altruism level, that is, on the extent to which she shares the regulator's concern for the jointly produced outcome. Therefore, contracts consist of a transfer from the regulator to the agent and of a regulator's input in the joint production. To sum up, we face a problem of moral hazard together with adverse selection.

We derive and analyze the optimal contract in three regimes. We refer to first-, second- and third-best contracts, respectively, without any agency problem, with moral hazard only and contracts with moral hazard *cum* adverse selection. We proceed in this way because it enables to study progressively how the information asymmetry impacts on the regulator-agent relationship. With the first-best contract, regulator's input, effort and transfer do not vary with the altruism level and they are decreasing in the shadow cost of public funds. Altruism does not affect the first-best solution because the latter satisfies a binding agent's limited liability constraint, which does not include her altruistic benefits. As long as the solution satisfies the limited liability constraint, it also satisfies the participation constraint, which adds to

[1] *Medical technologies can cover elements such as access to personalized medicine techniques*

the former the altruistic component. The first-best contract would still be optimal if the only agency problem was adverse selection. With the second-best contract, we show that for selfish agents, effort and transfer increase with altruism and the regulator's input is invariant to changes in altruism. The regulator's input serves as an incentive for a selfish agent's effort and the transfer exactly compensates the selfish agent for her personal cost of effort. For the altruistic agents, the second-best solution coincides with the first-best one. The regulator uses the transfer to restrict the altruistic agent's effort to the first-best one. Higher levels of agent's effort would be too costly to the regulator because of limited liability. The regulator's input in turn plays no role as an incentive for effort, when we consider the second-best contract for an altruistic agent. Even so, the effort of an altruistic agent is higher than the one of a selfish agent, and the second-best transfer is thus higher for an altruistic agent. If the regulator proposes the second-best contracts to the agents without being able to distinguish them according to their degree of altruism, then the selfish agents would pretend to be altruistic to enjoy the higher transfer designed for altruistic agents. To avoid such a selection issue, we turn to the third-best analysis and we derive separating contracts to maximize welfare. These third-best contracts specify a higher transfer for the altruistic agents and a higher regulator's input for the selfish agent.

An important driver of our results is our modeling of the agent's participation and limited liability constraints. The difference between the participation constraint and the limited liability constraint is the altruistic part of the agent's utility. A classical participation constraint grants the agent a utility that is not lower than a reservation utility, which we normalize to zero. This classical participation constraint includes the altruistic component of the agent's utility. Such a participation constraint appears in the analysis by Jack (2005). It implicitly assumes that the altruistic satisfaction of contributing to the outcome can only be obtained within the principal-agent relationship. It also implicitly assumes that an altruistic agent is ready to pay to enter the principal-agent relationship. We rather opt for a stronger version of the participation constraint. To avoid any confusion, we call it a limited liability constraint and it grants the agent a payment that is not lower than her personal cost of effort. It is equivalent to the minimum profit constraint in Chone and Ma (2011). The only difference with a classical participation constraint is that it does not include the altruistic part of the utility. By construction, it implies the classical participation constraint. We can interpret this limited liability constraint as a legal constraint according to the idea that work must pay. We can also interpret it as a participation constraint if we consider that the altruistic part of the utility enters not only the agent's utility within the principal-agent relationship, but also her reservation utility. This approach is valid if the agent can contribute to the outcome outside of the principal-agent relationship. In this case, individual rationality would make her volunteer rather than pay for working. We show that without any limited liability constraint, the third-best contracts would specify (weakly) higher transfers for the selfish agents and (weakly) higher regulator's inputs for the altruistic agents. This result about transfers reflects the opportunity for a regulator to use job satisfaction rather than payment when agents are altruistic and no limited liability is granted. This is in line with Jack (2005) who derives self-selecting contracts with higher payments for the selfish agents than for the altruistic ones. In the present paper instead, because of the limited liability constraints, the transfer must be higher for the agents exerting higher levels of effort, that is, the altruistic ones. With adverse selection, the regulator compensates this difference between payments for altruistic and selfish agents by a difference in regulator's inputs, this time in favor of the selfish agents. A higher regulator's input for selfish agents will act as an incentive for their effort. Instead, the regulator's input for the altruistic agents plays no role as an incentive for effort since their effort, which is costly to the principal because of the limited liability, is restricted by the level of transfer.

The rest of the paper is organized as follows. In the following subsection, we describe the related literature. In section 2, we set up the model. In Section 3, we analyze optimal contracts when there is no asymmetry of information. In section 4, we analyze optimal contracts with a focus on moral hazard: the agents' effort is not contractible. In section 5, we derive optimal contracts with a focus on both moral hazard and adverse selection: the agent has private information on her altruism and his effort is not contractible. We conclude in section 6. Proofs are in Appendix.

1.1.1 Related literature

The approach our paper uses to capture the issue of regulation and altruism is to consider that the regulator and the agent participate in a joint production with private information on the agent's altruism and non-contractibility of both effort and outcome. In this literature review, we discuss how different is our model compared to classical principal-agent models used in the literature.

The first strand is the literature on joint production and productivity enhancing investments (Dor and Watson (1995); Jelovac and Macho-Stadler (2002); Canidio and Gall (2019)). In a recent paper, Canidio and Gall (2019) studied how the principal decides on the provision of non-monetary rewards (such as perks) in a model that seeks to link agents' career concern and their incentives. Perks are used both as a form of remuneration and a way to affect agent's optimal task choice. The paper compares to ours, as perks play in their setting, the same role as our principal input. Both capture productivity enhancing investments made by the principal. Jelovac and Macho-Stadler (2002) consider a model in which two agents, a hospital and a physician, both participate in the production of health services. The agency problem the paper deals with is two-sided moral hazard. There is moral hazard on the physician's effort and on the hospital's investment. The aim of the paper is to analyze and compare the performance of two possible organizations for health services: a centralized structure in which the regulator contracts with both agents (hospital and physician) and a decentralized structure in which he contracts with the hospital only, delegating to this latter the authority to contract with physicians. Unlike Jelovac and Macho-Stadler however, we account for altruism by integrating it in the objective of the agent. While their paper deals with two sided moral hazard on the two inputs of the joint production function, our model rather considers moral hazard on one input and adverse selection on the altruism parameter that enters the agent's objective function.

The second strand of the literature to which our paper relates is the literature on adverse selection on altruism. Many papers in the literature consider moral hazard with altruism. However, heterogeneity in altruism is relatively new and scarce in the literature (see Macho-Stadler and Perez-Castrillo (2016)'s discussions on extensions of moral hazard models and Hoszegi (2014)'s review on behavioral contract theory). Experimental studies report that agents differ in their levels of altruism. For example, Godager and Weisen (2013) report on heterogeneity in physician altruism while Brosig-Koch *et al.* (2017) report heterogeneity in altruism among medical and non-medical subjects. Some other theoretical works integrate the altruistic component in their analysis without explicitly considering its heterogeneity in the population (see Biglaiser and Ma (2007) and Naegelen and Mougeot (2011) among others).

Theoretical papers to which our study directly relates are Jack (2005), Chone and Ma (2011), Liu and Ma (2013) and Barigozzi and Burani (2016). Jack (2005) studies optimal contracts under both adverse selection on altruism and moral hazard on agent's effort. He studies how the regulator can use the cost-sharing parameter and the transfer to deal with these two informational problems. Chone and Ma (2011) consider a physician-consumer relationship with heterogeneity on both altruism and patient's benefit. They show that, without heterogeneity on consumer's benefit, the regulator can impose the first-best quantity even under unknown altruism. Our analysis confirms Chone and Ma (2011)'s results if our setting accounted for unknown altruism only. In Liu and Ma (2013)'s setting, the physician-consumer relationship also takes into account commitment, risk aversion and insurance. They show that the first best (same transfer regardless of the provider's altruism) can be implemented in a "commitment game" where the physician commits to treatment decisions at contracting. Screening appears only in the non-commitment game where more altruistic providers earn positive profits. Barigozzi and Burani (2016) analyze optimal contracts between a hospital and health professionals privately informed about their ability and altruism. They introduce difference in ownership structures as well as mission of hospitals. The first big difference between our paper and these four papers is that we have a joint production framework. In our framework, the regulator and the agent both contribute to the non-contractible outcome. The second difference is the fact that we do not assume that the variable under the agent's control is contractible. Assuming that the regulator can neither contract on agent's effort nor on the produced outcome makes our analysis different from what has been done in earlier work in the literature using principal-agent frameworks. Therefore, our results really complement the results of all these papers with this alternative setting that recognizes that outcome cannot be accurately measured, liability is limited and production is joined.

1.2 Model

A regulator contracts with an agent to jointly produce a non-negative outcome $S(a, e)$ such as health or education. We consider that the regulator and the agent contribute to this outcome with non-negative and costly input a and effort e, respectively. The regulator's input can take the form of hospitals or computing facilities, for example, while the agent's effort is a professional effort. The agent always observes the regulator's input a while the regulator does not observe the agent's effort e. The setting so far is very similar to the one in Jelovac and Macho-Stadler (2002).

Both the regulator and the agent care for the outcome $S(a, e)$ but not to the same extent. The agent weights her utility from the outcome $S(a, e)$ by a non-negative parameter θ. An agent with a high θ is more altruistic or more mission-oriented than an agent with a low θ.

We reasonably assume that the outcome function $S(a, e)$ is increasing in both input and effort and it is concave: $S_a > 0$, $S_e > 0$, $S_{aa} < 0$, $S_{ee} < 0$ and $S_{aa}S_{ee} - (S_{ae})^2 > 0$. We also assume that input and effort are complementary in the production of the outcome S: $S_{ae} \geq 0$. To derive some of our results, it is useful to approximate the ratio S_{ae}/S_{ee} by a negative constant: $S_{ae}/S_{ee} = -k$, with $k > 0$.[2] The unit cost of input a and effort e are c and v, respectively. We denote by λ the shadow cost of public funds ($\lambda > 0$).

[2]That is, $S(a, e)$ can be any function of the type $S(a, e) = g(a) + h(e) + kae$, with $g_a > 0$, $h_e > 0$, $g_{aa} < 0$, $h_{ee} = -1$ and $k > 0$. The following quadratic function is compatible with our assumptions: $S(a, e) = \gamma_1 a + \gamma_2 e - \frac{1}{2}[a^2 + e^2 - 2kae]$.

The payment from the regulator to the agent is a simple transfer t. We rule out more sophisticated incentive payments because we consider that outcome and agent's effort cannot be accurately measured. Accordingly, we define the agent's utility and the regulator's welfare as follows:

$$U = \theta S(a, e) + t - ve; \tag{1.1}$$

$$W = S(a, e) - (1 + \lambda)(t + ca) + t - ve. \tag{1.2}$$

Equation (1.1) is the agent's utility function. It consists of her transfer t net of personal cost of effort ve, plus her direct benefit from the joint outcome $\theta S(a, e)$. Equation (1.2) is the regulator's welfare function. It is written as the sum of the surplus $S(a, e)$ and the provider's financial surplus net of regulator's direct costs. Note that to avoid double counting, the regulator does not take into the welfare function the direct benefit from the outcome derived by the agent. Equation (2) also reflects that the regulator weights the utility of more altruistic agents the same as that of more selfish agents. Writing Equations (1) and (2) as such illustrates that some agents (low θ) care less than the regulator does, while others (high θ) care more.

1.3 First-best contract

We derive in this section the optimal contract if the regulator can observe and contract upon the agent's effort e. It serves as a benchmark for our analysis.

The regulator chooses both the agent's effort e and his own input a, as well as the level of compensation t for the agent, so as to maximize welfare. He must ensure that the agent is willing and able to participate. We consider that the agent is willing to sign the contract only if her resulting utility is at least her reservation utility, normalized to zero for simplicity: $\theta S(a, e) + t - ve \geq 0$. However, the agent participates only if her limited liability allows her to: $t - ve \geq 0$. In our model, the agent's limited liability constraint always implies her participation constraint. Accordingly, our first-best problem is the following one:

$$\max_{a,e,t} W = S(a, e) - (1 + \lambda)(t + ca) + t - ve \tag{P1}$$

s.t.

$$t - ve \geq 0. \tag{LLC}$$

which is equivalent to:

$$\max_{a,e} W = S(a, e) - (1 + \lambda)(ve + ca), \tag{P1'}$$

with the limited liability constraint (LLC) binding: $t = ve$.

The first-best input a^{FB} and effort e^{FB} are given by the first-order conditions for an interior solution and we summarize the first-best solution in the following lemma.

Lemma 1. ***First-best solution analysis***

1. *The first-best solution (a^{FB}, e^{FB}, t^{FB}) is given by:*
   - $S_a(a^{FB}, e^{FB}) = (1 + \lambda)c$
   - $S_e(a^{FB}, e^{FB}) = (1 + \lambda)v$
   - $t^{FB} = ve^{FB}$
2. *a^{FB}, e^{FB} and t^{FB} are decreasing in the shadow cost of public funds.*
3. *a^{FB}, e^{FB} and t^{FB} are invariant to changes in altruism θ.*

The conditions defining the first-best input and effort reflect the traditional equality between marginal utility and marginal cost. The first-best transfer in Lemma 1 is given by the agent's limited liability constraint. This constraint binds so that the first-best transfer exactly compensates the disutility of first-best effort. The agent earns no rent. Lemma 1 also states that, first-best input, effort and transfer decrease with the shadow cost of public funds. This is intuitive since a higher burden on public expenses increases the marginal cost of effort and input. The variation in transfer just compensates for the variation in effort due to a change in the shadow cost of public funds. The first-best solution is invariant to the level of altruism of the agent because the altruism parameter θ appears neither in the limited liability constraint nor in the welfare function. Then, if the problem was only adverse selection (unknown altruism) without moral hazard (non-contractible effort), the first-best contract would be optimal and bunching. This is in line with the result of Chone and Ma (2011) that, without heterogeneity on consumer's benefit, the regulator can impose the first-best quantity even under unknown altruism.

1.4 Second-best contract

The first-best solution does not account for situations in which, the agent's decision is not a contractible variable. We consider in the present section optimal contracts in a second-best regime where the regulator cannot contract upon either the agent's endogenous effort or the joint outcome.[3] This is a case of moral hazard and the

[3] If instead, the outcome S was contractible, then effort e could be recovered from the level of $S(a, e)$. If such a level of effort was contractible too, then we would not have a proper moral hazard problem

regulator anticipates that the agent can behave strategically if she is proposed the contract.

This situation is similar to a game in which the agent moves after the regulator and determines her optimal level of effort in stage 2 and the regulator takes into account this additional incentive constraint when he decides the transfer and his own level of input in stage 1.

The agent's problem is to choose the level of effort which maximizes her utility function (1) subject to her limited liability constraint (LLC). The solution to this constrained maximization of the agent's utility is either interior or constrained by (LLC) in case transfer t is too low. Formally, this solution implicitly defines the best-reply effort of the agent, $e = e^{br}(a, t, \theta)$, which is given by:

$$e = e^{br}(a, t, \theta) = Min\left\{e^*(a, \theta), \frac{t}{v}\right\} \tag{ICC}$$

where

$$e^*(a, \theta) \text{ is implicitely defined by } \theta S_e(a, e^*(a, \theta)) - v = 0. \tag{1.3}$$

If transfer t is high enough ($t \geq ve^*(a, \theta)$), then the solution is interior and the agent exerts effort $e = e^*(a, \theta)$ so that the marginal cost, v, equates the marginal utility, $\theta S_e(e^*, a)$, she derives from it. We notice at this stage that this effort increases with input a. The regulator's input a can thus be used as an incentive for an agent's affort. It is also increasing in altruism. These relationships are all very intuitive and formally, they are given by:

$$\frac{\partial e^*}{\partial a} = -\frac{S_{ae}}{S_{ee}} = k \geq 0; \tag{1.4}$$

$$\frac{\partial e^*}{\partial \theta} = -\frac{S_e}{\theta S_{ee}} = -\frac{v}{\theta^2 S_{ee}} > 0. \tag{1.5}$$

If instead, the transfer t is not high enough ($t \leq ve^*(a, \theta)$), then we have a corner solution because (LLC) binds and the agent's decision on effort is constrained so as not to make losses: $e = \frac{t}{v}$.

The regulator's problem at stage 1 is to maximize the welfare function subject to the agent's limited liability constraint (LLC) and the incentive compatibility con-

and the solution to the whole problem would be straightforward. If instead, such a level of effort was not contractible, then we could make the transfer t depend on whether outcome S is high or low, with effort (and possibly the input a) influencing the probability of having a high versus a low value for the outcome S. Such a setting would be closer to the standard setting (see chapter 7 in Laffont and Martimort, 2002) and informational rents would appear with moral hazard, uncertainty and limited liability all together. However, informational rents would be lower due to altruism.

straint (ICC). The problem of the regulator is given by:

$$\max_{a,t} W = S(a,e) - (1+\lambda)(t+ca) + t - ve \tag{P2}$$

s.t

$$t - ve \geq 0; \tag{LLC}$$

$$e = e^{br}(a,t,\theta). \tag{ICC}$$

We summarize in the following lemma the analysis of the second-best solution.

Lemma 2. ***Second-best solution analysis***

1. *When* $(1+\lambda)\theta < 1$,
   - *the second-best solution* (a^{SB}, e^{SB}, t^{SB}) *is given by:*
     - $a^{SB} = a^{FB}$
     - $S_e(a^{SB}, e^{SB}) = \frac{1}{\theta} v$
     - $t^{SB} = ve^{SB}$
   - *comparative statics:*
     - a^{SB}, e^{SB} *and* t^{SB} *are decreasing in the shadow cost of public funds*
     - a^{SB} *is invariant to changes in altruism*
     - e^{SB} *and* t^{SB} *are increasing in altruism*
2. *When* $(1+\lambda)\theta \geq 1$, *the second-best solution* (a^{SB}, e^{SB}, t^{SB}) *coincides with the first-best one* (a^{FB}, e^{FB}, t^{FB}).

When altruism and the shadow cost of public funds are low enough ($(1+\lambda)\theta < 1$), then the second-best effort increases with altruism since altruism increases the marginal benefit of effort to the agent without affecting the marginal cost of it. The second-best transfer also increases with altruism to compensate the agent for her higher effort. Input is invariant to changes in altruism and it always coincides with its first-best level. However, this double result is conditionned by the quadratic form of $S(a,e)$. In a more general framework, input a would vary with altruism.

Note that when $(1+\lambda)\theta = 1$, the agent is what we call a “perfect agent”. She intervenes in the production of S exactly as the regulator would have done if there were no delegation. When her concern for patient is stronger than a perfect agent’s one ($(1+\lambda)\theta > 1$), the regulator adjusts the transfer t so as to constrain the agent to exert the first-best level of effort. This is why the second-best and the first-best solution are equivalent when the agent is highly altruistic.

It is interesting to see how agent's incentives change the optimal payment, effort and input. For this purpose, we compare the first-best and the second-best solutions. We summarize the comparison in the following lemma:

Lemma 3. ***Comparison of the first-best and the second-best solutions***

- $a^{SB} = a^{FB}$
- $(e^{SB}, t^{SB}) < (e^{FB}, t^{FB})$ *if and only if* $(1 + \lambda)\theta < 1$
- $(e^{SB}, t^{SB}) = (e^{FB}, t^{FB})$ *if and only if* $(1 + \lambda)\theta \geq 1$

The comparison presented in Lemma 3 shows to what extent moral hazard can be an issue for the regulator. When altruism and the shadow cost of public funds are low ($(1 + \lambda)\theta < 1$), the level of effort chosen by the agent is distorted downward compared to the first-best contract. In fact, effort is increasing in altruism when decided by the agent (Lemma 2) while it is decreasing in the shadow cost of public funds when decided by the regulator (Lemma 1). This arises because the agent's marginal benefit from effort increases with altruism while the marginal regulator's cost of effort increases with the shadow cost of public funds when the regulator decides effort. Conversely, as we already mentionned, when the agent's altruism is high enough ($(1 + \lambda)\theta > 1$), the regulator adjusts the transfer t so as to constrain the agent to exert the first-best level of effort. So far, the assumption about the quadratic form of $S(a, e)$ is without loss of generality. However, such a quadratic form is the very reason why, in all cases, the regulator's input is the same no matter whether the agent's effort is contractible or not. This is indeed a limiting factor. Nevertheless, the quadratic form greatly simplifies the analysis without affecting the generality of the results, except for the persistent optimality of the first-best regulator's input so far.

Lemma 2 also states that, a more benevolent agent is given a (weakly) higher transfer to compensate for a higher effort:

$$\frac{\partial t^{SB}}{\partial \theta} \geq 0. \tag{1.6}$$

If the regulator proposes the second-best contract to an heterogeneous population of agents and is able to tell agents apart according to their altruism, then very altruistic agents ($(1 + \lambda)\theta \geq 1$) would work more and they would earn higher transfers than less altruistic agents ($(1 + \lambda)\theta < 1$). The regulator's input is the same for both types. If the regulator cannot distinguish between agents according to their altruism, less altruistic agents would pretend to be highly altruistic to earn a higher transfer. Therefore, the regulator can propose different contracts depending on the type of the agent (whether she is very altruistic or not). We study in the next section optimal contracts in a third-best regime where we have moral hazard and heterogeneity among agents according to their concerns for the outcome.

1.5 Third-best contracts

We now consider that the regulator contracts with a population of agents who differ in their level of altruism and that he cannot tell the agents apart according to their altruism. To keep the analysis simple and interesting, we consider two different types of agents: altruistic agents who are more altruistic than a regulator's "perfect agent" ($\theta = \theta_1 > \frac{1}{1+\lambda}$) and selfish agents who are less altruistic than a "perfect agent" ($\theta = \theta_0 < \frac{1}{1+\lambda}$). We also refer to type 1 and type 0 to distinguish between them.

In this section, we consider that screening is possible and we assume that the regulator chooses a level of input together with a level of transfer for each type of agent: (a_1, t_1) for the altruistic type of agent and (a_0, t_0) for the selfish type of agent. As is typical in adverse selection problems, the regulator tailors contracts so that each type of agent exactly selects the one made for her. We assume that the regulator knows by experience that there is a proportion α of type 1 and $1-\alpha$ of type 0 in the population of agents. The regulator also anticipates moral hazard. That is, the regulator expects that agent's effort is $e = e^{br}(a, t, \theta)$, as defined by (ICC). The regulator solves the following problem:

$$\begin{aligned}\max_{a_0,a_1,t_0,t_1} W =&\alpha \left\{S(a_1, e^{br}(a_1, t_1, \theta_1)) - (1+\lambda)(t_1 + ca_1) + t_1 - ve^{br}(a_1, t_1, \theta_1)\right\} + \\ &(1-\alpha)\left\{S(a_0, e^{br}(a_0, t_0, \theta_0)) - (1+\lambda)(t_0 + ca_0) + t_0 - ve^{br}(a_0, t_0, \theta_0)\right\}\end{aligned} \tag{P3}$$

s.t

$$t_0 - ve^{br}(a_0, t_0, \theta_0) \geq 0; \tag{LLC0}$$

$$t_1 - ve^{br}(a_1, t_1, \theta_1) \geq 0; \tag{LLC1}$$

$$\begin{aligned}&\theta_0 S(a_0, e^{br}(a_0, t_0, \theta_0)) + t_0 - ve^{br}(a_0, t_0, \theta_0) \geq \\ &\theta_0 S(a_1, e^{br}(a_1, t_1, \theta_0)) + t_1 - ve^{br}(a_1, t_1, \theta_0);\end{aligned} \tag{ICC0}$$

$$\begin{aligned}&\theta_1 S(a_1, e^{br}(a_1, t_1, \theta_1)) + t_1 - ve^{br}(a_1, t_1, \theta_1) \geq \\ &\theta_1 S(a_0, e^{br}(a_0, t_0, \theta_1)) + t_0 - ve^{br}(a_0, t_0, \theta_1).\end{aligned} \tag{ICC1}$$

The expected welfare is written as the weighted sum of the total surplus. The two first constraints, (LLC0) and (LLC1), are limited liability constraints for type 0 and type 1, respectively. As already mentioned in the previous sections, they ensure that each agent agrees to sign the contract. The two following constraints, (ICC0) and (ICC1), are adverse selection incentive compatibility constraints for type 0 and type 1, respectively. They are set to ensure that each agent prefers the contract designed for her type. For $i \in \{0, 1\}$, type i agent must have a higher utility when she selects the contract that consists of transfer t_i and regulator's input a_i, as compared to when she chooses the other contract made for type $j \neq i$.

The last discussion in Section 4 referred to a self-selection issue when the regulator proposes the menu of second-best contracts (a_0^{SB}, e_0^{SB}, t_0^{SB}) and (a_1^{SB}, e_1^{SB}, t_1^{SB}) to the agents without being able to distinguish their types. Indeed, a type 0 agent prefers the contract (a_1^{SB}, e_1^{SB}, t_1^{SB}) designed for type 1 to the contract (a_0^{SB}, e_0^{SB}, t_0^{SB}) designed for herself. In terms of incentive compatibility constraint, this is equivalent to saying that the second-best solution violates the constraint (ICC0), which is the incentive compatibility constraint for the type 0 agent.

The third-best solution (a_0^{TB}, e_0^{TB}, t_0^{TB}, a_1^{TB}, e_1^{TB}, t_1^{TB}) is the solution to the regulator's problem (P3) and the following lemma characterizes its important properties.

Lemma 4. ***Properties of the third-best solution***

The third-best solution (a_0^{TB}, e_0^{TB}, t_0^{TB}, a_1^{TB}, e_1^{TB}, t_1^{TB}) satisfies

1.
   - $S_a(a_0^{TB}, e_0^{TB}) < S_a(a_0^{SB}, e_0^{SB}) \leq S_a(a^{FB}, e^{FB}) = S_a(a_1^{SB}, e_1^{SB}) < S_a(a_1^{TB}, e_1^{TB})$
   - $S_e(a^{FB}, e^{FB}) = S_e(a_1^{SB}, e_1^{SB}) < S_e(a_1^{TB}, e_1^{TB})$
   - $S_e(a^{FB}, e^{FB}) = S_e(a_1^{SB}, e_1^{SB}) < S_e(a_0^{SB}, e_0^{SB}) = S_e(a_0^{TB}, e_0^{TB})$
2.
   - $a_0^{TB} > a_1^{TB}$
   - $t_0^{TB} < t_1^{TB}$
   - $e_0^{TB} < e_1^{TB}$
3.
   - $e_0^{TB} = e^*(a_0^{TB}, \theta_0) \leq \frac{t_0^{TB}}{v}$
   - $e_1^{TB} = \frac{t_1^{TB}}{v} < e^*(a_1^{TB}, \theta_1)$
   - *either* $\frac{t_1^{TB}}{v} < e^*(a_1^{TB}, \theta_0)$ *or* $\frac{t_1^{TB}}{v} > e^*(a_1^{TB}, \theta_0)$

As in the second-best analysis, we confirm here that $e_0^{TB} = e^*(a_0^{TB}, \theta_0)$ and $e_1^{TB} = \frac{t_1^{TB}}{v}$. Again, the regulator uses the transfer t_1 to restrain an altruistic agent from exerting an effort that is high and too costly from the regulator's perspective because of the agent's limited liability constraint. With such a constraint on the choice of the altruistic agent, the input a_1 has no role to play as an incentive for a higher effort.

Conversely, for a selfish agent, the regulator uses the input a_0 to give her an incentive to exert a higher effort and the transfer more than compensates her for such an effort. The reason for such third-best contracts is that the selfish agent will not prefer the contract designed for the altruistic agent.

When the altruistic agent's effort is constrained by her transfer, then the slope of her indifference curve in the (a, t) space is negative but higher than the negative slope of the indifference curve of a selfish agent:

- $\frac{\partial t}{\partial a} = -\frac{\theta S_a(a, \frac{t}{v})}{\frac{\theta}{v} S_e(a, \frac{t}{v})} > -\theta S_a(a, e^*(a, \theta))$ when $U = \theta S(a, \frac{t}{v})$ is constant and $t < ve^*(a, \theta)$
- $\frac{\partial t}{\partial a} = -\theta S_a(a, e^*(a, \theta))$ when $U = t + \theta S(a, e^*(a, \theta)) - ve^*(a, \theta)$ is constant

This is equivalent to saying that, an altruistic agent indifference curve crosses the selfish agent's one from below. Therefore, to satisfy both (ICC0) and (ICC1), we need $a_0 - a_1 \geq 0 \geq t_0 - t_1$. This is consistent with the fact that input is not used to give incentives to the altruistic type. If instead, our analysis did not consider any limited liability constraint, just as in Jack (2005), then the regulator would not use the transfer t to constrain the effort of an altruistic agent. The input a would work as an incentive for effort for both a selfish and an altruistic agent, and even more so for the latter since she cares more. In such a case, the indifference curve of an altruistic agent would be more negative than the one of a selfish agent; the former would cross the latter from above. As a consequence, self-selection through (ICC0) and (ICC1) would require $a_0 - a_1 \leq 0 \leq t_0 - t_1$. This reasonning illustrates why our third-best contracts pay an altruistic agent better than a selfish one, while in Jack (2005), the opposite holds. The main rationale behind this difference is that without any limited liability constraint, the satisfaction of an altruistic agent from contributing to the production can be used to replace part of the transfer. These intuitions are consistent with the full comparison of the main variables of interest across all cases that the next proposition provides to complete the analysis.

Proposition 1. ***Comparison between first-, second- and third-best solutions***

- *Ranking of regulator's input levels:* $a_1^{TB} < a^{FB} = a_0^{SB} = a_1^{SB} < a_0^{TB}$
- *Ranking of agents' effort levels:* $e_0^{SB} < e_0^{TB} < e_1^{TB} < e^{FB} = e_1^{SB}$
- *Ranking of transfer levels:* $t_0^{TB} < t_1^{TB} < t^{FB} = t_1^{SB}$

All our proofs are based on the quadratic form of the production function $S(a, e)$, which is not without loss of generality. In particular, when we do not condition $S(a, e)$ to be quadratic, there is no reason for the regulator's input to be the same in both the first-best and the second-best analysis, and for both a type 1 and a type 0 agent. However, what we consider to be robust in our analysis is (1) that the transfer to an altruistic agent limits her choice of effort in the second-best analysis as well as in the third-best one and, because of this, that (2) the satisfaction of the self-selection constraints for the third-best contracts requires $a_0 - a_1 \geq 0 \geq t_0 - t_1$ rather than $a_0 - a_1 \leq 0 \leq t_0 - t_1$. The quadratic form of $S(a, e)$ does not affect the generality of the graphical argument that precedes Proposition 1 either.

Both moral hazard and adverse selection appear to matter in our setting even if no informational rents are due, because of the non-contractibility of outcome S and because of the specificities of our limited liability constraints, which impose that work must pay no matter whether the agent is altruistic or not. Indeed, the second-best solution with moral hazard only, is different from the first-best outcome, at least when altruism is low enough. Moreover, the solution to the mixed third-best problem is different from the solution to the adverse selection problem alone, which coincides with the first-best.

A shut-down policy is never optimal in our setting. Indeed, a shut-down policy would consider only one contract to attract only one type of agent. However, both types of agents have the possibility to adjust their level of effort so as to be better off with a contract than without it. At worst, they earn exactly the cost of their effort. Formally, the indirect utility of an agent writes as follows:

$$U(a, t, \theta) = \theta S(a, e^{br}(a, t, \theta)) + t - v e^{br}(a, t, \theta). \tag{1.7}$$

Following (ICC) defined in Section 4, if $\frac{t}{v} > e^*(a, \theta)$, then $e^{br}(a, t, \theta) = e^*(a, \theta)$ and thus $U(a, t, \theta) > \theta S(a, e * (a, \theta)) \geq 0$. If instead $\frac{t}{v} \leq e^*(a, \theta)$, then $e^{br}(a, t, \theta) = \frac{t}{v}$ and thus $U(a, t, \theta) = \theta S(a, \frac{t}{v}) \geq 0$. In both cases, an agent is at least as well off signing the contract than not signing it, no matter her level of altruism. Therefore, there is no contract that would attract only one type of agent.

1.6 Concluding remarks

In this paper, we analyze a particular class of regulator-agent relationships. The main features of our model are joint participation to the non-contractible outcome, different non-contractible concern for the outcome being produced, limited liability and non-contractible effort. Transfer and regulator's input are the two instruments to solve information asymmetry. This setting is relevant for sectors of activities such as education and health.[4]

[4] *Our normative analysis implicitly assumes that access to input such as personalized medicine techniques, hospitals, classrooms, etc. can be made divisible.*

Comparing first- and second-best solutions to analyze the effect of moral hazard, we show that a selfish agent's effort as well as the corresponding transfer are distorted downward and the regulator's input works as an incentive for effort by the selfish agent. For a given level of regulator's input, the altruistic agent instead would prefer to exert a higher effort than the first-best one. However, such an effort would be too costly to the regulator because of limited liability and therefore the regulator pays the altruistic agent exactly the first-best level of transfer so as to restrict the altruistic agent's choice of effort to not make losses. In our analysis of the consequences of combined adverse selection and moral hazard, we derive a separating contract to maximize welfare; the transfer is higher for an altruistic agent than for a selfish one and the regulator's input is higher for a selfish agent than for an altruistic one. Such a result contrasts with an existing one by Jack (2005) according to which, the selfish are paid better than the altruistic. It is the limited liability constraint that explains the difference between our results and Jack (2005)'s. Our main result also contrasts with the one of Brekke and Nyborg (2010) who conclude that motivated agents' wages must be kept strictly lower than their marginal productivity while overinvestment in equipment can be justified as a means to attract motivated agents.

The whole analysis with mixed moral hazard and adverse selection relies on having one type that is more altruistic than a regulator's perfect agent and one that is less. If instead, the two types were less altruistic than the regulator's perfect agent, then we may not have a priori ruled out bunching. However, this is outside the scope of this paper and we leave this for future research.

1.7 Appendices

1.7.1 Proof of Lemma 1

Proof. The first-best solution under point 1 directly follows the maximization of the regulator's welfare (P1'). The comparative statics under points 2 and 3 result from the total differentiation of the second-best solution under point 1. In particular,

- $\frac{\partial a^{FB}}{\partial \lambda} = \frac{cS_{ee} - vS_{ea}}{S_{aa}S_{ee} - (S_{ae})^2} < 0$
- $\frac{\partial e^{FB}}{\partial \lambda} = \frac{vS_{aa} - cS_{ea}}{S_{aa}S_{ee} - (S_{ae})^2} < 0$
- $\frac{\partial t^{FB}}{\partial \lambda} = v\frac{\partial e^{FB}}{\partial \lambda} < 0$

□

1.7.2 Proof of Lemma 2

Proof. The second-best solution under points 1 and 2 follows the maximization of the regulator's welfare (P2). We know from (ICC) that the agent's best-reply effort can be either an interior solution ($e^{br} = e^*(a,\theta)$) or a corner solution ($e^{br} = \frac{t}{v}$). Therefore, we rewrite (P2) for both cases in order to consider all relevant candidates for a solution.

If the agent's best-reply effort is interior, then we rewrite (P2) as (P2i):

$$\max_{a,t} W = S(a, e^*(a,\theta)) - (1+\lambda)(t+ca) + t - ve^*(a,\theta) \tag{P2i}$$

s.t

$$t - ve^*(a,\theta) \geq 0;$$

where $e^*(a,\theta)$ is implicitely defined by $\theta S_e(a, e^*(a,\theta)) - v = 0$ (see Equation (3) in Section 4).

The solution to (P2i) is denoted (a^{SB}, e^{SB}, t^{SB}) and it is given by:

- $S_a(a^{SB}, e^{SB}) = (1+\lambda)c - kv(\frac{1}{\theta} - (1+\lambda))$
- $S_e(a^{SB}, e^{SB}) = \frac{1}{\theta}v$
- $t^{SB} = ve^{SB}$

Total differentiation of this solution leads to:

- $\frac{\partial a^{SB}}{\partial \theta} = 0$
- $\frac{\partial e^{SB}}{\partial \theta} = -\frac{v}{\theta^2 S_{ee}} > 0$

In particular, when $(1+\lambda)\theta = 1$, this solution candidate (a^{SB}, e^{SB}, t^{SB}) coincides with the first-best solution (a^{FB}, e^{FB}, t^{FB}). Therefore,

- $a^{SB} = a^{FB}, \forall\theta,$
- $e^{SB} > e^{FB} \iff (1+\lambda)\theta > 1$
- $e^{SB} = e^{FB} \iff (1+\lambda)\theta = 1$
- $e^{SB} < e^{FB} \iff (1+\lambda)\theta < 1$

The regulator's indirect utility corresponding to this solution to (P2i) is thus:

$$W^{SB} = S(a^{SB}, e^{SB}) - (1+\lambda)(ve^{SB} + ca^{SB}). \tag{1.8}$$

If instead, the agent's best-reply effort is constrained by (LCC), then $e^{br} = \frac{t}{v}$ and we rewrite (P2) as (P2c):

$$\max_{a,t} W = S(a, \frac{t}{v}) - (1+\lambda)(t + ca) \tag{P2c}$$

s.t

$$t - ve^*(a,\theta) \leq 0;$$

where $e^*(a,\theta)$ is implicitely defined by $\theta S_e(a, e^*(a,\theta)) - v = 0$.

The solution to (P2c) can be unconstrained, in which case it coincides with the first-best solution:

- $S_a(a^{FB}, e^{FB}) = (1+\lambda)c$
- $S_e(a^{FB}, e^{FB}) = (1+\lambda)v$
- $t^{FB} = ve^{FB}$

For the constraint to be satisfied, the following condition must hold: $t^{FB} - ve^*(a^{FB},\theta) \leq 0$. Given that $a^{SB} = a^{FB}$ and $t^{FB} = ve^{FB}$, this constraint reduces to $e^{FB} \leq e^{SB}$, which itself is equivalent to $(1+\lambda)\theta \geq 1$.

The regulator's indirect utility corresponding to this unconstrained solution to (P2c) is thus:

$$W^{FB} = S(a^{FB}, e^{FB}) - (1+\lambda)(ve^{FB} + ca^{FB}). \tag{1.9}$$

Alternatively, the solution to (P2c) can be constrained by $t - ve^*(a,\theta) \leq 0$, in which case it coincides with the solution to (P2i):

- $S_a(a^{SB}, e^{SB}) = (1+\lambda)c - kv(\frac{1}{\theta} - (1+\lambda))$
- $S_e(a^{SB}, e^{SB}) = \frac{1}{\theta}v$
- $t^{SB} = ve^{SB}$

To summarize, we have two candidates for the solution, either (a^{FB}, e^{FB}, t^{FB}) under the condition $(1+\lambda)\theta \geq 1$, or (a^{SB}, e^{SB}, t^{SB}), with $a^{FB} = a^{SB}$. To compare W^{FB} and

W^{SB}, we recognize that (a^{FB}, e^{FB}) maximizes $W = S(a,e) - (1+\lambda)(ve+ca)$. Therefore, $W^{FB} > W^{SB}$ and the best candidate for a solution is (a^{FB}, e^{FB}, t^{FB}) whenever it is feasible, that is, under the condition $(1+\lambda)\theta \geq 1$. Otherwise, when $(1+\lambda)\theta \leq 1$, the only candidate for a solution is (a^{SB}, e^{SB}, t^{SB}).

The comparative statics under point 1 result from the total differentiation of the second-best solution under point 1.

In particular, when $(1+\lambda)\theta < 1$, we have

- $\frac{\partial a^{SB}}{\partial \lambda} = -\frac{1}{S_{ae}} \frac{v(S_{ae})^2 - cS_{ae}S_{ee}}{S_{aa}S_{ee} - (S_{ae})^2} < 0$
- $\frac{\partial e^{SB}}{\partial \lambda} = \frac{1}{S_{ee}} \frac{v(S_{ae})^2 - cS_{ae}S_{ee}}{S_{aa}S_{ee} - (S_{ae})^2} < 0$
- $\frac{\partial t^{SB}}{\partial \lambda} = v\frac{\partial e^{SB}}{\partial \lambda} < 0$

and

- $\frac{\partial a^{SB}}{\partial \theta} = 0$
- $\frac{\partial e^{SB}}{\partial \theta} = -\frac{v}{\theta^2 S_{ee}} > 0$
- $\frac{\partial t^{SB}}{\partial \theta} = -\frac{v^2}{\theta^2 S_{ee}} > 0$

Conversely, if $(1+\lambda)\theta \geq 1$, then the comparative statics are as in Lemma 1. □

1.7.3 Proof of Lemma 3

Proof. When $(1+\lambda)\theta \geq 1$, the second-best solution coincides with the first-best solution. Moreover, Lemma 2 states that a^{SB} is invariant to changes in altruism θ while e^{SB} and t^{SB} both increase with θ when $(1+\lambda)\theta < 1$. Using this result together with the approximation $S_{ae}/S_{ee} = -k$, directly yields the result in Lemma 3. □

1.7.4 Proof of Lemma 4

Proof. We *a priori* assume that $ve^*(a_i, \theta_1) \geq t_i \geq ve^*(a_i, \theta_0)$, for $i = \{0, 1\}$, so as to have $e_0 = e^*(a_i, \theta)$ for $\theta = \theta_0$ and $e_1 = \frac{t_i}{v}$ for $\theta = \theta_1$, as in the second-best. We include the type 0's limited liability constraint into our maximization program ($t_0 \geq ve^*(a_0, \theta_0)$) and we leave the other three inequalities for *ex post* verification.

Since the second-best solution violates (ICC0), we consider that such an incentive compatibility constraint must bind when solving (P3) and thus we substitute the binding (ICC0) into (P3). As is usual with adverse selection problems, we leave the constraint (ICC1) outside of the maximization program for now and we check it *ex post*.

We can thus simplify (P3) as follows:

$$
\begin{aligned}
\max_{a_0,a_1,t_1} W =&\alpha \left\{ S(a_1, \frac{t_1}{v}) - (1+\lambda)(t_1 + ca_1) \right\} \\
&+ (1-\alpha)\{(1+\lambda\theta_0)S(a_0, e^*(a_0,\theta_0)) - (1+\lambda)(ve^*(a_0,\theta_0) + ca_0)\} \\
&- (1-\alpha)\lambda\{t_1 + \theta_0 S(a_1, e^*(a_1,\theta_0)) - ve^*(a_1,\theta_0)\} \qquad \text{(P3')}
\end{aligned}
$$

s.t

$$
t_1 + \theta_0 S(a_1, e^*(a_1,\theta_0)) - ve^*(a_1,\theta_0) - \theta_0 S(a_0, e^*(a_0,\theta_0)) \geq 0. \qquad \text{(LLC0)}
$$

To solve (P3'), we consider two cases: Either (LLC0) does not bind, or it does.

1.7.4.1 case 1

If (LLC0) does not bind, then the first-order conditions directly imply:

- $S_a(a_0^{TB}, e^*(a_0^{TB}, \theta_0)) = \frac{1+\lambda}{1+\lambda\theta_0}c - \frac{1-\theta_0}{\theta_0(1+\lambda\theta_0)}kv < S_a(a_0^{SB}, e^*(a_0^{SB}, \theta_0))$
- $S_a(a_1^{TB}, \frac{t_1^{TB}}{v}) = (1+\lambda)c + \frac{(1-\alpha)\lambda\theta_0}{\alpha} S_a(a_1^{TB}, e^*(a_1^{TB}, \theta_0)) > S_a(a^{FB}, e^{FB})$
- $S_e(a_1^{TB}, \frac{t_1^{TB}}{v}) = (1+\frac{\lambda}{\alpha})v > S_e(a^{FB}, e^{FB})$

Furthermore, $e_0^{TB} = e^*(a_0^{TB}, \theta_0)$ implies

- $S_e(a_0^{TB}, e^*(a_0^{TB}, \theta_0)) = \frac{v}{\theta_0} = S_e(a_0^{SB}, e_0^{SB})$

This series of four inequalities together with the constraints (in case they are satisfied) imply the following comparison between $(a_0^{TB}, t_0^{TB}, e_0^{TB})$ and $(a_1^{TB}, t_1^{TB}, e_1^{TB})$:

- $a_0^{TB} > a_1^{TB}$
- $t_0^{TB} < t_1^{TB}$
- $e_0^{TB} < e_1^{TB}$

It is easy to check that *a priori* constraints $t_1 \leq ve^*(a_1^{TB}, \theta_1)$ and $t_0 \leq ve^*(a_0^{TB}, \theta_1)$ both hold.

However, the constraint $t_1 \geq ve^*(a_1^{TB}, \theta_0)$ holds if and only if $\theta_0(1+\frac{\lambda}{\alpha}) \leq 1$. Therefore, we need to check in a separate case (case 3 and case 4, depending on whether (LLC0) binds or not)) what happens when the opposite constraint holds: $t_1 \leq ve^*(a_1^{TB}, \theta_0)$.

Last, (ICC1) holds as long as (ICC0) holds together with $a_0^{TB} - a_1^{TB} > 0 > t_0^{TB} - t_1^{TB}$.

1.7.4.2 case 2

If (LLC0) binds, then

- $t_0^{TB} = ve^*(a_0^{TB}, \theta_0)$.

Substituting t_0^{TB} into binding (ICC0), we obtain

- $t_1^{TB} = ve^*(a_0^{TB}, \theta_0) + (\theta_0 S(a_0, e^*(a_0, \theta_0)) - ve^*(a_0, \theta_0)) - (\theta_0 S(a_1, e^*(a_1, \theta_0)) - ve^*(a_1, \theta_0))$

and the first-order conditions directly imply:

- $S_a(a_0^{TB}, e^*(a_0^{TB}, \theta_0)) \left\{ 1 - \alpha + \alpha\theta_0(\frac{1}{v} S_e(a_1^{TB}, \frac{t_1^{TB}}{v}) - (1+\lambda)) \right\} - (1-\alpha)(1+\lambda)c + \left\{ \frac{1-\alpha}{\theta_0} - (1+\lambda) + \frac{\alpha}{v} S_e(a_1^{TB}, \frac{t_1^{TB}}{v}) \right\} kv = 0$

- $S_a(a_1^{TB}, \frac{t_1^{TB}}{v}) - (1+\lambda)c - \theta_0 S_a(a_1^{TB}, e^*(a_1^{TB}, \theta_0)) \left\{ \frac{1}{v} S_e(a_1^{TB}, \frac{t_1^{TB}}{v}) - (1+\lambda) \right\} = 0$

Again, these conditions imply either $t_1 > t_0$ and $a_1 > a_0$, which we reject because it is not compatible with (ICCO), or what is stated in Lemma 4:

- $S_a(a_0^{TB}, e_0^{TB}) < S_a(a_0^{SB}, e_0^{SB}) \leq S_a(a^{FB}, e^{FB}) = S_a(a_1^{SB}, e_1^{SB}) < S_a(a_1^{TB}, e_1^{TB})$

- $S_e(a^{FB}, e^{FB}) = S_e(a_1^{SB}, e_1^{SB}) < S_e(a_1^{TB}, e_1^{TB})$

- $S_e(a^{FB}, e^{FB}) = S_e(a_1^{SB}, e_1^{SB}) < S_e(a_0^{SB}, e_0^{SB}) = S_e(a_0^{TB}, e_0^{TB})$

We now depart from the conditions on t_i, $i = \{0, 1\}$, that prevail in the second-best analysis and we assume instead $t_1 \leq ve^*(a_1, \theta_0)$. The main difference is the type 0's choice of effort, should she choose the contract designed for the type 1 agent. Therefore, we modify (ICC0) accordingly:

$$t_0 + \theta_0 S(a_0, e^*(a_0, \theta_0)) - ve^*(a_0, \theta_0) - \theta_0 S(a_1, \frac{t_1}{v}) = 0 \tag{ICC0}$$

We maintain the other three *a priori* assumptions on t_i, $i = \{0, 1\}$. We include (LLC0) in the maximization program and leave the other three for *ex post* verification. We do the same for (ICC1).

Therefore, our maximization program now writes as follows:

$$\max_{a_0,a_1,t_1} W = \{\alpha - \lambda\theta_0(1-\alpha)\} S(a_1, \frac{t_1}{v}) - \alpha(1+\lambda)(t_1 + ca_1)$$
$$+ (1-\alpha)\{(1+\lambda\theta_0)S(a_0, e^*(a_0,\theta_0)) - (1+\lambda)(ve^*(a_0,\theta_0) + ca_0)\} \quad \text{(P3")}$$

s.t

$$S(a_1, \frac{t_1}{v}) - S(a_0, e^*(a_0,\theta_0)) \geq 0; \quad \text{(LLC0)}$$

Again, to solve (P3"), we consider two cases: Either (LLC0) does not bind, or it does.

1.7.4.3 case 3

If (LLC0) does not bind, then the first-order conditions directly imply:

- $S_a(a_0^{TB}, e^*(a_0^{TB}, \theta_0)) = \frac{1+\lambda}{1+\lambda\theta_0}(c+kv) - \frac{kv}{\theta_0} < S_a(a_0^{SB}, e^*(a_0^{SB}, \theta_0))$
- $S_a(a_1^{TB}, \frac{t_1^{TB}}{v}) = \frac{\alpha}{\alpha-\lambda\theta_0(1-\alpha)}(1+\lambda)c > S_a(a^{FB}, e^{FB})$
- $S_e(a_1^{TB}, \frac{t_1^{TB}}{v}) = \frac{\alpha}{\alpha-\lambda\theta_0(1-\alpha)}(1+\lambda)v > S_e(a^{FB}, e^{FB})$

Furthermore, $e_0^{TB} = e^*(a_0^{TB}, \theta_0)$ implies

- $S_e(a_0^{TB}, e^*(a_0^{TB}, \theta_0)) = \frac{v}{\theta_0} = S_e(a_0^{SB}, e_0^{SB})$

This series of four inequalities together with the constraints (in case they are satisfied) imply the following comparison between $(a_0^{TB}, t_0^{TB}, e_0^{TB})$ and $(a_1^{TB}, t_1^{TB}, e_1^{TB})$:

- $a_0^{TB} > a_1^{TB}$
- $t_0^{TB} < t_1^{TB}$
- $e_0^{TB} < e_1^{TB}$

It is easy to check that *a priori* constraints $t_1 \leq ve^*(a_1^{TB}, \theta_1)$ and $t_0 \leq ve^*(a_0^{TB}, \theta_1)$ both hold.

However, the constraint $t_1 \leq ve^*(a_1^{TB}, \theta_0)$ holds if and only if $\theta_0(1+\frac{\lambda}{\alpha}) \geq 1$, which perfectly complements case 1.

Last, (ICC1) holds as long as (ICC0) holds together with $a_0^{TB} - a_1^{TB} > 0 > t_0^{TB} - t_1^{TB}$.

1.7.4.4 case 4

If (LLC0) binds, then the first-order conditions associated with (P3”) imply the following inequalities (with μ denoting the lagrange multiplier associated to (LLC0)):

- $S_a(a_0^{TB}, e^*(a_0^{TB}, \theta_0)) = \frac{1-\alpha}{1-\alpha+\mu}(1+\lambda)(c+kv) - \frac{kv}{\theta_0} < S_a(a_0^{SB}, e^*(a_0^{SB}, \theta_0))$
- $S_a(a_1^{TB}, \frac{t_1^{TB}}{v}) = \frac{\alpha}{\alpha-\mu}(1+\lambda)c > S_a(a^{FB}, e^{FB})$
- $S_e(a_1^{TB}, \frac{t_1^{TB}}{v}) = \frac{\alpha}{\alpha-\mu}(1+\lambda)v > S_e(a^{FB}, e^{FB})$

Furthermore, $e_0^{TB} = e^*(a_0^{TB}, \theta_0)$ implies

- $S_e(a_0^{TB}, e^*(a_0^{TB}, \theta_0)) = \frac{v}{\theta_0} = S_e(a_0^{SB}, e_0^{SB})$

This series of four inequalities together with the constraints (in case they are satisfied) imply the following comparison between $(a_0^{TB}, t_0^{TB}, e_0^{TB})$ and $(a_1^{TB}, t_1^{TB}, e_1^{TB})$:

- $a_0^{TB} > a_1^{TB}$
- $t_0^{TB} < t_1^{TB}$
- $e_0^{TB} < e_1^{TB}$

The *a priori* constraints are satisfied again.

1.7.4.5 case 5

To be complete and because (PC0) can possibly bind, we must consider the alternative $t_0 < ve^*(a_0, \theta_0)$, so that $e_0^{TB} = \frac{t_0}{v}$. Solving the regulator's maximization program with $e_i^{TB} = \frac{t_i}{v}$, $i = \{0, 1\}$, would lead to the first-best solution, which contradicts $t_0 < ve^*(a_0, \theta_0)$. Therefore, the present case presents no relevant candidate for the third-best contracts solution.

To sum up, all relevant cases (1 to 4) lead to the same properties stated in Lemma 4. No constraint concerning $e_0^{TB} = e^*(a_0^{TB}, \theta_0)$ is ever violated and no constraint concerning $e_1^{TB} = \frac{t_1}{v}$ does ever bind, which conforts our results in Lemma 4.

□

1.7.5 Proof of Proposition 1

Proof. Straightforward from the preceding lemmata. □

References

Barigozzi, Francesca, and Burani, Nadia. 2016. Competition and screening with motivated health professionals. Journal of Health Economics, 50, 358-371.

Biglaiser, Gary, and Ma, Albert Ching-to. 2007. Moonlighting: public service and private practice. The RAND Journal of Economics, 38(4), 1113-1133.

Brekke, Kjell Arne and Nyborg, Karine. 2010. Selfish bakers, caring nurses? A model of work motivation. Journal of Economic Behavior & Organization 75, 377-394.

Brosig-Koch, Jeannette, Hennig-Schmidt, Heike, Kairies-Schwarz, Nadja, and Wiesen, Daniel. 2017. The Effects of Introducing Mixed Payment Systems for Physicians: Experimental Evidence. Health Economics, 26(2), 243-262.

Canidio, Andrea, and Gall, Thomas. 2019. Rewarding idleness. Journal of Public Economic Theory, to appear, doi.org/10.1111/jpet.12362.

Chone, Philippe, and Ma, Ching-to Albert. 2011. Optimal health care contract under physician agency. Annals of Economics and Statistics/Annales d'Economie et de Statistique, 229-256.

Dor, Avi, and Watson, Harry. 1995. The hospital-physician interaction in US hospitals: evolving payment schemes and their incentives. European Economic Review, 39(3), 795-802.

Godager, Geir, and Wiesen, Daniel. 2013. Profit or Patients' Health Benefit? Exploring the Heterogeneity in Physician Altruism. Journal of Health Economics, 32(6), 1105-1116.

Jack, William. 2005. Purchasing health care services from providers with unknown altruism. Journal of health economics, 24(1), 73-93.

Jelovac, Izabela, and Macho-Stadler, Ines. 2002. Comparing organizational structures in health services. Journal of Economic Behavior and Organization, 49(4), 501-522.

Koszegi, Botond. 2014. Behavioral contract theory. Journal of Economic Literature, 52(4), 1075-1118.

Laffont, Jean-Jacques, and Martimort, David. 2002. The theory of incentives: the principal- agent model. Princeton university press.

Lagarde, M. and D. Blaauw (2014). Pro-social preferences and self-selection into jobs: Evidence from South African nurses. Journal of Economic Behavior & Organization 107, Part A: 136-152.

Liu, Ting, and Ma, Ching-to Albert. 2013. Health insurance, treatment plan, and delegation to altruistic physician. Journal of Economic Behavior and Organization, 85, 79-96.

Macho-Stadler, Ines, and Perez Castrillo, David. 2018. Moral hazard: Base models and two extensions, in the "Handbook of Game Theory and Industrial Organization Vol 1," L.C. Corchon and M.A. Marini (eds.), Edward Elgar, 2018.

Naegelen, Florence, and Mougeot, Michel. 2011. Power of incentives with motivated agents in public organizations. Journal of Public Economic Theory, 13(3), 391-416.

Watt, Helen M.G., Richardson, Paul W., and Wilkins, Kari. 2014. Profiles of professional engagement and career development aspirations among USA preservice teachers. International Journal of Educational Research, 65(Supplement C), 23-40.

Chapter 2

Physicians' incentives to adopt personalized medicine: experimental evidence

This chapter is based on a joint research with David Bardey (Los Andes University and Toulouse School of Economics) and Bruno Ventelou. I started working on this paper in winter 2016. I proposed the research idea to Bruno and David and we designed the experiment and drafted the paper together.

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 as a "commitment device". However, quantification of our results suggests that its positive impact is not strong enough to justify generalizing paid access to personalized medicine.

2.1 Introduction

Personalized medicine (PM) 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, with the issue not appearing to be only one of cost. The literature has documented reasons for this under-use of personalized medicine techniques. A first argument is found in Antoñanzas et al., (2015) who report that there exists an uncertainty about personalized medicine test results. This uncertainty is likely to prevent physicians from relying on personalized medicine tests. A second argument is given by Howard et al., (2017) who argue that payment schemes might not provide enough incentives to adopt these new technologies. They find that physicians paid under a FFS plan tend to under-use PM technologies even when it is available for free. The inter-play between access to PM technologies (free *vs.* alternative modes) and physicians' payment systems will be the heart of our contribution. 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 design an experiment to replicate the physician's trade-offs under the option of personalized medicine information. As implemented, subjects are placed in a real-effort task game, as per Green, (2014); Bejarano et al., (2017) and, less directly, Lagarde and Blaauw (2017). In our experiment, ninety-five prospective physicians perform a task simulating the option of access to information likely to help them to take better care of a patient. First, in order to imitate the relationship between the physician and the patient, our game is similar to that of Green (2014), consisting in proofreading short texts with potential positive benefits for a third party (the "patient") when the texts are corrected well. Second, we design 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 making 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" is defined in the texts and only actions within this subset are considered to generate potential benefits for the third party. Thus, the proofreader's efficiency crucially depends on the informational input on priority sentences. Precise information on which sentences are "priority" is 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 to make corrections in priority sentences differ, as do motivations for buying personalized medicine. The prospective physicians are subjected to different payment schemes in a mix of within/between design. We explore three pure payment schemes: fee-for-service (FFS), capitation payment (CAP) and payment-for-performance (P4P). CAP is designed as a payment per "treated" text. FFS is designed to reward the physician based on the quantity of services provided (number of words corrected in the text). Lastly, P4P is designed as a payment for a minimum number of appropriate corrections in priority sentences. Our empirical strategy involves a two-stage panel least-square estimation, used to compare the behavior of subjects who buy personalized medicine information with that of subjects who do not buy it, under the different payment schemes.

Our main results can be summarized as follows. First, as expected, our experiment reveals that P4P –a quality-oriented remuneration scheme– yields stronger incentives to prospective physicians to buy personalized medicine techniques than FFS or CAP. In line with this first result, it seems that our subjects are also sensitive to financial incentives in their patient-care activities: while they treat more patients when paid by CAP, they perform more medical interventions under an FFS payment scheme (already in Green (2014)). We also find that CAP and P4P tend to generate similar incentives regarding the number of interventions; however, P4P is 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 paid. 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 subjects buy information, they make much better use of it, compared to the situation where access to personalized medicine information is free for 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, the regulator should partially subsidize it, opting for a P4P scheme to enhance the adoption of relevant tools.

2.1.1 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 gate-keeping 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 U.S breast cancer Medicare patients, so as to identify patients who are highly responsive to the IMRT option. Physicians work either in free-standing clinics (where their FFS payment plan also includes a monetary reward for treating patients with IMRT), 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.

Counterfactual is 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 payments. Over the last decade, a burgeoning literature has used experimental economics to study physician payments: 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); Bejarano et al., (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) reports 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, precisely, the former tends to be more "patient-oriented" than the latter.

In the light of this literature, we will opt for the recruitment of advanced medical students (prospective physicians) as our experimental subject pool. From a methodological point of view, we follow Green, (2014), Bejarano et al., (2017) and, less directly, Lagarde and Blaauw, (2017), who design a real-effort task experiment rather than the 'declared-effort' used in earlier work on physician payment schemes. Equivalence between real and chosen effort is proven for altruistic behaviors in gift-exchange games (Brüggen and Strobel, 2007). However, 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.[1] In Green's experiment, the optimal quantity of services depend 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.

2.2 Data and methodology

We imitate the physician-patient relationship by using an experimental game having two sequential phases: phase 1 organized for patients, and phase 2 for physicians. Before phase 1 and 2, we have selected 48 short texts, 36 of primary-school level and 12 of first-year secondary-school level. Each selected text contains words with errors (spelling, syntax, vocabulary).

[1] 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 in order to better focus on the issue of personalized medicine.

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

The aim of phase 1 is to "materialize" patients who will enter the experimentally-created physician-patient relationship. In experiments on physician behaviors, the physician should normally be able to identify the patient who will also be benefiting from his effort. In the first experiment on physician behavior for example, Hennig Schmidt et al., (2011) use abstracts patients in the lab, but their subjects know that gains generated for these abstract patients will benefit "real patients outside the lab". In these types of designs (this setting is also used by Brosig-Koch et al., [2013], Hennig-Schmidt and Wiesen (2014) among others), the total amount generated for "patients" is given to an entity outside the lab, a hospital for example. The advantage of this kind of experimental settings is that the experimenter can use "abstract" patients in the lab, therefore not rely on a "proper phase 1" with "real patients". This procedure might however be at the cost of measuring "artificially" low levels of physician altruism.

We opt for a sequential game with patients playing phase 1. This is similar to the designs of Green, (2014) and Bejarano et al., (2017). In these designs, patients express their "symptoms" in a first phase, and in a second phase, physicians address these symptoms under different payment mechanisms. At the end of phase 2, the patient receives the experimentally-generated gains, which represents his health benefit.

48 short texts are selected prior to organizing phase 1. These texts are then given to a group of 8 subjects (a set of 6 texts per subject). Each of the 8 subjects has to highlight words. Those words are the ones displayed to prospective physicians playing phase 2. Health benefits are computed based only on corrections done on those highlighted words. The main difference between our design and Green (2014)'s one is that, rather than letting subjects choose the words, we indicate those words to them. Proceeding this way allows to control "patient-heterogeneity" and we can better focus on the issue of personalized medicine. In Figure 2.1, we present the timeline of the main steps of the experiment.

Phase 1 is organized with 8 subjects, students from the department of economics of Aix-Marseille University. They receive 2 sets of 48 texts. Some words (both correct and incorrect) are in bold in one set of texts, not in the other set. The task of each of the 8 subjects is to highlight manually, on the unmarked set of texts, words that are in bold on another set. They use a yellow highlighter for this task. They are told that they are participating in a 2-phase game in which they were playing the first phase. To further ensure incentive compatibility, we will inform them that an additional payment will be generated by other subjects playing phase 2 of the game.

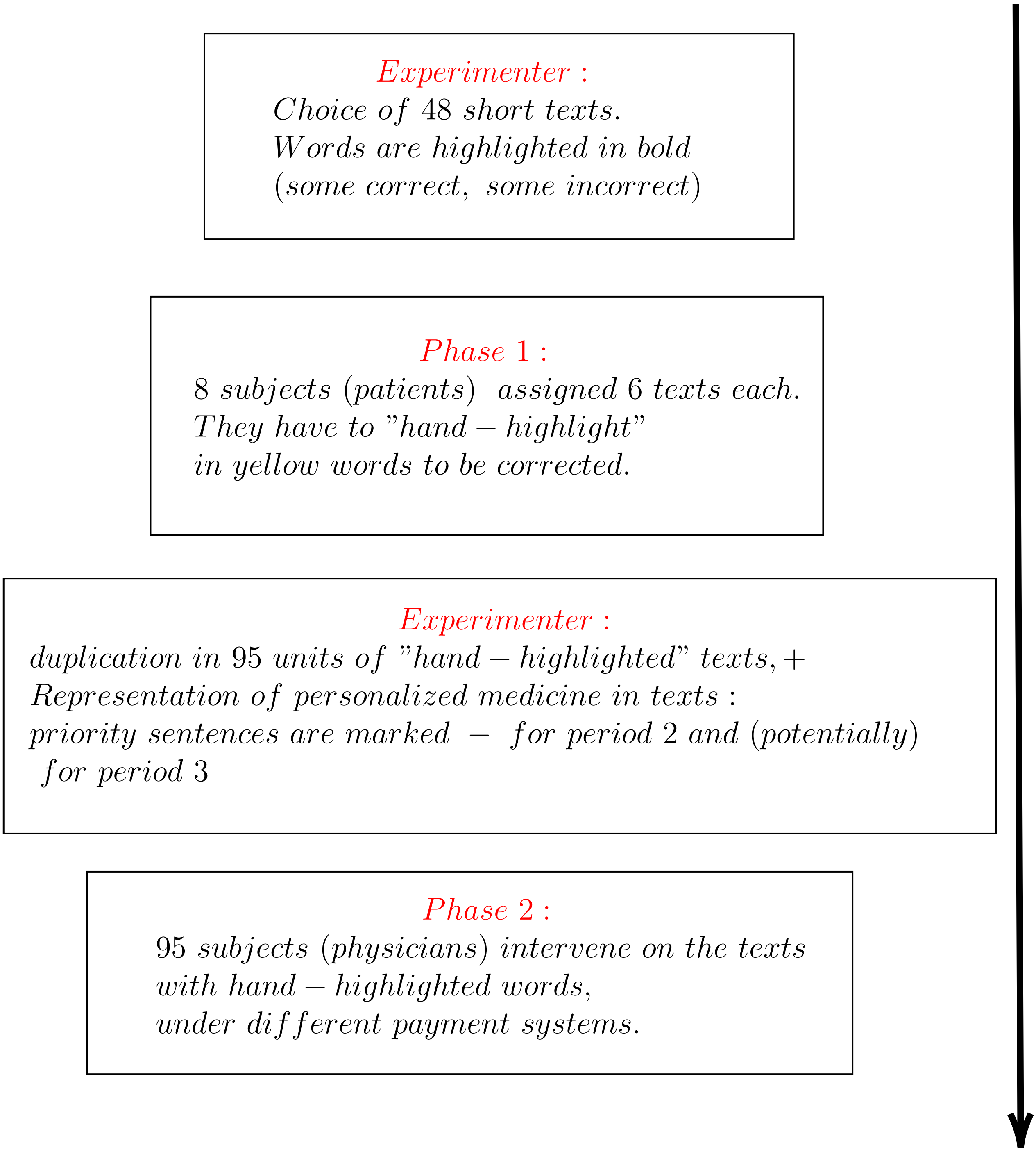


Figure 2.1 – Timeline of the main steps of the experiment

For this session, each of the 8 subjects is given a fixed endowment of €10. Each "bold-word" missed in the text incurs 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 the instructions.

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

In phase 2, we have ran different experimental sessions with advanced medical students playing the role of physicians. We implement 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 introduce a representation of personalized medicine by including access to information on priority sentences (underlined). The timeline of phase 2 is summarized in Figure 2.2.

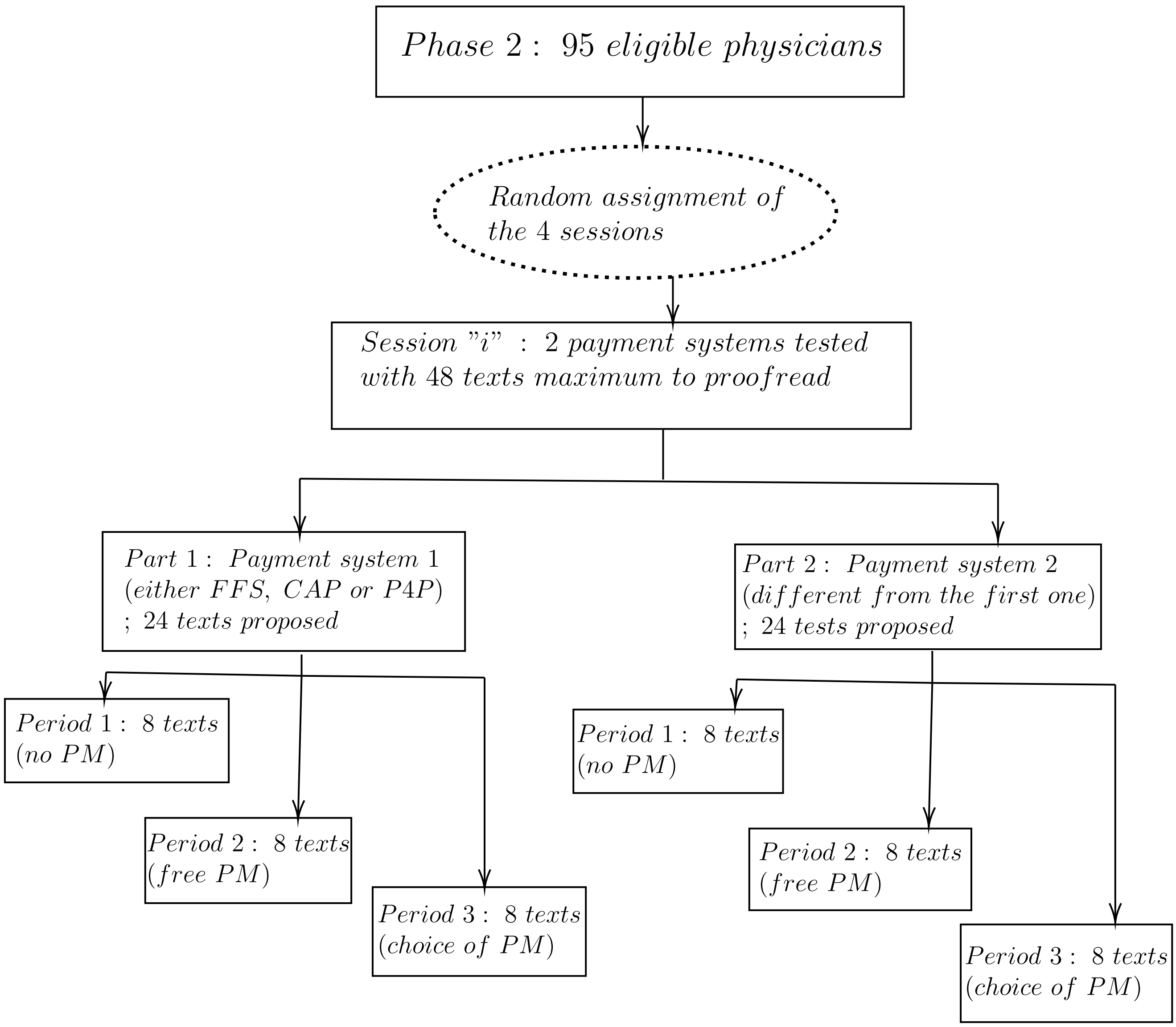


Figure 2.2 – Timeline of phase 2

Each treatment contains three successive periods of proofreading corresponding to three informational contexts:

- **Period 1:** 8 texts are presented without showing priority sentences, corresponding to a situation where personalized medicine is not available.
- **Period 2:** 8 texts are presented with priority sentences underlined, corresponding to a situation where personalized medicine is accessible free of charge.
- **Period 3:** the physician first has 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 corrects the texts, playing the game either as in period 1 or 2. If he chooses to have access to priority sentences, he is charged a fixed €0.50 per treated text (a text is considered treated if there is at least one correction). He is not charged otherwise. This last period captures a situation where the physician chooses whether or not to buy personalized medicine.

> 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.

> 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 2.3 – Examples of texts given to physicians in periods 1 and 2.

In Figure 2.3, 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 are either as in period 1 or in period 2. In appendix (table 2.11), we give a summary of all the data that are presented to each subject: total number of highlighted words, total number of wrong words, total number of correct words highlighted.

Physicians have 5 minutes per period to correct 8 short texts. They are free to allocate their time on the texts as they wish, including not altering some of them. For each treatment (payment mechanism), 24 texts are proposed (8 per period), so physicians can work on up to 48 texts per experimental session. Treatment variables are Capitation payment (CAP), Pay-For-Performance (P4P) and Fee-For-Service (FFS). Under CAP, the physician is 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 earn €0.30 per intervention, again regardless of appropriacy. Each text has a minimum of 6 highlighted words and a maximum of 12, thus the earnings range per text under FFS is between €1.80 and €3.60. Under P4P, the physician earns €2.50 per text if 80% of words in priority sentences are correctly written at the end of the proofreading, and nothing otherwise. The priority sentences contain between 5 and 9 words, and their positions in the texts vary.

To avoid portfolio strategies, we remunerate 2 periods chosen randomly (one period for the first treatment and one period for the second treatment). To ensure that physicians would earn approximately the same amounts, we ran a pilot experiment with different payment parameters. Based on these pilots, we chose the payment parameters used in the paper. There is equivalence for example between either (1) doing one action per text on the 6 proposed texts in CAP; or (2) intervening on 80% of all the highlighted words in a given period (doing approximately 5 words per text); or (3) correcting appropriately 4 texts in P4P.

Table 2.1 – Summary of the experimental payment mechanism parameters

Experimental condition	Payment parameter
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 is presented to physicians as a game in 2 parts. In each part, the physicians plays the 3 periods and they are informed of the payment system at the beginning of each part. We randomize the order of P4P.

The 4 sessions of Phase 2 have taken place between January and March 2017. All our 95 prospective physicians are advanced medical students from Aix-Marseille and Nice Universities. As mentioned in the introduction, medical students are chosen

Table 2.2 – 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

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: Physicians know that their actions can generate a financial gain for their subject-1 counterparts who has highlighted words in the texts in Phase-1 of the dual game. This represents the altruistic part of medical activity.[2] Subject-1 counterparts receive €5 if 90% of the highlighted words in priority sentences are corrected appropriately, and nothing otherwise.

Recruitment procedure: To ensure anonymity of registrations, our advanced medical students were invited through the student's association. The invitation message contained the date 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 was needed during the experiment: pens, sets of texts and instructions. There was enough space between subjects to avoid peer-influence in actions and decisions. 15 minutes was allowed to instructions' reading and completion of a comprehension test on instructions (results are available upon request). To ensure anonymity with respect to the experimenter and

[2]We can identify three main ways to introduce altruistic preferences in experiments on physicians' behavior: (i) the experimenter informs the subject playing the role of physician that his game-generated gains for "patients" will benefit real patients outside the lab; (ii) the experimenter explicitly recruits subjects to play the role of patients, those subjects receiving the experimentally-generated gains for "patients"; (iii) the experimenter runs a dictator game or any other simple game that can enable to measure physician altruism. Cases (i) and (ii) are incentive-compatible ways of generating altruism used by authors cited in the introduction. However, as mentioned above, 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 experiment, we explicitly introduce patients as subjects-1 from the first phase.

the students' association, subjects' earnings have been delivered to them by the university accountants.

At the end of each experimental session, we also 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 (categorical variable named *TECHNO* in the econometric analysis). Of this additional set, the first two variables are used as controls for the analysis, while *TECHNO* is also used as an instrumental variable.

The mean age in our sample is 22 years old. Our subject pool is made of 57% female and 58% of subjects are in year 4 of their medical school or above. The minimum and maximum earnings per physicians were respectively €6.20 and €35.40, with a mean and a median around €20. Out of the 95 subjects playing physicians' role, 5 did not collect their earnings. All subjects-1 collected their additionally-generated payments.

2.2.3 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 being 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.[3] 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

[3] First yielding a free access to personalized medicine before the physician makes his choice in the costly setting captures real word situations where the physician first learns the benefits of the technology and then decides whether he wants to invest in it. This roll-out of the experiment has the advantage that the physician is arguably making an informed decision (what our experiment captures). This roll-out is however at the expense of not knowing what would have happened if the physician was first proposed access to personalized medicine, without knowing potential benefits for him. In such cases, we hypothesize that physician's attitude towards risk might be at play in physician's decision to invest in personalized medicine. These cases are left for future research.

cost to physicians and the benefit to patients are monetized.[4] Table 2.3 summarizes how our experimental settings correspond to real-life medical settings.

Table 2.3 – Correspondence between experimental and real-life personalized medicine settings

	In experimental setting	**In real-life setting**
Period 1	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	Priority sentence shown, physician can target/focus interventions	Subset of relevant symptoms shown, physician can target/focus interventions
Period 3	Are you willing to buy the information on priority sentences?	Are you willing to buy (/spend time on obtaining) PM information?
Payment schemes	-% quality of overall text -per intervention -per text	-P4P -FFS -Capitation

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

[4] This monetization is, in our view, the first 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 second limitation that we see is the fact that our P4P is defined as a remuneration system on its own. This choice enabled us to keep experimental instructions readable. Because of our within-design, only a limited risk was however incurred in the P4P condition. Ensuring a minimum gain (by having a back-up FFS or CAP payment system) has made our analysis comparable to what has been done earlier.

uses the information related to the priority sentence to choose the best course of action.

2.3 Results

We focus on two issues to study the role of physicians' incentives. 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?

2.3.1 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 table 2.4, we report the decision to buy the information on priority sentences by payment mechanism.

Table 2.4 – Decision to buy information and payment mechanisms

Decisions and Payment systems	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 2.4 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 is further investigated using a Probit model. We hypothesize 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 preference 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 2.5 reveal that there is a positive and statistically significant association between the purchase decision and the preference of physicians for innovative technologies (Reference for interpretation: very likely). Other variables are 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 2.5 – Variables affecting the decision to buy information on priority sentences

	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)
TECHNO *Weakly*	0.838^{*} (0.450)
Controls included	*Yes*
Constant	-2.526 (1.610)
Observations	190
Log Likelihood	-110.488
Akaike Inf. Crit.	238.976
Notes:	***Significant at the 1 percent level.
	**Significant at the 5 percent level.
	*Significant at the 10 percent level.

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.

2.3.2 Result 2: Access to personalized medicine 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 generate similar outcomes in terms of number of interventions, while FFS and P4P yield 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.

2.3.2.1 Free access to personalized medicine 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 it's common, we first report descriptive statistics and complement these by estimating an econometric model to provide further evidence. The econometric model is the following:

$$y_{itT} = c + \alpha_i + \beta_i Pay_{iT} + \gamma_i INFO_{it} + \theta_i Pay_{iT} * INFO_{it} + \Theta_i X_i + \epsilon_{itT} \tag{2.1}$$

In equation (2.1):

- 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, 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. We provide balance table in appendix that shows differences in these skills across sessions. We therefore control for these differences (see table 2.14 for performance at high school; table 2.12 for the self-declared writing skills and table 2.13 for age).

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.

2.3.2.2 Free access to personalized medicine information and physicians' lack of focus (degree)

The focus variable allows us to capture how physicians orient their intervention with the informational tool at their disposal. We measure focus 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 the extent to which the physician focuses on the patient's problem. Figure 2.4 and Table 2.6 present both descriptive statistics and results of our estimation.

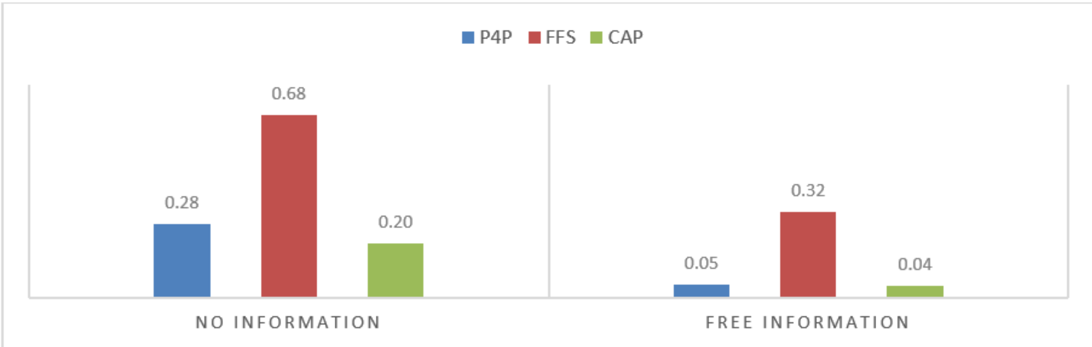


Figure 2.4 – Free access to information and physicians' degree of focus.

Table 2.6 – Impact of free information and payment mechanisms on degree of focus

	Focus (rate of interventions outside priority sentences)		
	Panel linear		
FFS (Ref: P4P)	0.341*** (0.036)	0.303*** (0.047)	0.371*** (0.056)
CAP (Ref: P4P)	−0.039 (0.036)	−0.048 (0.036)	−0.081* (0.047)
INFO (Ref: No info)	−0.244*** (0.025)	−0.244*** (0.025)	−0.227*** (0.035)
Controls included?	*No*	*Yes*	*Yes*
INFO in the FFS payment system			−0.136** (0.060)
INFO in the CAP payment system			0.067 (0.059)
Constant	0.284*** (0.025)	0.511* (0.293)	0.503* (0.294)
Observations	190	190	190
R^2	0.522	0.527	0.549
Adjusted R^2	0.514	0.509	0.527
F Statistic	67.603*** (df = 3; 186)	29.022*** (df = 7; 182)	24.388*** (df = 9; 180)

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Remember that correcting words outside priority sentences is not of any benefit to patients.[5] When information is available, the degree of focus as measured by the

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

intervention rate outside priority sentences is lower whatever the payment scheme (columns 1 and 2 of Table 6, variable '*INFO*'; and Figure-2.4-histogram bars in "No information" *vs* bars in "Free information"). From Figure 2.4, 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 2.6 and Figure 2.4). Thus, we can conclude that physicians really used information to focus their interventions on the priority sentences.

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 2.6, column 3). Despite their financial incentives, when physicians have access to personalized medicine information, they reduce their interventions outside priority sentences even though their income is increasing in the number of interventions. In the context of FFS, this result clearly reveals that our physicians have a more complex objective than mere profit maximization: they are behaving altruistically.[6]

2.3.2.3 Free access to personalized medicine information and physician' 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 2.7 and Figure 2.5. Table 2.7 only presents estimation results for the ratio of well-treated to treated texts. Econometric results on the well-treated texts are very similar.

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

[6] 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).

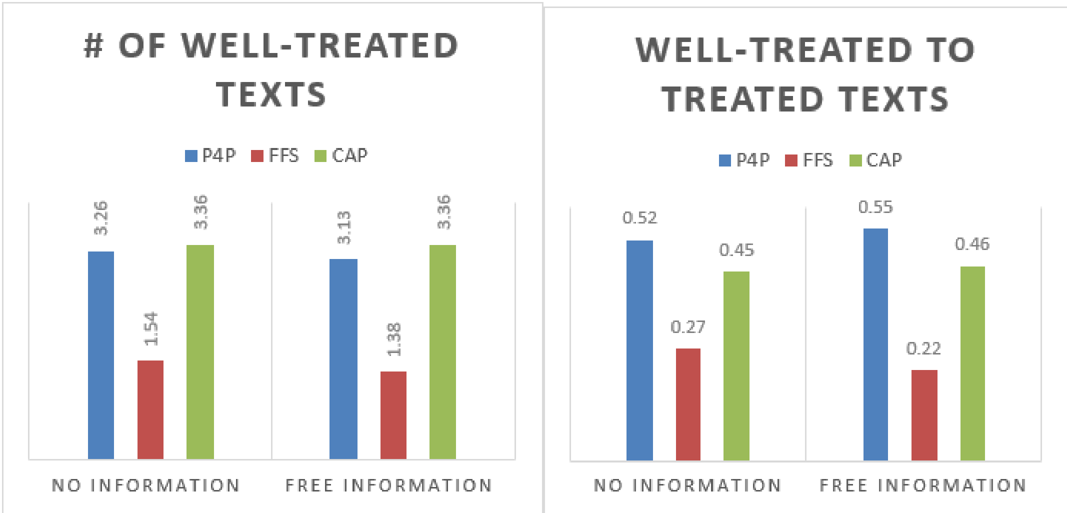


Figure 2.5 – Free information and physicians' number and rate of well-treated texts

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 2.7 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.

2.3.3 Paid access to personalized medicine information and physicians' qualitative outcomes

Access to information on priority sentences is available but has to be paid for in period 3. Due to the impact on benefits that we introduce, many factors might 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 model 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 information buyers and non-buyers, we provide estimates of a 2-stage Instrumental Variable Probit model. The Probit model estimated in the "first step"

Table 2.7 – Impact of free information and payment mechanisms on number of well-treated texts

	rate of well-treated texts	
	Panel linear	
FFS (Ref: P4P)	−0.297*** (0.048)	−0.263*** (0.063)
CAP (Ref: P4P)	−0.083* (0.047)	−0.073 (0.048)
INFO (Ref: No info)	0.003 (0.027)	0.003 (0.027)
Controls included?	*No*	*Yes*
Constant	0.536*** (0.031)	0.130 (0.389)
Observations	190	190
R^2	0.172	0.183
Adjusted R^2	0.159	0.151
F Statistic	12.912*** (df = 3; 186)	5.811*** (df = 7; 182)

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

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 the tables below are second-step regression results and corrected for the endogeneity of the decision to buy information on priority sentences.[7]

Formally, we estimate the following set of equations:

[7] 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.

$$BUYINFO_{iT} = c + \mu_i Pay_{iT} + \eta_i TECHNO_i + \rho_i X_i + v_{iT} \tag{2.3.3}$$

$$y_{i,T} = c + \alpha_i + \beta_i Pay_{i,T} + \gamma_i \widehat{BUYINFO}_{iT} + \theta_i Pay_{iT} * \widehat{BUYINFO}_{iT} + \Theta_i X_i + \epsilon_{itT} \tag{2.3}$$

Equation (2.3.3) was previously estimated and results provided in Table 4. In equation (2.3) :

- 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 pure 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 preference 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 2.5). Our second argument for using this instrument is the fact that it does not affect our outcome variables (focus of actions, rate of well treated patients). We also check that we don't have potential confounding factors that could affect both the instrument and the outcome variables to be sure that the exclusion restriction is verified. We have checked for risk aversion and we have also used a proxy of self-confidence. Both are not linked neither to our instrument nor to our dependent variables.

2.3.3.1 Paid access to personalized medicine information and physicians' lack of focus (degree)

We summarize our results on focus in Figure 2.6 and Table 2.8.

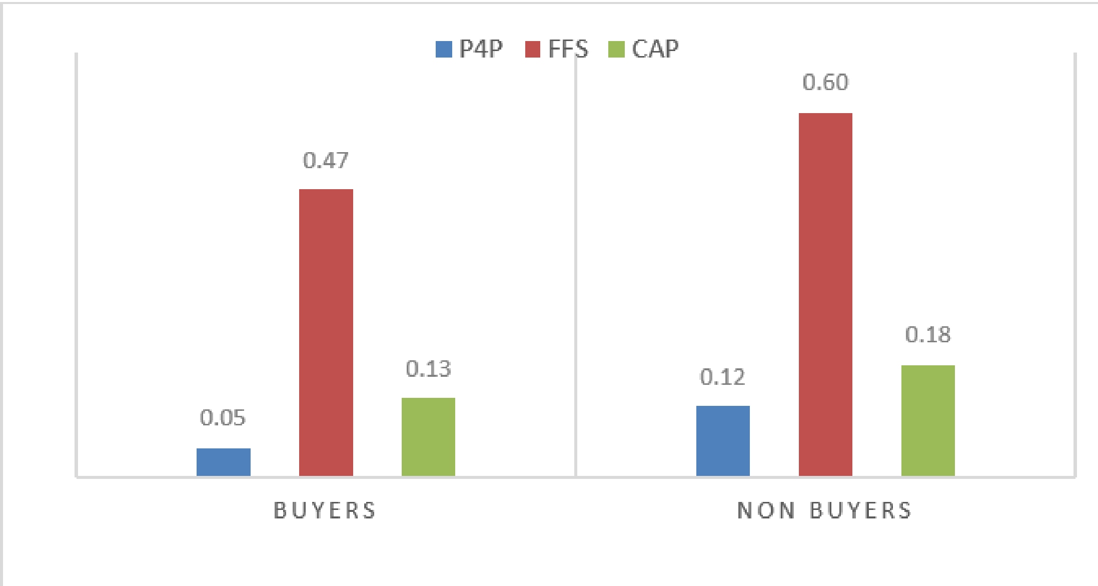


Figure 2.6 – Free information and physicians' rate of actions outside of priority sentences

When considering the effect of information, the 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 2.8 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.[8] A commitment device effect is consistent with the fact that the rate of intervention outside priority sentences decreases by 0.14% when access to personalized medicine is free (column 3 of Table 2.6), whereas under paid access to

[8]This behavioral effect should be observed in CAP but is less visible in our data. In table 2.8, 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).

Table 2.8 – Impact of buying information and payment mechanisms on the degree of lack focus

	Focus (rate of interventions outside priority sentences)		
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)
Control 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)

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

information, we observe a decrease of 1.4%. It appears that our physicians tend to make better use of information when they had to pay for it.

2.3.3.2 Paid access to personalized medicine information and physicians' treatment of texts

Results on the two other indicators of quality are summarized in Figure 2.7 and Table 2.9. In Table 2.9, an interesting difference appears for the percentage of well-treated texts: 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 2.9). 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 2.10 summarizes all our descriptive results on the qualitative variables.

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 is accessible, and whether this access is free or has to be bought. This table shows that P4P and CAP generate very similar incentives, except for the focus variable, where P4P does better

Figure 2.7 – Paid access to information and number and rate of well-treated texts

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: should the access to personalized medicine be free of charge or paid?

Table 2.9 – Impact of buying information and payment mechanisms on ratio of well-treated to treated texts

	Ratio of well-treated to treated texts
	OLS
FFS (Ref: P4P)	0.001 (0.089)
CAP (Ref: P4P)	0.121 (0.079)
BUYINFO (Ref: Non-buyers)	0.391* (0.206)
Constant	−0.187 (0.278)
Observations	190
R^2	0.147
Adjusted R^2	0.114
Residual Std. Error	0.237 (df = 182)
F Statistic	4.482*** (df = 7; 182)
Notes:	***Significant at the 1 percent level.
	**Significant at the 5 percent level.
	*Significant at the 10 percent level.

Table 2.10 – Payment scheme ranking according to information structure

	No info	With free info	With bought info (comparison of buyers)	With bought info (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

2.4 Quantification exercise: 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 can yield a social benefit: physicians better exploit information they had to pay for.[9] 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 informa-

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

tion (because of the commitment-device effect described earlier), but the drawback of limiting access to information to those physicians who are 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 the degree of lack focus, as it is the only variable for which interactions with payment schemes often appeared significant. We will also limit our simple analysis to P4P and FFS. Capitation does not allow for comparisons, because there is not enough variation between free and paid access to information.

We suppose that we have a community made 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:

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

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 give: [10]

$$\Delta H^P = B^P.\Delta h^P(Info = 1) + \Delta B^P * \Delta h^P(Info = 1) - h^P(Info = 0)]. \qquad (2.2)$$

The complete effect of charging for information is given by equation (2.2). 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:

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

$$== B^{FFS} * \text{(slopes in Table 2.8)} + \Delta B^{FFS} * \text{ (slope in Table 2.6 for the var. Info)}$$
$$== 9 * (-1.414) - 40 * (-0.227 + -0.136)$$
$$= +1.794.$$

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

$$== B^{P4P} * \text{(slopes in Table 2.8) } + \Delta B^{P4P} * \text{(slope in Table 2.6 for the var. Info)}$$
$$== 55 * (0) - 40 * (-0.227)$$
$$= +9,08.$$

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.

2.5 Concluding remarks

This article reports results from an experiment on physicians' incentives to use personalized medicine techniques. Our experimental design uses 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 modify 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. We thus recreate the fundamental trade-offs of an agent (the physician) deciding on access to an informational technology like personalized medicine. This framework not only allows 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 an 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 find 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 is 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 find 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 rate 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 conclude that this is consistent with a "commitment device". Using a simple quantification framework to assess the consequences of a generalized paid access to personalized medicine, we find that charging for information is not desirable in P4P and FFS. While the experiment provides 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.

2.6 Appendix

2.6.1 Appendix 1: Summary of all the data on texts presented to subjects

Table 2.11 – Summary of the random assignment of texts to periods and payment mechanisms

	Highlighted words	Wrong words	Correct words	Mistakes PS	Correct words PS
Period 1	92	31	61	24	44
FFS or CAP	42	14	28	12	21
P4P	50	17	33	12	23
Period 2	90	30	60	21	48
FFS or CAP	45	15	30	10	25
P4P	45	15	30	11	23
Period 3	92	32	60	23	46
FFS or CAP	48	16	32	11	24
P4P	44	16	28	12	22

PS: priority sentences.

2.6.2 Appendix 2: Balance checks on our control variables

In tables below, NS stands for "Non-significant" significance at a 5% level of a t-test.

Table 2.12 – Variable Level in spelling ("note en dictée")

	Session 1	Session 2	Session 3	Session 4	Means
Session 1					12.71
Session 2	NS				12.81
Session 3	***	***			10.08
Session 4	NS	NS	NS		11.44

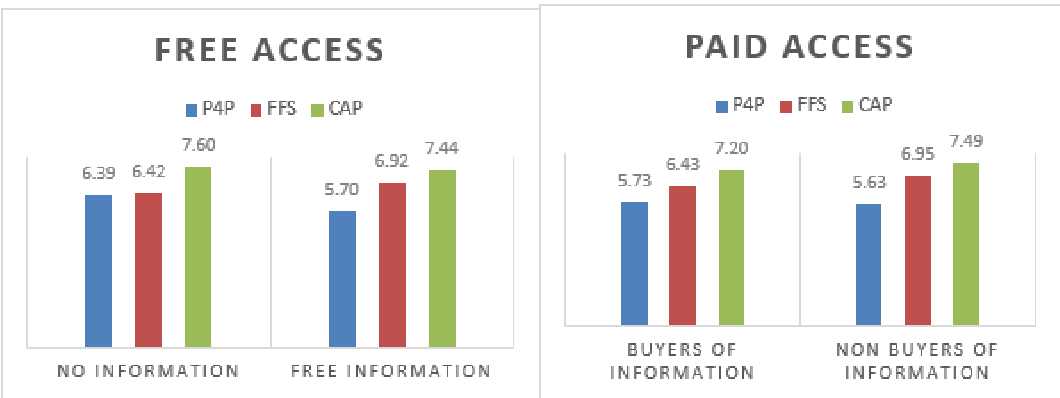
Table 2.13 – Variable age

	Session 1	Session 2	Session 3	Session 4	Means
Session 1					20.42
Session 2	***				23.19
Session 3	***	***			22.4
Session 4	***	NS	NS		22.6

Table 2.14 – Variable Grade at the end of high school ("note au Bac")

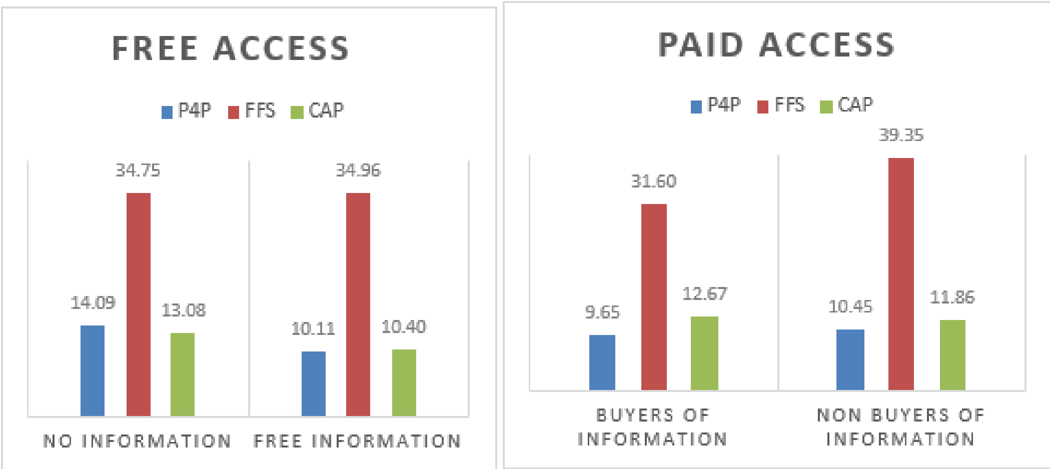
	Session 1	Session 2	Session 3	Session 4	Means
Session 1					15.60
Session 2	***				13.92
Session 3	NS	***			15.24
Session 4	***	NS	NS		14.42

2.6.3 Appendix 3: Summary of quantitative results



Free access to information and physicians' number of texts treated

We summarize all our results dealing with these two quantitative variables in the following table



Free access to information and number of services

Table 2.15 – Payment scheme ranking according to information structure

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

2.6.4 Appendix 2: Experimental protocol (free translation).

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.

If you have any question at any time during the experiment, please raise your hand. This first part is composed of 3 periods. 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 the priority sentences are, 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:

- 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.

- 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.

- 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:

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.

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

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 (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 deduct some expenditures for the period depending on your choice.
  - 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.

Your earnings (Payment per action) – FFS –

You will receive in this part a payment per action. This “Per-action Payment” is a fixed remuneration per action on highlighted words, regardless of whether these corrections are appropriate.

- Earnings in period 1 = €0.30 * Total number of highlighted words on which you have worked.
- Earnings in period 2 = €0.03 * Total number of highlighted words on which you have worked.
- Earnings in period 3 = €0.30 * Total number of highlighted words on which you have worked. From this sum we will deduct some expenditures for the period depending on your choice.
  - 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.

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.

Bejarano, H., Green, E.P., Rassenti, S., 2017. Payment scheme self-selection in the credence goods market: An experimental study. J. Econ. Behav. Organ. 142, 396–403.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies, N., Wiesen, D., 2013. How Effective are Pay-for-Performance Incentives for Physicians?-A Laboratory Experiment.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., Wiesen, D., 2017. The effects of introducing mixed payment systems for physicians: experimental evidence. Health Econ. 26, 243–262.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., Wiesen, D., 2016a. Physician performance pay: Evidence from a laboratory experiment. Ruhr Economic Papers.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., Wiesen, D., 2016b. Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision. J. Econ. Behav. Organ. 131, 17–23.

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., Hennig-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

Chapter 3

Physician responses to previous exposure to P4P incentives: Experimental evidence

We study short term effect on quantity and quality of care of a withdraw of Pay-For-Performance (P4P) incentives. We experimentally mimic the physician-patient relationship by using a text proofreading assignment with prospective physicians as the subject pool. Physicians' exposure to P4P incentives is randomly decided to be either anterior or posterior to Fee-For-Service (FFS) or CAP (Capitation) payment systems. We compare "treated physicians" (those exposed to P4P incentives before FFS or CAP payment systems) to "control physicians" (those exposed to FFS or CAP payment systems in the first place). P4P is constructed to remunerate only proofreading actions that are beneficial to the patient. We find that previous exposure to P4P incentives increases physician's focus on patient needs: In short, FFS and CAP physicians are less likely to randomly allocate their effort when they are previously exposed to a P4P incentive. This positive effect of exposure to P4P incentives is however offset by a decrease in quality observed for treated physicians in CAP, which we interpret as a decrease in intrinsic motivation. Policy makers should be aware of this potential tension between focus on patient needs and erosion of intrinsic motivation when designing and promoting P4P approaches.

3.1 Introduction

We use an experimental design to study how previous exposure to pay-for-performance (P4P) incentives later affect physicians' quantity and quality of care. The way physicians respond to financial incentives is a relatively new and burgeoning research area in experimental health economics. Incentive properties of more-classic payment systems like fee-for-service, capitation, salary and fund-holding are well documented. It is less the case for the effect of P4P schemes where robust evidence is rare, due to the methodological challenge of the identification strategy. Many countries have in fact attempted to reform their payment systems by introducing performance pay schemes for general practitioners, on top of their usual payment parameters: Examples include the U.S "Premier Hospital Quality Incentive Demonstration" program (started in 2003), the Germany "Disease Management Program" (started in 2002), the French "Contrat d'Amelioration des Pratiques Individuelles (CAPI)" (2009) later replaced by the "Remuneration sur Objectifs de Sante Publique (ROSP)" (2012) and the United Kingdom "Quality and Outcomes Framework" (started in 2004), the latter being probably the most known and most studied. This tendency of attempting P4P reforms is probably inspired by the overall positive effect of P4P initiatives in the education sector (Neal, 2011). In the health economics literature however, there is a need to document how withdrawing P4P indicators may later affect physicians' quantity and quality of provided care.

We use an experimental procedure where some prospective physicians first experience the P4P system and later are exposed to either FFS or CAP conditions. We refer to this latter group as the "treated physicians". We compare their care quantity and quality decisions to those of a "control" group of physicians who rather first experience FFS or CAP conditions. The FFS condition rewards the physician for every single action taken, regardless of whether it is appropriate for the patient. The CAP rewards the physician for each patient for whom at least one action is taken. The P4P condition imposes a threshold of correct actions for the physician to get a reward for the enlisted patient.

Our experiment uses prospective physicians as the subject pool. In the experiment, each physician is always observed under two different conditions, each representing an institutional payment system (either P4P, FFS or CAP). The order in which physicians are likely to begin with a "P4P" condition is randomly decided: Two groups of physicians have P4P as their first experimental condition (therefore having either CAP or FFS as the second experimental condition), while two other groups begin with either FFS or CAP and later experience P4P as the second experimental condition. We exploit this differential and random exposure to P4P payments to study how quantity and quality of care are later affected by a withdrawal of P4P incentives. In fact, a “control” group is provided by physicians who did not have P4P as their first experimental condition.

We use a text proofreading assignment to mimic the physician-patient relationship. The prospective physician is the proofreader. In the design, the highlighted words to proofread represent all the possible patient symptoms, as expressed by patients (highlighters). A feature of our design is that the physician knows the selected areas in texts where his correct actions will always maximize potential patient health benefits. If the appropriate proofreading of words in these areas is always benefit enhancing for patients, it is profit maximizing for physicians only depending on the payment plan. The P4P system is experimentally designed to push physicians to make appropriate actions in these areas. Therefore, it is also profit maximizing to take appropriate actions in the selected areas that maximize patient benefits. However, in the CAP condition, the physician’s gain is independent of the location and the appropriateness of his actions. Furthermore, he is economically incentivized to make at most one action. Lastly, in the FFS condition, it is rather profit maximizing to proofread words everywhere in the text, including in areas that are not indicated as being relevant for patients. This design therefore leaves us with the the possibility to capture insights on non-financial motives in physicians’ decisions and we can also assess physicians’ capacity to focus on what is relevant for patients prior and post-P4P incentives.

This question is linked to different strands of the literature in health economics. First, our paper is related to the research demonstrating the importance of non-monetary incentives for physicians. The large part of this literature focuses on physician altruism, its relevance and its distribution in the population. The main take-away here is that altruism matters and physicians are heterogeneous with respect to altruism: Empirical evidence on physician altruism in the literature can be gathered depending on the methodology of the study (Galizzi et al., (2015)). There are studies that use surveys and interviews (see Allaby, (2003); Pawlikowski et al., (2012) among others); others that rely on discrete choice experiments (see Scott, (2001); Rizzo and Zeckhauser, (2003); Scott and Sivey, (2017) among others); another group that uses field experiments (Kolstad, (2011); Serra et al., (2011); Smith et al., (2012)); and a last group (the most recent one) made of studies that have used laboratory experiments to touch upon physician altruism (Hennig-Schmidt et al., 2011; Brosig-Koch et al., 2013; Godager and Wiesen, 2013; Kesternich et al., 2015; Green, 2014; Godager et al., 2016; Brosig-Koch et al., 2016b, 2016a, 2017)

Our paper directly relates to this last part of the literature that has used laboratory games as a way to study physicians’ behaviors. There are many take-away from this literature. One of them is the fact that the subject pool matters for the study of physician altruism. Hennig-Schmidt and Wiesen, (2014) for example find that non-medical subject pool tends to behave less altruistically than medical subject pool. This finding has inspired our choice of using “prospective physicians” as subject pool for this experiment. Another element that is worth highlighting from this literature is the kind of tasks that experimenters have used to study physician payment systems. The majority of these scholars has used chosen effort experiments where only “intention” to make effort is collected (Brosig-Koch et al., 2013, 2016b, 2017; Godager et al., 2016; Godager and Wiesen, 2013; Hennig-Schmidt et al., 2011; Hennig-Schmidt and Wiesen, 2014). The particularities of this kind of experiments are the fact that physician choices are hypothetical *per se* and profit functions (benefit functions) of each physician (each patient) are convex (concave). Closed to this kind of experimental procedures, has also emerged real effort experimental designs (Green, 2014; Lagarde and Blaauw, 2017; Bejarano et al., 2017; Bardey et al., 2018) where the physician’s effort is inferred based on his actions during the game. We use a proofreading assistance game where prospective physicians correct highlighted words in texts. This game was first proposed by Green, (2014) and has also been used by Bejarano et al., (2017) and Bardey et al., (2018). The game is suitable to mimic the physician-patient relationship.

A second strand of the related literature assesses the effectiveness of P4P initiatives. Rosenthal and Frank, (2006) propose a review of empirical evidence on P4P and Emmert et al., (2012); Eijkenaar et al., (2013) propose systematic reviews. There is room to improve the efficiency of health systems and one way in which many countries have moved is to introduce P4P reforms for physicians. Evaluations of these initiatives in different countries conclude to modest (if any) impact (see Sicsic et al., (2012); Saint-Lary and Sicsic, (2015); Herbst et al., (2018) among others). We remind that, in FFS, CAP, salary, fundholding systems, etc., the reward of quality is not as explicitly incentivized as it is in P4P systems. P4P has been introduced in many countries on top of a “base” payment system already in place (mainly FFS in France for example and mainly CAP in the U.K). The literature identifies at least two challenges that can be faced by a regulator introducing a P4P reform. First of all, rewarding certain dimensions of quality can be detrimental to other dimensions of quality that are unrewarded, resulting in an overall lower level of quality. Secondly, explicitly paying for performance may deteriorate physician intrinsic motives (Brosig-Koch et al., 2016a).

For the first point on the rewarded/unrewarded dimensions of quality, Mullen et al., (2010) and Li et al., (2014) have evaluated the effectiveness of P4P.[1] The Li et al., (2014) study exploits a natural experiment in the province of Ontario, Canada, to identify the impact of pay-for-performance (P4P) incentives on the provision of targeted primary care services. The results indicate that responses are modest. Mullen et al., (2010) have used data from the performance reports of medical groups contracting on a capitated basis with a large network HMO in California to study the effectiveness of P4P. They use performance reports of medical groups before and after the implementation of two P4P programs in California. The authors compare the performance of these groups to other medical groups that were not affected by either program. They fail to find evidence that a large P4P initiative either results in major quality improvement or notable disruption in quality. Our work relates to these two studies mainly to the last one. In terms of design, we are close to Mullen et al., (2010), as we are able to identify a control group of physicians (those who have not experienced a P4P initiative, as they had been paid only by the traditional base payment system, FFS or CAP). We can investigate how their behavior differs from that of the treated group of physicians who have been confronted to the P4P.

For the second point on how P4P may affect physician's motivations, previous studies report the "hidden" costs of incentives (Ellis and McGuire, 1986; Kreps, 1997; Chalkley and Malcomson, 1998; Jack, 2005; Benabou and Tirole, 2006; Sliwka, 2007; Maynard, 2012). The message of this literature is that introducing "prices" (A P4P approach being perceived as a price) in the public sector may crowd-out people's intrinsic values for quality. For the study of physician behaviors, Siciliani, (2009), Brosig-Koch et al., (2013) and Brosig-Koch et al., (2016b) study the impact of paying physicians using a P4P system. Siciliani, (2009) uses a theoretical approach while Brosig-Koch et al., (2013) uses an experimental approach and Brosig-Koch et al., (2016b) a field experiment. Both papers find evidence of a crowding-out of intrinsic motivation. Our work relates to this literature that seeks to evaluate if crowding out of physician intrinsic motivation exists.

The remaining of the paper is organized as follows: The design is presented in section 2. Results are reported in section 3. We briefly conclude in section 4.

[1] There are other papers on rewarded/unrewarded dimensions of quality, those suggesting that physicians game with P4P indicators (see Gravelle et al., (2010); Clemens and Gottlieb, (2014) among others).

3.2 Design

Our data comes from the real effort experimental game described in chapter 2. In that chapter, we analyze how physician incentives affect their likelihood of adopting, for free or possibly for a cost personalized medicine techniques. Main features about the design to be kept in mind are the following:

Reminder 1: The design builds on earlier work using real effort tasks in experimental studies on physicians' behaviors.[2] We used a 2-Phase game to mimic the physician-patient relationship. In Phase 1, the patient declares his symptoms (by highlighting words in texts) and in Phase 2, the physician provides proofreading assistance under different payment conditions.[3]

Before organizing phase 1, we have first selected 48 short texts and choose for each text, a series of candidates words for potential proofreading in Phase 2. Those "candidate words" comprise both truly misspelled words and words that are well written. In all texts, we underline an area of "priority sentences", such that only physicians' efforts in these sentences are effective for the patient welfare. Phase 2 is the most important part of this experiment. Two sessions are organized with P4P as the first payment system and two others with P4P as the second payment system (the other payment parameters being either FFS or CAP).

Reminder 2: For each payment system, the game is displayed in three periods of 5 minutes each as follows:

1. A *Period 1* in which prospective physicians are proposed 8 texts with only highlighted words to proofread;

2. A *Period 2* in which prospective physicians are proposed 8 texts with highlighted words and indications on the priority sentences to proofread: underlined sentences visible;

3. A *Period 3* in which prospective physicians are first invited to make a choice between two sets of 8 texts per set: in one set, it is possible to see highlighted words only (similar to period 1) and in the other set it is possible to see both highlighted words and underlined sentences (similar to period 2). The choice of texts having both underlined and highlighted words is associated with a little reduction on final gains.

[2] See Bejarano et al., (2017); Lagarde and Blaauw, (2017); Green, (2014)

[3] Green (2014) and Bejarano et al., (2017) have used a similar game to study physician behaviors (Green, 2014) and payment self-selection in credence good markets (Bejarano et al., 2017).

Reminder 3: In our experiment, P4P is either anterior or posterior to FFS or CAP payment. In the FFS treatment, the physician receives €0.30 per highlighted word that he corrects, regardless of whether his correction is appropriate. In the CAP treatment, he receives €1.75 per text on which he makes at least one correction, regardless of whether it is correct. In the P4P treatment, the physician receives €2.50 if by the end of his intervention, the text is such that, 80% of words in underlined sentences are appropriately written. If that's not the case, then the physician does not receiving a payment for the text. The gains of the subject sum up between the two parts, with one paid period being chosen at random for each payment condition. Prospective physicians have earned throughout the experiment a minimum and a maximum amount of €6.20 and €35.40 respectively.

Reminder 4: To introduce other-regarding behaviors, each prospective physician is informed that he can generate €5 for the patient (who first highlighted the words in yellow) if he corrects appropriately all the wrong words in the priority sentences (we allow a liberty of one wrong word). If the threshold of wrong words to generate €5 is not attained, the patient does not receive anything. Between 1 and 4 wrong words (depending on the text) are needed to be corrected to generate €5.

In this paper, our focus will only be on physician behaviors in Period 2 of the game in FFS and CAP payment systems. Behaviors observed in these periods for each payment scheme are sufficient to address our research question. Only period 2 is chosen because behaviors are unambiguous in this period, compared to 1 and 3 where they are: in Period 1, ambiguity comes from the fact that the physician does not know the "priority sentences", and in period 3 the introduction of a choice creates an endogenous decision which will bias a direct between-subject comparison.

3.3 Results

3.3.1 Question 0: Quality of the randomization

Our data has an original feature which makes it possible for us to investigate our research question: the design allows us to have a "treated" group and a "control" group. The treated group receives the P4P in part 1 of the experiment; the counterfactual of their behavior is provided by the two other groups that are not treated by a P4P scheme in Part 1 (those who are "treated" by Capitation or Fee-For-Service). Table 3.1 summarizes the order of exposure to P4P by group:

One confounding factor that needs to be accounted for is the "learning effect" or the "fatigue effect". We test for these effects by comparing behaviors in Part 1 with

Table 3.1 – Order of exposure to P4P

	Part 1	**Part 2**
Control group	CAP or FFS	P4P
Treatment group	P4P	CAP or FFS

those in Part 2 for all the subjects of our experiment in all the payment systems that we study for this experiment. Note that, Table 3.2 that summarizes our comparison is not limited to Period 2 data, as below:

Table 3.2 – Comparison of the percentage of actions per payment system

	Part 1	**Part 2**	**P-value**
P4P	0.32	0.32	0.71
FFS	0.42	0.43	0.67
CAP	0.21	0.21	0.32

Table 3.2 is built for the percentage of action observed in all the payment systems, depending on whether it is proposed in Part 1 or in Part 2. It shows unsurprisingly that FFS is associated with the highest percentage of action, followed by P4P and CAP in the last position. What is of interest for us in this table it is the fact that there is not a systematic difference between Part 1 and 2 for each payment method.

Our proxy of "work intensity" is not showing a pattern between the Part 1 and 2. We can confidently study how P4P explains differences between control and treated physicians in Period 2 of the game, as this is the only varying factor between subjects. Our main results presented below will cover one indicator related to work intensity (quantity) and another one related to the quality of care.

3.3.2 Question 1: Does exposure to P4P impact physicians' quantity of care? How different is the effect under CAP and FFS?

One way in which one would report on physician behaviors is by looking at the quantity of care services that they offer. The quantity of services captures how many actions the physician takes in his attempt to treat the patient. We proxy the quantity of services in our experimental data by using the overall number of corrections. Remember that, the only way in which the physician could generate gain for the patient is by making corrections on wrong words.

To report on how exposure to P4P incentives affects physician quantity of services, we first show a histogram comparing the number of corrections for the treatment and control group in Period 2 (Figure 3.1), and a table summarizing average comparisons for the number of corrections (Table 3.3).

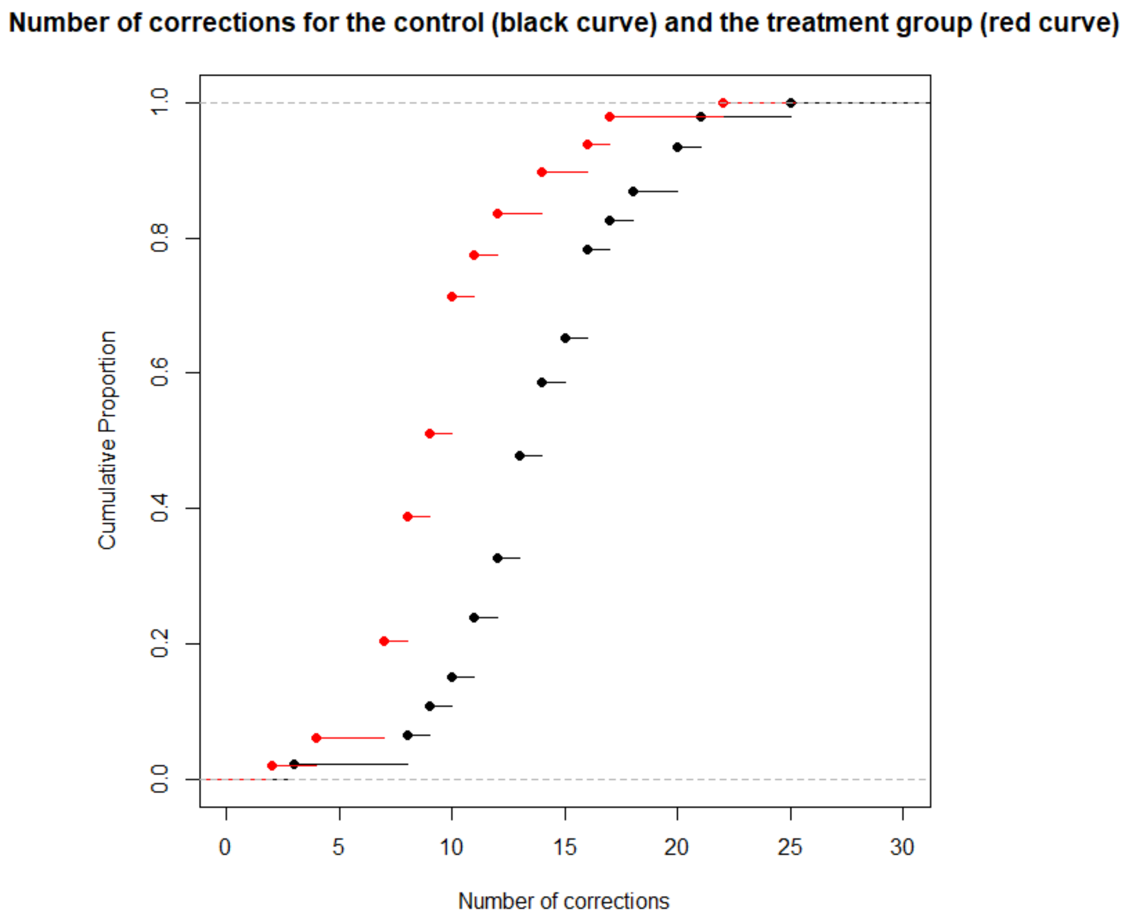


Figure 3.1 – Histograms showing the cumulative distributions of the number of corrections for the control and the treatment group.

The histogram in Figure 3.1 shows the cumulative distributions of the number of corrections for control (back curve) and treated (red curve) physicians. From the histogram, we can see a dominance of the curve of the cumulative distribution

Table 3.3 – Summary of mean comparisons for the number of corrections

.

Treatment group	Average	Control group	Average	Difference	Pvalue
P4P – FFS/CAP	9,878	FFS/CAP – P4P	14,087	29.9%	0.00
P4P – CAP	9,520	CAP – P4P	13,619	30,1 %	0.00
P4P – FFS	10,250	FFS – P4P	14,480	29,2 %	0.00

of actions of the control group, over that of the treatment group. The quantity of actions therefore seems to be lower after the withdrawal of P4P incentives. The modal behavior is in the interval [10-15[corrections in the control group and it falls in the interval [5-10[actions in the treatment group. Control physicians offer in average 14.09 corrections, *vs.* 9.88 words for treated physicians (Table 3.3), with the difference being highly significant. In average, there is therefore a 30% decrease in the number of corrections after the withdrawal of P4P incentives.

We adjust overall behaviors estimating Ordinary Least Squares (OLS) models. First, we estimate a general model that has the order of exposure to P4P incentives as an explanatory variable, controlling for gender and age. Secondly, we also estimate models for clusters of CAP and FFS physicians separately. Results are reported in Table 3.4.

The table shows that, overall treated physicians do less corrections than control physicians (those who have FFS or CAP in the first place) (column 1, Table 3.4). On average, the former group does 4.2 less actions than the latter. This effect was already seen in the histogram and the table presented above.

We decompose this overall effect for the sub-groups of FFS and CAP physicians in models 2 and 3 (columns 3 and 4, Table 3.4). It appears that it is approximately the same order of difference: CAP physicians in the control group, perform in average 3.88 actions more than CAP physicians in the treatment group, while FFS physicians in the control group perform 3.55 actions more than those in the treatment group. Differences observed in Table 3.3 persist, after controlling for gender and age. These differences are interesting because we seem to observe the same effect, with overall the same order of magnitude even though subjects are under remuneration systems carrying different incentive properties. We remind that we are in Period 2 where the physician sees the areas where his effort is likely to maximize patient benefits.

Table 3.4 – Physician number of corrections for the control and treatment groups

	Number of corrections		
		OLS	
	Pooled data(1)	CAP only (2)	FFS only (3)
Treated group P4P first (dummy)	-4.114^{***}		
	(0.819)		
Treated group CAP second (dummy)		-3.881^{***}	
		(0.912)	
Treated group FFS second (dummy)			-3.551^{**}
			(1.644)
Age	-0.143	0.017	0.175
	(0.243)	(0.328)	(0.526)
Sex	-2.261^{***}	-1.418	-3.325^{**}
	(0.769)	(0.928)	(1.340)
Constant	18.187^{***}	13.703^{*}	11.897
	(5.538)	(7.710)	(11.614)
Observations	95	46	49
R^2	0.307	0.362	0.293
Adjusted R^2	0.284	0.317	0.246
Residual Std. Error	3.676 (df = 91)	2.996 (df = 42)	4.268 (df = 45)
F Statistic	13.449^{***} (df = 3; 91)	7.949^{***} (df = 3; 42)	6.225^{***} (df = 3; 45)
Notes:	***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.		

Regardless of the payment system, this first result suggests that withdrawal of P4P incentives has short term consequences on the quantity of care given. Thanks to our design, we are able to better understand the origins of these differences between control and treated physicians. The rationale for having control-physicians providing more quantity care than treated physicians differs according to the payment condition. In the FFS system, this lower care quantity for treated physicians is likely an evidence that physicians are probably oriented towards doing more quality after the withdrawal of a P4P incentive. Remember that, the optimal strategy in FFS is to provide as many services as possible in the allocated time-frame. All these provided services are beneficial for the patient only if they are correct and appear in a given area. Physicians exposed to a remuneration system that explicitly rewards their actions in the best patient interest reduce the FFS-embedded tendency to make too much actions. We will assess in the next section whether this decrease in quantity has resulted in overall higher quality. In the CAP system on the other hand, the significant difference between groups on the quantity of their given care is not straightforwardly explainable as for physicians under the FFS system. In CAP in fact, the physician is not economically incentivized to make several actions in the patient interest, nor is he incentivized to make those actions in the relevant areas. The interpretation of this difference therefore should also account for the areas where actions are done.

We further compare the quantity of care services that are taken in underlined areas by physicians across payment systems. We first report on the number of useless corrections. The number of useless corrections measures the number of corrections done in non-underlined sentences (non-useful areas for the patient). These sentences are clearly unlikely to generate financial gains to the patient. Remember that in P4P, physicians are incentivized to provide zero actions on non-underlined sentences, while in CAP and FFS, they are not explicitly incentivized to do that. In CAP, they can work in this non-underlined area to generate money for themselves (as part of the minimum needed to have the CAP payment), and in FFS, they can work in this non-underlined area to maximize their payments. We provide in Table 3.5, the summary of our descriptive statistics for each of the groups:

Table 3.5 – Summary of mean comparisons for the number of useless corrections in FFS and CAP

.

Treatment group	Average	Control group	Average	difference	Pvalue
P4P – FFS/CAP	1.367	FFS/CAP – P4P	3.500	60.94%	0.00
P4P – CAP	1.840	CAP – P4P	3.048	39.6 %	0.00
P4P – FFS	0.875	FFS – P4P	3.880	77.4 %	0.00

Table 3.5 shows that treated physicians tend to do less useless actions than control physicians. Overall, there is a 61% drop in the number of useless actions after the withdrawal of P4P incentives. We estimate as previously three models and we report our results in the Table 3.6 :

There are 2 elements that are worth mentioning from Table 3.6: first, the table globally shows that physicians tend to significantly reduce the number of their “useless care actions” when they are first exposed to a P4P system (column 1, Table 3.6). The second observation from the table is that FFS control physicians perform

Table 3.6 – Physician number of useless actions for the control and treatment groups

	Number of useless corrections		
	OLS		
	Pooled data(1)	CAP only (2)	FFS only (3)
Treated group			
P4P first (dummy)	-2.133^{***}		
	(0.368)		
Control group			
CAP first (dummy)		-1.167^{*}	
		(0.581)	
Control group			
FFS first (dummy)			-3.005^{***}
			(0.609)
Age		-0.010	-0.049
		(0.209)	(0.195)
Sex		-0.319	-0.976^{*}
		(0.592)	(0.496)
Constant	3.500^{***}	3.386	5.376
	(0.264)	(4.916)	(4.303)
Observations	95	46	49
R^2	0.265	0.104	0.523
Adjusted R^2	0.257	0.040	0.491
Residual Std. Error	1.793 (df = 93)	1.910 (df = 42)	1.581 (df = 45)
F Statistic	33.577^{***} (df = 1; 93)	1.619 (df = 3; 42)	16.446^{***} (df = 3; 45)

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

more "useless" care actions compared to CAP control physicians. The reason why the positive effect of withdrawal of P4P incentives is higher for the FFS system can be explained by the fact that FFS is already associated with very high useless actions for control physicians, who have not experienced a P4P incentive. Once again, it is a profit maximizing strategy to make useless care actions in FFS.

The fact that P4P exposure also decreases useless corrections for CAP physicians is an interesting result. The rationale for the explanation is probably related to the fact that CAP physicians can proofread words in either underlined or non-underlined sentences. All the actions that are taken in non-underlined areas are useless from the patient perspective. From the physician perspective however, in CAP, useless actions can be re-classified into two categories, regarding how they relate to his earnings: (i) useless actions that are desirable from the physician perspective (in the sense that these actions allow the physician to earn his CAP gain); (ii) useless actions that are not necessarily desirable from the physician perspective (in the sense that, these actions are taken on top of actions also made in underlined areas). We capture which kind of useless actions is decreased in CAP using the two categories above mentioned. In Table 3.7, we provide the total number of corrections done in non-underlined areas, conditional on having done corrections in underlined areas (case (ii) above), and we also show the total number of corrections done in non-underlined areas only, with no actions in underlined areas (case (i) above).

Table 3.7 – Attempt to capture the importance of undesirable effort in CAP.

	Control	**Treatment**	**pvalue**
Useless actions in non-underlined areas, with no actions in underlined areas (case (i))	64	46	0.03
Useless actions in non-underlined areas, but with actions also done in underlined areas (case (ii))	16	3	0.03

Physicians work on both underlined and non-underlined words in CAP. However, the overall reduction in the number of useless corrections as mentioned above (39% of reduction, cf. Table 3.5) is more a reduction in the number of corrections done in non-underlined areas, without actions being taken in underlined areas. A withdrawal of P4P incentives seem therefore to increase the focus of actions by reducing the tendency of physicians to allocate their effort randomly in CAP. In fact, in CAP, control physicians are probably more likely to randomly decide on which word(s) they will put their effort. Exposing them to P4P incentives reinforce the importance of making actions that are beneficial for both the physician and the patient. We measure the importance of this focus for both treatment and control, by looking at the ratio of useful corrections to the total number of actions.

Table 3.8 – P4P and physician focus

	Rate of useful corrections		
	OLS		
	Pooled data(1)	CAP only (2)	FFS only (3)
Treated group			
P4P first (dummy)	0.103***		
	(0.026)		
Treated group			
CAP second (dummy)		0.050	
		(0.043)	
Treated group			
FFS second (dummy)			0.173***
			(0.034)
Age	−0.008	−0.002	−0.000
	(0.008)	(0.015)	(0.011)
Sex	0.030	0.011	0.043
	(0.025)	(0.043)	(0.028)
Constant	0.922***	0.821**	0.722***
	(0.179)	(0.361)	(0.242)
Observations	95	46	49
R^2	0.215	0.040	0.534
Adjusted R^2	0.189	-0.029	0.503
Residual Std. Error	0.119 (df = 91)	0.140 (df = 42)	0.089 (df = 45)
F Statistic	8.313*** (df = 3; 91)	0.576 (df = 3; 42)	17.177*** (df = 3; 45)

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Overall, treated physicians perform 10% more useful actions than control physicians. Treated CAP physicians perform 5% more useful actions than control CAP physicians, but this difference is not statistically significant. The focus effect seems to be driven by the FFS system where treated physicians perform 17% more useful actions than control physicians.

In conclusion, withdrawing P4P incentives tend to lead to less quantity care for patients, this difference being mainly a decrease in the percentage of useless care actions. P4P exposure has overall a positive effect on FFS physicians, as they increase the focus of their actions on what is relevant for the patient. The profit maximizing behavior that characterizes FFS physicians is therefore offset by a P4P exposure. In CAP, exposure to P4P incentives reduces the tendency that control CAP physicians might have to randomly allocate their effort.

In question 2, we focus on quality. Impact on quantity suggests that, although P4P incentives increases the focus of actions, there is still a large part of physician services which is done in the underlined area (two thirds in CAP and one third in FFS). As a reminder, correct actions in underlined areas only are taken into consideration to generate patient benefits. While in P4P the physician is explicitly incentivized to work on underlined wrong words, in FFS or CAP, it is not the case. Regardless of the payment system, the criteria to generate money for the patient is the same, that of making good corrections in the underlined area. In question 2 below, we look at how P4P incentives affect the level of quality that the physician reach.

3.3.3 Question 2: Does exposure to P4P incentives impact physician quality of actions? How does this vary for FFS and CAP physicians?

Improving the quality of care is an important objective for policy makers. This experiment allows us to study how exposure to P4P incentives affect the level of quality of care. We measure this quality by looking at the number of well treated patients. Each text corrected in an appropriate way leads to a generation of €5 to the first subject (patient). Our first variable counts the number of texts on which the physician has generated €5 out of the 8 proposed texts in Period 2.

We represent in Figure 3.2 the histogram of our quality variable for control and treated physicians and in Table 3.9, we summarize our comparison of means for control and treated groups.

Table 3.9 – Summary of mean comparisons for the number of well treated patients

.

Treatment group	Average	Control group	Average	difference	Pvalue
P4P – FFS/CAP	2.592	FFS/CAP – P4P	3.261	20.51%	0.01
P4P – CAP	2.320	CAP – P4P	3.476	33.2 %	0.00
P4P – FFS	2.875	FFS – P4P	3.080	6.6 %	0.58

The cumulative distribution does not reveal a clear domination of one of the groups, but the modal number of well treated patients is 3 for the control group and 2 for the treatment group. At this stage, all observations are pooled together for FFS and CAP. The average quantities are respectively 3.26 and 2.59 for the treated and control groups. The table above shows that there is a decreasing number of well

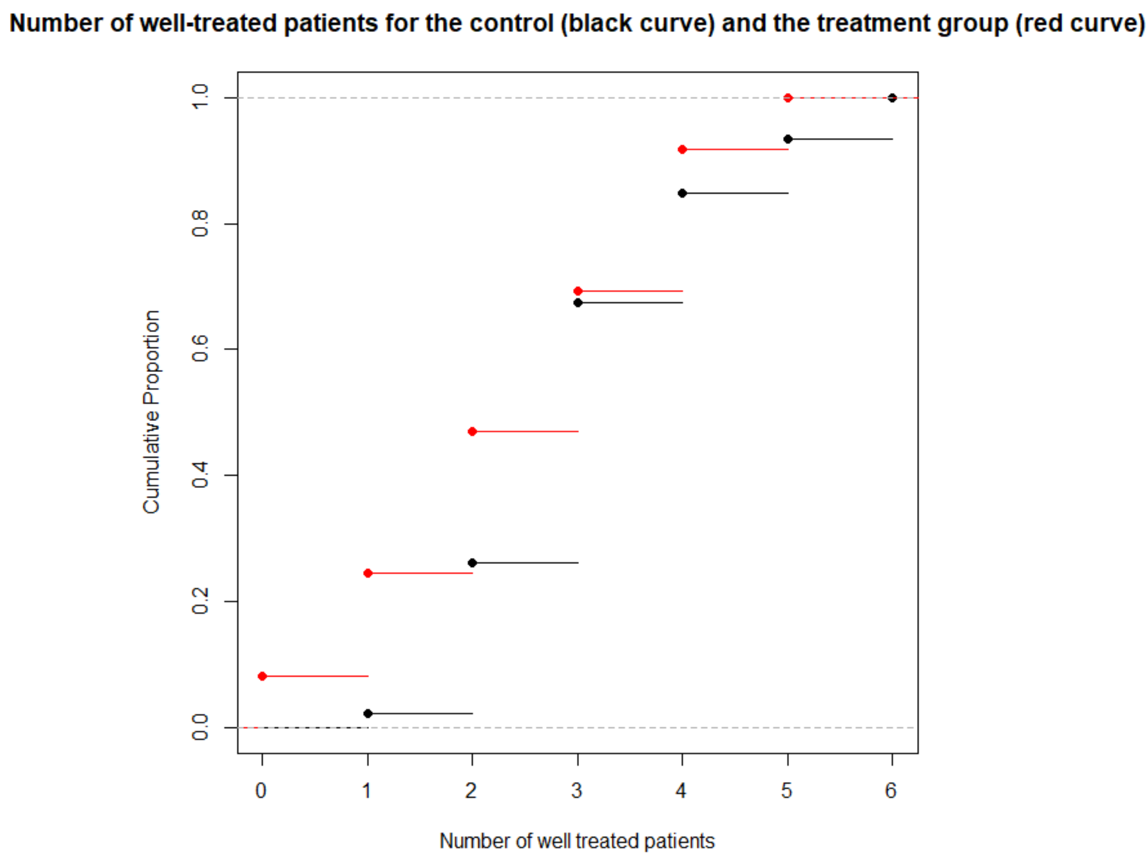


Figure 3.2 – Histograms showing the number of well treated patients for control and treatment group.

treated patients, with the effect mainly driven by CAP. We estimate as previously three models with controls and we summarize our results in Table 3.10.

The Table shows that the treated group does less quality than the control group (column 1, Table 3.10). In other words, on average, physicians who have CAP or FFS first treat 0.70 patient more than those who were exposed to the P4P. Exposure to P4P incentives therefore seems to decrease the quantity of well treated patients. When we disaggregate this effect by payment system, it appears that the overall effect is driven only by the CAP system as reported also in descriptive statistics. There is a negative difference in FFS, but this is not statistically significant

In CAP, control physicians treat 1.14 patient more than those who were exposed to P4P incentives. This result complements our analysis on physician number of services. Control CAP doing more quality than treated CAP physicians is an interesting result and implies that the reduction in overall quantity of care observed above has been detrimental to quality.

Control CAP physicians treat more patients than treated CAP likely because they are intrinsically motivated for quality. Control CAP Physicians correct more words for each patient, even though they are incentivized to provide only a minimal number of

Table 3.10 – Physician number of well treated patients for the control and treatment groups in FFS and CAP

	Number of well treated patients		
	OLS		
	Pooled data(1)	CAP only (2)	FFS only (3)
Treated group P4P first (dummy)	-0.698^{**}		
	(0.295)		
Treated group CAP second (dummy)		-1.135^{***}	
		(0.408)	
Treated group FFS second (dummy)			-0.084
			(0.508)
Age	-0.038	-0.043	0.057
	(0.087)	(0.147)	(0.163)
Sex	-0.151	-0.317	-0.082
	(0.277)	(0.416)	(0.414)
Constant	4.192^{**}	4.585	1.837
	(1.996)	(3.453)	(3.588)
Observations	95	46	49
R^2	0.067	0.178	0.009
Adjusted R^2	0.037	0.119	-0.057
Residual Std. Error	1.325 (df = 91)	1.342 (df = 42)	1.319 (df = 45)
F Statistic	2.190^{*} (df = 3; 91)	3.025^{**} (df = 3; 42)	0.141 (df = 3; 45)
Notes:	***Significant at the 1 percent level.		
	**Significant at the 5 percent level.		
	*Significant at the 10 percent level.		

services. This result is therefore showing an erosion of intrinsic motives when treated by P4P incentives. Treated CAP physicians are likely still affected by the performance criteria that was set in the P4P system, even after its withdrawal. Physicians are therefore less inclined to make "benevolent" actions after an exposure to performance criteria. This is detrimental to quality of care, measured here by the number of patients well treated.

We can go a little bit further in understanding how P4P exposure affect intrinsic motivation. In fact, intrinsic motivation is a broader term that covers elements such as altruism, professionalism, reputation concern etc. In CAP, we are able to identify physician altruism. In the CAP system in fact, the physician is economically incentivized to provide the minimum number of corrections. He can earn the capitated amount if he corrects one highlighted word in the text, regardless of whether it is correct. There is therefore no economic incentive to correct more than one word. Even though correcting one word is not always enough to generate full patient benefit, exerting extra-effort by correcting more than one word is however a possible demonstration of physician altruism. It is even more so, if there is persistence of this behavior on more than one text. The implicit assumption here is that effort is costly for the physician, which we can support by the fact that the physician was free to allocate his 5-minute time frame as wished, including enjoying leisure.

We measure altruistic intentions by looking at the number of texts for which the physician has done more than one action. We make a count of this variable and we use P4P exposure to predict the number of texts on which the physician is likely to be altruistic. We use a linear model to quantify this impact of exposure to P4P incentives on the demonstration of altruism. Our results are reported in Table 3.11

Table 3.11 – Physician altruism before and after the P4P treatment

	Number of texts on which more than one action is taken
	OLS
Treated group having P4P first	-0.983^{*} (0.518)
Age	-0.076 (0.187)
Sex	0.241 (0.528)
Constant	4.817 (4.382)
Observations	46
R^2	0.081
Adjusted R^2	0.015
Residual Std. Error	1.703 (df = 42)
F Statistic	1.227 (df = 3; 42)
Notes:	***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

It appears that, the random order in which physicians are exposed to a P4P incentive affects their altruistic intentions. Physicians in the control group (not affected by P4P), seem to be showing a likely altruistic behavior on one text more than those in the treatment group. Complementing what we already found above, P4P seems to be crowding-out intrinsic motivation for quality, especially altruism.

Previous studies have already reported the "hidden costs of incentives", using some elements of social psychology. Deci (1971), experimentally studying the effect of external rewards on intrinsic motivation report that intrinsic motivation tend to decrease when financial reward is used as an incentive. More recent work (Kreps, 1997; Benabou and Tirole, 2006; Siciliani, 2009) also demonstrate the fact that extrinsic motivation can crowd out intrinsic motivation. The general idea of this literature is the fact that introducing financial rewards pushes people to make a "cognitive reevaluation" of the activity, from one which is intrinsically motivated to one which is motivated by an anticipation of financial reward (Deci, 1971). In experimental health economics, Brosig-Koch et al., (2013) and Brosig-Koch et al., (2016b) use controlled laboratory and field experiments to study the effect of introducing pay for performance. Features of their design (baseline payment system being either FFS or CAP and then performance-based bonus on top of baseline payment system) compare to ours. One of their finding on the existence of some intrinsic motivation' crowd-out corroborates what we find in this work. Overall, explanation for motivation crowd-out is generally the fact that people perceive those "additional" incentives as "controlling". In the medical context, physicians might see P4P incentives as an attempt to control their "natural" inclination to provide "care".

3.4 Concluding remarks

In this paper, we have studied how exposure to P4P incentives affects physician quantity and quality of care. Using an experimental game in which prospective physicians are randomly exposed to P4P incentives, we have studied how this affects the number of services that they offer, as well as the quality of their practice. Physicians exposed to P4P incentives experience a specific payment system where only their actions benefiting the patient are rewarded. P4P is therefore more explicitly rewarding the physician to focus on the patient needs than FFS and CAP do. A strength of this analysis is the fact that, through the design, we can identify actions that are clearly not benefiting the patient, that is, proofreading interventions that are done in non-underlined areas. While in P4P the physician is incentivized not to focus on these actions, in FFS and CAP, there are no economic incentives preventing the physician from taking these kinds of actions. Our results are twofold:

- First, on the quantity of services, we find that, P4P tends to decrease the overall quantity of services. This decrease in quantity is observed both in FFS and CAP. We find that, exposure to P4P incentives mainly decreases the quantity of services that do not benefit patients. In this sense, it increases the focus on physicians in both CAP and FFS, with the increase in focus being higher for FFS physicians.

- Second, on the quality generated by physicians, we find that P4P exposure decreases quality for treated CAP physicians. Considering that CAP physicians are incentivized to provide the minimum number of corrections, we interpret this decrease as evidence of a crowding-out intrinsic motivation for quality. Examining which kind of intrinsic motivations is crowded out, we find P4P exposure decreases the likelihood of willingness to behave altruistically towards the patient.

Our results corroborate earlier work finding a modest effect (if any) of P4P reforms (Herbst et al., 2018; Michel-Lepage and Ventelou, 2016; Saint-Lary and Sicsic, 2015; Sicsic et al., 2012). After P4P, physicians focus more on what is relevant for the patient, but they are also less intrinsically motivated. Our design does not however allow us to study which of the two effects dominate the other. We leave this for future research. Policy makers should be aware of this tension between focus on patients' needs and intrinsic motivation crowding out when designing and promoting P4P approaches.

References

Allaby, M.A.K., 2003. Doctors for the Poor in Urban Nepal. Trop. Doct. 33, 83–85. https://doi.org/10.1177/004947550303300209

Bardey, D., Kembou Nzale, S., Ventelou, B., 2018. Physicians' incentives to adopt personalized medicine: experimental evidence. Doc. CEDE.

Bejarano, H., Green, E.P., Rassenti, S., 2017. Payment scheme self-selection in the credence goods market: An experimental study. J. Econ. Behav. Organ. 142, 396–403.

Benabou, R., Tirole, J., 2006. Incentives and prosocial behavior. Am. Econ. Rev. 96, 1652–1678.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies, N., Wiesen, D., 2013. How Effective are Pay-for-Performance Incentives for Physicians?-A Laboratory Experiment.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., Wiesen, D., 2017. The effects of introducing mixed payment systems for physicians: experimental evidence. Health Econ. 26, 243–262.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., Wiesen, D., 2016a. Physician performance pay: Evidence from a laboratory experiment. Ruhr Economic Papers.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., Wiesen, D., 2016b. Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision. J. Econ. Behav. Organ. 131, 17–23.

Chalkley, M., Malcomson, J.M., 1998. Contracting for health services when patient demand does not reflect quality. J. Health Econ. 17, 1–19.

Clemens, J., Gottlieb, J.D., 2014. Do physicians' financial incentives affect medical treatment and patient health? Am. Econ. Rev. 104, 1320–49.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. Journal of personality and Social Psychology, 18(1), 105.

Eijkenaar, F., Emmert, M., Scheppach, M., Schöffski, O., 2013. Effects of pay for performance in health care: a systematic review of systematic reviews. Health Policy 110, 115–130.

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

Emmert, M., Eijkenaar, F., Kemter, H., Esslinger, A.S., Schöffski, O., 2012. Economic evaluation of pay-for-performance in health care: a systematic review. Eur. J. Health Econ. 13, 755–767.

Galizzi, M.M., Tammi, T., Godager, G., Linnosmaa, I., Wiesen, D., 2015. Provider altruism in health economics. Työpap. 2015004.

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

Godager, G., Wiesen, D., 2013. Profit or patients' health benefit? Exploring the heterogeneity in physician altruism. J. Health Econ. 32, 1105–1116.

Gravelle, H., Sutton, M., Ma, A., 2010. Doctor behaviour under a pay for performance contract: treating, cheating and case finding? Econ. J. 120, F129–F156.

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

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.

Herbst, T., Foerster, J., Emmert, M., 2018. The impact of pay-for-performance on the quality of care in ophthalmology: Empirical evidence from Germany. Health Policy 122, 667–673.

Jack, W., 2005. Purchasing health care services from providers with unknown altruism. J. Health Econ. 24, 73–93.

Kesternich, I., Schumacher, H., Winter, J., 2015. Professional norms and physician behavior: homo oeconomicus or homo hippocraticus? J. Public Econ. 131, 1–11.

Kolstad, J.R., 2011. How to make rural jobs more attractive to health workers. Findings from a discrete choice experiment in Tanzania. Health Econ. 20, 196–211. https://doi.org/10.1002/hec.1581

Kreps, D.M., 1997. Intrinsic motivation and extrinsic incentives. Am. Econ. Rev. 87, 359–364.

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

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

Maynard, A., 2012. The powers and pitfalls of payment for performance. Health Econ. 21, 3–12.

Michel-Lepage, A., Ventelou, B., 2016. The true impact of the French pay-for-performance program on physicians' benzodiazepines prescription behavior. Eur. J. Health Econ. 17, 723–732. https://doi.org/10.1007/s10198-015-0717-6

Mullen, K.J., Frank, R.G., Rosenthal, M.B., 2010. Can you get what you pay for? Pay-for-performance and the quality of healthcare providers. Rand J. Econ. 41, 64–91.

Neal, D., 2011. Chapter 6 - The Design of Performance Pay in Education, in: Hanushek, E.A., Machin, S., Woessmann, L. (Eds.), Handbook of the Economics of Education, Handbook of The Economics of Education. Elsevier, pp. 495–550. https://doi.org/10.1016/B978-0-444-53444-6.00006-7

Pawlikowski, J., Sak, J.J., Marczewski, K., 2012. Physicians' religiosity and attitudes towards patients. Ann. Agric. Environ. Med. 19, 503–507.

Rosenthal, M.B., Frank, R.G., 2006. What is the empirical basis for paying for quality in health care? Med. Care Res. Rev. 63, 135–157.

Saint-Lary, O., Sicsic, J., 2015. Impact of a pay for performance programme on French GPs' consultation length. Health Policy 119, 417–426.

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

Scott, A., Sivey, P.M., 2017. Motivation and Competition in Health Care. SSRN Electron. J. https://doi.org/10.2139/ssrn.2905491

Serra, D., Serneels, P., Barr, A., 2011. Intrinsic motivations and the non-profit health sector: Evidence from Ethiopia. Personal. Individ. Differ. 51, 309–314.

Siciliani, L., 2009. Paying for performance and motivation crowding out. Econ. Lett. 103, 68–71.

Sicsic, J., Le Vaillant, M., Franc, C., 2012. Intrinsic and extrinsic motivations in primary care: An explanatory study among French general practitioners. Health Policy 108, 140–148.

Sliwka, D., 2007. Trust as a signal of a social norm and the hidden costs of incentive schemes. Am. Econ. Rev. 97, 999–1012.

Smith, R., Lagarde, M., Blaauw, D., Goodman, C., English, M., Mullei, K., Pagaiya, N., Tangcharoensathien, V., Erasmus, E., Hanson, K., 2012. Appealing to altruism: an alternative strategy to address the health workforce crisis in developing countries? J. Public Health 35, 164–170.

General Conclusion

This thesis uses both theoretical and experimental tools to study healthcare providers' payment systems. Our overall contribution can be summed up in the following points.

Joint production of health or education services in a context with limited liability and adverse selection on altruism: There is an abundant literature in contract theory which studies optimal contract with moral hazard on altruism. The study of heterogeneity in altruism is however new and relatively few theoretical papers have accounted for that. Theoretical settings that allow for joint production and productivity enhancing investments are also relatively scarce. Thanks to recent investigations in health economics, it is documented that physicians vary according to their level of ethics (professionalism, altruism, concern for reputation, etc.). The possibilities offered by personalized medicine also reinforce the necessity to account for joint-production frameworks. The normative analysis proposed in this dissertation, combines meaningful assumptions on joint production, adverse selection on altruism and limited liability. Our use of limited liability rather than "classic" participation constraints also provides interesting insights. The limited liability constraint imposes that work must be paid, regardless of whether the healthcare provider is altruistic. This differs from the "classic" participation constraint used so far in the literature and which implicitly implies that the agent's satisfaction is possible only in a principal-agent relationship. Under the conditions of our setting, we show that the optimal separating contract entails a higher transfer to the altruistic agent and a higher principal's input for the selfish agent. These results contrast with existing ones and the main driver of this difference is the use of a limited liability in our case. Our assumption on the levels of altruism is interesting and provides a benchmark for future work that can compare our results with what we get in a framework where altruism is defined to be below that of the perfect agent, for both types.

Physicians' incentives to use personalized medicine techniques: Personalized medicine seems promising as it allows to make a profiling of patients according to their best predicted response. The literature documents reasons for why personalized medicine can only be poorly adopted. One of the main given reason is the fact that current payment systems for physicians might not provide enough incentives to adopt personalized medicine techniques. It is documented that physicians paid under a fee-for-service plan even tend to under-use available and free personalized medicine technologies. We use an experimental framework to create the trade-off that the physician faces when confronted to the adoption of personalized medicine. We compare a free and a costly provision of personalized medicine technologies and we study the role of physician payment systems in their decisions to adopt these tools. Earlier work has demonstrated the feasibility of modeling physicians' trade-offs by the use of experimental tools. Our experimental setting has provided a meaningful context to the study of physician incentives when confronted to personalized medicine. We find that, overall, physicians would tend to better use personalized medicine technologies when it comes at a cost. When personalized medicine access is given for free, pay-for-performance, capitation and fee-for-service systems, tend to provide very similar incentives on a set of outcomes. Pay-for-performance however is associated with a higher likelihood of acquiring at a cost personalized medicine tools. Two areas of improvement can be identified for this work: (i) first, we can think of a different timing in which the physician makes her decision to access at a cost personalized medicine technologies. Our current timing considers that the physician first sees the benefits of using personalized medicine, and later decides whether she wants costly access. A different timing where the physician takes her decision without knowing the benefits of the technology could be thought of. In this last case, we hypothesize that elements such as physicians' attitudes toward risks would be potential predictors of physicians' decisions to adopt costly personalized medicine, the latter decision being perceived as any kind of risky investment. (ii) A second area of improvement is the way in which we implement the pay-for-performance condition in the lab. In our experiment, the pay-for-performance is used for all physicians in combination with either fee-for-service or capitation, because of our within-subject design. The physician's overall gain is therefore the sum of his payments in each of the experimental conditions. This procedure has the advantage that the physician does not incur all the financial burden that is embedded in a pay-for-performance system. Future work could however consider payment systems that are more unified, having in one experimental condition, a combination of both pay-for-performance and other classical payment systems (salary, capitation, fee-for-service). Our results on the incentive role of payment systems are robust to these limitations. Policy makers can use our work to design more P4P initiatives in order to enhance costly adoption and potential benefits of personalized medicine technologies.

Effect of previous exposure to pay-for-performance initiatives: Pay-for-performance seems to be of high interest for many policy-makers around the world. There is an abundant literature in both developing and developed countries on effects of pay-for-performance initiatives. This literature overall documents a modest (if any) effect of pay-for-performance initiatives and also highlights the fact that there is a persistence of the effect of pay-for-performance initiatives on physicians behaviors. Building on a framework where access to pay-for-performance initiatives is randomly decided to be either anterior or posterior to fee-for-service or capitation, we report on the effect of pay-for-performance on physicians' quantity and quality of care. Our main result suggests that, exposure to pay-for-performance incentives decreases the quantity and the quality of care provided by physicians. The decrease in quantity is overall not always bad. We have both a decrease in the quantity of useless actions and useful actions. By leading to a decrease in the quantity of useless actions, pay-for-performance exposure is therefore increasing the focus of physicians' actions. The decrease in quality is likely an indication that pay-for-performance exposure destroys physicians' intrinsic motivation for quality. Our experimental setting has the advantage that we can distinguish useless and useful actions to better study the effect of pay-for-performance exposure. Our result that pay-for-performance increases physician focus but destroys their intrinsic motivation provides a promising avenue for future research. Policy makers should be aware of the existence of this trade-off when designing and promoting pay-for-performance initiatives.

Overall, our PhD work has studied three main ingredients that in our view are among the most important ones to improve health systems performance: (i) payment systems for healthcare professionals; (ii) effort enhancing investments like personalized medicine access; (iii) role of healthcare professionals' ethics and overall code of conduct.

We have a couple of research prospects for the future. Apart from expanding our current research agenda on health workers incentives, we are also likely to collaborate on a research project that will study experimentally the impact of performance-based financing for health facilities in Ivory Coast. In general, our research agenda will expand to questions related to education and early childhood development in developing countries.

Abstract

We study incentive properties of healthcare providers' contracts in different contexts and using a range of methods. Stability and sustainability of health systems are highly dependent upon the contribution of healthcare providers. In the regulator-healthcare provider relationship, such elements as (i) healthcare providers norms and ethics (their altruism, concern for reputation, professionalism, etc.); (ii) payment systems in place; and (iii) infrastructures and technologies available and their effective use, are critical factors for a successful relationship. In this thesis, we study healthcare providers' payment systems accounting for the 2 other elements above-mentioned. In Chapter 1, we propose a "regulator-agent" model with adverse selection on altruism, limited liability and a possible effort enhancing input provided by the regulator. In such a context (characterizing the healthcare sector with a free access to personalized medicine, for example), we show that the optimal contracts entails higher salaries for the altruistic agents and higher effort-enhancing technology for selfish agents. In Chapter 2, we propose an experiment in which healthcare providers can access for free or at a given cost, personalized medicine techniques. For different payment systems, we assess the likelihood of making the decision to access personalized medicine when it is paid. We also focus on how healthcare providers use these technologies, depending on whether their access is free or costly. We find that healthcare providers tend to make better use of personalized medicine techniques when they acquire it at a cost. In Chapter 3, we study incentive properties of performance pay systems. Using the same experiment as in Chapter 2, we find that performance-based systems increase healthcare providers' attention on what is relevant to the patient. It however destroys their intrinsic motivation. This PhD work increases our understanding of optimal contracts' properties in contexts where the regulator can provide effort-enhancing technologies to healthcare professionals.

Keywords: Regulation, Healthcare providers, Altruism, Pay-for-performance, Personalized medicine, Laboratory experiment

JEL classification: I18, J33, L24, C9

Résumé

Nous étudions les propriétés incitatives des contrats des médecins dans différents contextes et en mobilisant différentes méthodes. L'efficience et la soutenabilité des systèmes de santé dépendent fortement de la contribution des médecins. Dans la relation régulateur-médecins, certains facteurs importants sont essentiels pour maximiser les avantages de la relation. Des éléments tels que (i) les normes et l'éthique des médecins (leur altruisme, le souci de préserver leur réputation, le professionnalisme, etc.), (ii) les systèmes de paiement en place et (iii) les infrastructures et technologies disponibles et leur utilisation effective par les acteurs du système, sont des facteurs déterminants pour le succès de cette relation. Notre thèse étudie les systèmes de paiement des médecins en tenant compte des deux autres facteurs sus-cités. Au chapitre 1, nous proposons un cadre conceptuel global à l'aide d'un modèle "principal-agent" dans lequel il existe une sélection adverse sur l'altruisme des agents, une responsabilité limitée et une technologie fournie par le régulateur et pouvant améliorer la qualité de l'effort des médecins. Dans un tel contexte (caractérisant le secteur de la santé avec un accès à la médecine personnalisée par exemple), nous montrons que les contrats optimaux impliquent des salaires plus élevés pour les agents altruistes et une technologie d'amélioration de l'effort de meilleure qualité pour les agents égoïstes. Au chapitre 2, nous proposons une expérience dans laquelle les médecins peuvent accéder gratuitement ou à un coût donné à des techniques de médecine personnalisée. Nous évaluons pour différents systèmes de paiement, la probabilité que les médecins prennent la décision d'un accès payant à la médecine personnalisée, et nous nous concentrons également sur la manière dont ils utilisent ces technologies, selon que leur accès est gratuit ou coûteux. Nous trouvons que les médecins ont tendance à mieux utiliser les techniques de médecine personnalisée lorsqu'ils l'ont acquis à un coût. Au chapitre 3, nous étudions les propriétés incitatives des systèmes de rémunération à la performance. En utilisant la même expérience du chapitre 2, nous trouvons que les systèmes de paiement à la performance renforcent l'attention des médecins sur ce qui est pertinent pour le patient, mais sont associés à une érosion de leur motivation intrinsèque. De manière générale, nos travaux de doctorat permettent de mieux comprendre les propriétés des contrats optimaux dans des contextes où il y a la possibilité de fournir des techniques d'amélioration de l'effort aux médecins.

Mots-Clés: Régulation, Médecins, Altruisme, Paiement à la performance, Médecine personnalisée, Expérience de laboratoire

Classification JEL: I18, J33, L24, C9