December 11, 2019

Balance Billing as a Signaling Device

Damien Besancenot[*], Karine Lamiraud[†], Radu Vranceanu[‡]

Abstract

In France and some other countries, patients can chose between consulting a physician working in the regulated sector where, in general, fees are covered in full by the health insurance, or consulting a physician in the unregulated sector under a balance billing scheme. In the latter fees might not be fully covered by the health insurance, and patients pay out-of-pocket some positive amount. We analyze the signaling properties of this mechanism. If patients differ in their marginal cost of adhering to the treatment, and if this characteristic is private information to them, a mixed regulated and unregulated system including a balance billing scheme can help separating patient types. In this equilibrium, those patients with a high propensity to adhere to the treatment will receive a better attention from their physician, and vice-versa. We analyze the equilibria of the game and comment on their welfare properties.

Keywords: Balance billing, Treatment adherence, Signaling equilibrium, Physicians' beliefs, Reciprocation.

JEL Classification: I11, D82.

[*]University of Paris Descartes and LIRAES, 45 rue des Saints Pères, 75270 Paris. E-mail: damien.besancenot@parisdescartes.fr.

[†]ESSEC Business School and THEMA, 1 Av. Bernard Hirsch, 95021 Cergy, France. E-mail: lamiraud@essec.edu

[‡]ESSEC Business School and THEMA. 1 Av. Bernard Hirsch, 95021 Cergy, France. E-mail: vranceanu@essec.edu.

1 Introduction

In general, health authorities in the developed world regulate the market for outpatient care by setting a cap on fees that physicians can charge; furthermore, most of the times these regulated fees are reimbursed in full by the national health insurance scheme. However, in several countries, all or some categories of physicians are allowed to charge patients more than the regulated fee, a mechanism referred to as "balance billing", "bulk-billing" or "extra-billing". In this case, patients will have to pay out-of-pocket the extra fee or at least a part of it.

For instance, in France specialist physicians can opt between the legal framework of the "sector 1" and agree to provide medical services at a regulated fee, covered in full by a combination of public and private insurance schemes[1] , or the legal framework of "sector 2" under which they can charge an extra fee on top of the regulated fee. In 2017, 46% of the specialized physicians were registered with the sector 2 (DREES, 2018). If private insurance schemes do cover a large part of these extra expenses, some of them will be charged to the patient (see Clerc et al., 2012; Coudin et al., 2015, Dormont and Peron, 2016). The report of the DREES (2018) indicates that the uncovered fees totaled 2bn euros in 2017. Based on survey data from 2012, Dormont and Peron (2016) estimate the sector 2 out-of-pocket payment at 439 euros per patient*year. Belgium is running a similar system (Lecluyse et al., 2009).

In the US, before 1984 a significant number of physicians applied balance billing to beneficiaries of national health insurance for the 65+ year old (US Medicare). Between 1984 and 1990, a wave of regulations were adopted to prompt physicians to gradually abandon the balance billing system. If in 1984, balance billing amounted to 27% of the total out-of-pocket payments charged to Medicare beneficiaries, additional charges are now limited to 10% of the fee set by the federal health insurance (Kifmann and Scheuer, 2011). Interestingly, McKnight (2007) studied this transition and reported no variation in the quality of the provided service.[2]

General criticism against balance billing in these countries emphasizes that the system comes with high and unsustainable charges for patients with little value added for them, as today there

[1] Except for a one euro co-payment, introduced in 2005.

[2] See also Epp et al. (2000) for the presentation of balance billing in Canada.

is no evidence that physicians who charge supplementary fees provide better quality care than those who do not (Calconen and Van den Ven, 2019). High balance billing fees also contribute to mounting health expenses, and put additional stress on private insurance companies that cover these expenses. The list of benefits seems to be much shorter, pointing out in general the provision of higher income to physicians without an additional burden on the public spending.

This paper suggests that a well-designed balance billing system might provide a useful signaling mechanism for patients to sort according to their willingness to adhere to a treatment. It therefore provides a complementary explanation to the persistence of balance-billing in many countries. We analyze the functioning of a dual payment system for primary health care, including a regulated sector where cares are free of any supplement, and an unregulated sector which physicians are entitled to balance bill patients. The allocation of physicians to the sectors is exogenously given.[3] By paying a supplementary fee, depending on the equilibrium, patients might signal their willingness to adhere to the treatment. Our equilibrium analysis goes beyond the classical signaling model in which players can choose one action to signal their type (Spence, 1973; 2002), by including as an additional decision layer the optimal choice of effort by the patient. Thus, the equilibrium strategy of a player includes his billing strategy *and* the choice of effort. With two types of patients, the optimal effort of one type depends on the physicians' expected effort, which in turn depends on the equilibrium efforts of both type of patients. To our knowledge, there is no other theoretical analysis of the balance billing mechanism to show under which conditions this system can help sorting patient types, and prompt them to adjust their effort to adhere to a treatment.

The model is inspired by the French organization of the outcare patient market. We assume that physicians, who deliver the same service, can work either in the "sector 1" where fees are regulated and covered by the health insurances, or in the "sector 2" where physicians can charge a supplementary fee on top of the regulated fee which is not fully covered by the insurances (Coudin et al., 2015; Calconen and Van den Ven, 2019). We analyze in this context the game between a patient and his physician who interact in the joint production of the health care service.

[3] We do not analyze here physicians' long term decision on whether to work in sector 1 or in sector 2. See Besancenot and Vranceanu (2017) for an equilibrium search model where physicians optimally choose the sector of activity.

An impressive body of literature in social science and medicine was dedicated to the analysis of the patient-physician interaction, and how this interaction is impacting the actual and perceived quality of the treatment (inter alia: Parson, 1951; Buller and Stone, 1992; Charles et al., 1997; Heritage and Maynard, 2006).

One essential factor contributing to the success of a treatment is the patient's adherence to that treatment, i.e., the extent to which his behavior in terms of taking medication, following diets and changing lifestyle coincides with the received medical advice (inter alia, Davis, 1968; Haynes, 1979; Vermeire et al., 2001; DiMateo et al., 2002; DiMateo, 2004; Simpson et al., 2006; Iuga and McGuire, 2014). As a sheer example of how serious this issue can be, survey data reveal that seventy-five percent of Americans have trouble taking their medicine as directed; nonadherence to treatments accounted for an estimated 125,000 deaths annually and at least 10 percent of hospitalizations in the United States in 2012 (Benjamin, 2012). Following this literature, we will assume that patients differ in their propensity to adhere to the treatment (Giuffrida and Gravelle, 1998; Lamiraud and Geoffard, 2007), and this characteristic is private information to them. To keep the analysis simple, we assume that there are only two types of patients: those with a low and those with a high marginal cost of adhering to the treatment.

If the patient effort is important, so is the physician's investment in the relationship with the patient.[4] In our model, we follow Balsa and McGuire (2003) and Fichera et al. (2018) to assume that the health production function positively depends on both the effort provided by the patient to adhere to the treatment, and the effort provided by the physician in terms of attention, interest and time dedicated to his patient. Using observations from a large sample of interactions between English doctors and patients with cardiovascular diseases in 2004-2006, Fichera et al. (2018) verified that the two efforts complement each other.

It is commonly accepted that, beyond other more materialistic concerns, physicians care about the health of the patient (Arrow, 1963; Ellis and McGuire, 1986; Balsa and McGuire, 2003).[5]

[4] Many studies in the psychology of health care emphasize the role of physician's empathy toward the patient (e.g., Derksen et al., 2013; Kelm et al., 2014).

[5] Recent experimental studies corroborate the assumption of the altruisitic behavior of physicians (e.g., Godager and Wiesen, 2013; Kesternich et al. 2016).

Using a simple utility maximization framework that takes into account this dual goal and the cost of effort, we show that the physician's optimal behavior consists into aligning his/her own effort to the effort of the patient as he/she perceives it. Because the effort of the patient is unobservable, the physician will use the information provided by the payment strategy to update his/her beliefs about the type and effort of the patient.

The paper determines the equilibria of this game. Equilibrium is defined as a situation in which all patients chose their best billing and effort strategies, given the physicians' beliefs, and physicians' beliefs are correct given the best billing strategy and effort choice by the patients. As a most interesting case, the game presents a separating equilibrium in which only patients with a high propensity to adhere to the treatment will use the balance billing system, and patients with a low propensity to adhere to the treatment will use the regulated system (no charge). In this equilibrium, the "compliant" patient provides the high effort level, and the "reluctant" type provides the low effort level. Furthermore, physicians perfectly identify the two types and adapt their own effort to the effort of the patients. The welfare of the compliant patients is the highest. The separating equilibrium is not the only equilibrium of this game. If the balance billing fee is too high, the model presents a pooling equilibrium in which nobody pays it, and physicians cannot infer the type of patients from their billing strategy. The welfare of the "compliant" patients is lower in this equilibrium compared to the separating equilibrium.

The intuition underlying our main result is also present in Balsa and McGuire (2003) who explained how stereotypic beliefs can be supported as an equilibrium with health-care discrimination: because physicians believe that black patients are less compliant than the white patients, for some parameters of the problem it might be optimal for the white (black) patients to comply (and respectively not to comply). Our analysis is enriching this perspective, as it involves two decisions (whether to pay for the signal or not, and the effort level of each type). In Balsa and McGuire (2003), the type is directly observable (color of the skin); in our analysis the type is not observable; patients can signal themselves or not, then will adapt the effort accordingly. As a consequence, the belief formation mechanism is also more sophisticated, in particular for the mixed strategy equilibria.

Related to our work, a substantial number of papers in the industrial organization of the health care sector analyze balance billing and its welfare properties through the prism of price discrimination. Early models represented the physician as a monopolist providing a homogenous quality service who can price discriminate between patients with a different willingness to pay (Mitchell and Cromwell, 1982; Zuckerman and Holahan, 1991; Savage and Jones, 2004). In these models, balance billing can only increase the rents of physicians at the expense of patients. Feldman and Sloan (1988) argue that a monopolist physician subject to balance billing constraints would alter the quality of the service.

Glazer and McGuire (1993) analyzed monopolistic competition between two physicians who can engage in price and quality differentiation. They reveal the existence of a positive fee that maximizes social welfare, and show that restrictions on balance billing as applied in the US in the late 1980s would reduce quality of care for all patients, regardless of whether they pay an extra fee or not. Kifmann and Scheuer (2011) use the same model to show that a mixed system, with balance billing and fee-only patients, can increase patient welfare if the administrative costs of Medicare are sufficiently low. This relatively optimistic conclusion on the ability of balance billing to improve patients' welfare has been challenged by Jelovac (2015); she shows that if the physicians have imperfect information about patients' willingness to pay and must charge uniform fees, balance billing can increase inequalities in access to care and ultimately reduce social welfare. Gravelle et al. (2016) develop a n-player differentiation model *à la* Salop with price and quality differentiation, as applied to the Australian health care market, where balance billing is generalized.[6] They call attention on the risks related to the documented increasing market concentration that possibly leads to higher fees.[7]

Our analysis can also be related to the analysis of "dual practice" prevailing in many countries (e.g. the UK), where physicians can chose to practice either in the public sector, or in the private sector, or in both (see Eggleston and Bir, 2006; Barros and Siciliani, 2011). For instance, Kuhn

[6] In Australia patients pay a fee for each General Practitioner (GP) consultation. Physicians choose their fees freely. The national, tax financed, insurance scheme (Medicare) provides a subsidy for the cost of a consultation (the Medicare rebate). The patient pays the excess of the physician fee over the Medicare fee and these out-of-pocket co-payments by patients cannot be covered by insurance.

[7] Mu et al. (2018) bring empirical evidence showing that patients do not perceive quality as different among low price and high price medical services in Australia.

and Nuscheler (2013) study how a monopolist physician sets tariffs and quality (waiting time) when he can chose between a basic treatment at the regulated fee, and a sophisticated treatment at a higher, free price and reveal the conditions under which there is under/over provision of the health service as compared to the first-best allocation of resources.

The paper is organized as follows. The next section presents the game between patients and the physicians. Section 3 defines and analyses the equilibria of this game; policy implications follow. Section 4 is our conclusion.

2 Main assumptions

We study the strategic interaction between a patient and his physician in the joint production of the health care service. The patient can choose his effort in adhering to the treatment, and also whether to opt for the low fees in the "regulated sector" or the high fees in the "balance billing payment system".

We assume all patients have access to the regulated sector, in which the cost of the medical service is covered in full by the public and private health insurance. Patients can also choose to consult a physician working in the unregulated sector under a balance billing scheme. In the latter, physicians' fees would exceed fees in the regulated sector; let c denote the part of physician fees uncovered by the public and private insurance schemes under balance billing. This fee is assumed to be exogenously given; the implicit assumption is that the National Health Authorities can control it (at least can set a credible upper limit on it, as indicated in Calcoen and Van den Ven, 2019). Thus the payment strategy of the patient is $S \in \{c, \not{c}\}$, depending on whether he opts for the balance billing system (pays c) or the regulated system (pays nothing, $\not{c}$).

We consider an elementary health production function with two inputs. First, chances that the treatment is successful depend on a patient's own effort in adhering to the treatment. Second, the success of the treatment depends on the effort of the physician (including attention, time, interest). Denoting by H_{ij} the amount of health care delivered by the interaction between a physician j and a patient i, by e_i the patient's effort and by e_j the physician's effort, such a production function can take the standard Cobb-Douglas specification as suggested in Balsa and McGuire (2003):

$$H_{ij} = Ae_i e_j. \tag{1}$$

Since $\frac{\partial^2 H_{ij}}{\partial e_i \partial e_j} > 0$, this function features effort complementarity as revealed in the empirical analysis by Fichera et al. (2018).

Similar to Balsa and McGuire (2003), patients have the choice between two effort levels, a high effort level $e^h = 1$ and a low effort level, $e^l = 0$. Physician's effort is also defined in the interval $[0, 1]$.

Patients bear a cost of adhering to the treatment, assumed to be a linear function in the effort level; for an individual i, the adherence cost is $k_i(e_i)^2$, with $k_i > 0$. The mass of patients is normalized to one.

To keep the model tractable, we assume that the distribution of k_i is degenerated in two points of mass μ and respectively $(1-\mu)$: all type 1 patients have the same marginal cost of effort k_1 and all type 2 patients have the same marginal cost of effort k_2, with $k_2 > k_1$.[8]

The utility of the type i patient is thus:

$$U^i = Ae_i e_j - k_i(e_i)^2 - \mathbf{1}_c c, \text{ with } i \in \{1, 2\}, \tag{2}$$

expression in which the indicator variable $\mathbf{1}_c$ takes the value of 1 if the patient chooses the balance billing system and zero otherwise.

The physician j cares about the health of the patient (Ellis and McGuire, 1986; Balsa and McGuire, 2003). We assume that the utility function has the additively separable form:

$$V_j = \varphi p + (1-\varphi)Ae_i e_j - \beta(e_j)^2, \tag{3}$$

where $p > 0$ is the (constant) consultation fee (could be sector specific, but this is irrelevant in this case), e_i and e_j are the effort levels of the patient and physician respectively as defined before, $\beta(e_j)^2$ is a quadratic cost of effort for the physician and φ and $(1-\varphi)$ are the weights of the materialistic and respectively altruistic goals in the payoff of the physician.

[8] The properties of the solution would not change if we consider a non-degenerated distribution of costs instead.

However, the effort of the patient is unobservable, thus the physician uses as a guide his/her own expectations about his/her patient's effort, $E_j\left[e_i|S\right]$. The expected value conditional on the patient's observed payment strategy is:

$$EV_j[S] = \varphi p + (1-\varphi)AE_j\left[e_i|S\right]e_j - \beta\left(e_j\right)^2. \tag{4}$$

Then the first order condition for utility maximization determines the optimal effort of the physician simply as:

$$e_j = \frac{(1-\varphi)A}{2\beta}E_j\left[e_i|S\right],\ \text{with}\ \frac{(1-\varphi)A}{2\beta} \leq 1. \tag{5}$$

In the following, to avoid excessive complexity, we drop the index j from the expectations.

Physicians' beliefs are represented by the conditional probabilities Pr[type 1|S] and Pr[type 2|S] where S is the observed billing strategy $S \in \{c, \not{c}\}$.

Let e_1 denote the effort of the type 1 patient, and e_2 the effort of the type 2 patient, with $e_{1,2} \in \{0,1\}$. With these notations, the physician's expectations about the patient's effort are:

$$E\left[e|S\right] = \Pr[\text{type } 1|S]e_1(S) + \Pr[\text{type } 2|S]e_2(S). \tag{6}$$

In equilibrium, $e_1$and e_2 are the optimal effort levels of both types of patients.

Under these assumptions, the patient's utility becomes:

$$U^i = \gamma e_i E\left[e_i|S\right] - k_i(e_i)^2 - \mathbf{1}_c c,\ \text{with}\ i \in \{1,2\}\ \text{and}\ \gamma = \frac{(1-\varphi)A^2}{2\beta} \tag{7}$$

We will consider hereafter only the non-trivial case in which the marginal cost k_i is distributed above and below γ. This cost threshold determines two types: an "eager to comply" type 1 with $k_1 < \gamma$ and a "reluctant to comply" type 2 with $k_2 > \gamma$. Let μ be the frequency of type 1 patients in the total patient population; $(1-\mu)$ is the frequency of type 2 patients.

The sequence of decision is the following:

- At the outset of the game, Nature decides on patients' types.

- Patients chose their best billing strategy given their type.

- The physician observes the billing strategy, forms his beliefs about the type of patient (and his chosen effort) and finally decides about his own effort. At the same time, given the physician's beliefs, the patient chooses his optimal effort level.

The two latter decisions are taken simultaneously (Balsa and McGuire, 2003), as will be shown in the resolution steps. This analytical framework is the most meaningful for a one-shot patient-physician interaction. In the case of chronic diseases, characterized by a lasting relationship, a sequential approach including learning and reputation building, would be more appropriate (McGuire, 2001).

We can now study the equilibria of this game.

3 Equilibria of the game

An equilibrium is defined as a situation in which all patients chose the optimal billing and effort strategies given physicians' beliefs, and physicians beliefs are correct given the best billing strategy and effort choice by the patients.

This "dual strategy" problem presents a slightly higher degree of complexity compared to the traditional signalling model because, in the last stage of the game, patients and physicians decide simultaneously on the effort level and respectively the expected effort level given the billing strategy. As it will be shown in the next section, the optimal effort level of each type of patient types and the associated physician's beliefs are specific to the equilibrium. Specific to this model, the physician's expectations about the patient's effort depend on the optimal effort strategy of each type of patient, and the optimal effort strategy of each type depends on the optimal strategy of the other type by the intermediation of the physician's beliefs. As a consequence, the optimal effort of a type is equilibrium specific.

We will present in the main text the pure strategy equilibria of the game, namely a separating equilibrium, and the pooling equilibrium in which no patient opts for the balance billing system. We show in the Appendix 1 that the opposite pooling equilibrium in which all patients pay the fee $c > 0$ does not exist. Mixed strategy equilibria are analyzed in the Appendix 2 and 3.

3.1 The separating equilibrium

We analyze the separating equilibrium in which all type 1 patients (those with a low marginal cost of treatment adherence) choose the balance billing system, and all type 2 patients (with a high marginal adherence cost) choose the free-of-charge system.[9] In this case, the physician's beliefs are:

$$\begin{cases} \Pr[type\_1|c] = 1 \\ \Pr[type\_1|\not{c}] = 0 \end{cases}. \tag{8}$$

Because in this equilibrium the choice of the balance billing system signals unambiguously the type of patient, according to Eq. (6), the physician's expectations about the patient's effort contingent of the billing strategy of the latter are: $E[e|c] = e_1$ and $E[e|\not{c}] = e_2$.

Following the usual method in games with imperfect information, in a first step, we determine the *optimal efforts for each type* in this specific equilibrium. Then, we analyze the conditions necessary for this equilibrium to exist. Note that the steps used to analyze this equilibrium can be used to study any equilibrium of this game.

The utility of the type 1 patient who pays the fee c is:

$$\begin{aligned} U^1(e_1, c) &= \gamma e_1 E\left[e|c\right] - k_1 \left(e_1\right)^2 - c \\ &= \left(e_1\right)^2 \left(\gamma - k_1\right) - c \end{aligned}$$

Because $(\gamma - k_1) > 0$, the optimal effort strategy for type 1 patients is $e_1 = 1$. This strategy is preferred to the zero effort strategy regardless of the choice of effort level by type 2 patients.[10] With this optimal effort, the utility of a type 1 patient is:

$$U^1(1, c) = (\gamma - k_1) - c. \tag{9}$$

Turning to type 2 patients, we know that in this equilibrium, they do not pay c. The utility of a type 2 patient is:

$$\begin{aligned} U^2(e, \not{c}) &= \gamma e_2 E\left[e|\not{c}\right] - k_2 \left(e_2\right)^2 \\ &= \left(\gamma - k_2\right) \left(e_2\right)^2 \end{aligned}$$

We assumed that $k_2 > \gamma$: the optimal effort of the type 2 patient is $e_2 = 0$. For this optimal

[9] It can be shown that the opposite separating equilibrium in which the type 1 does not pay the fee and type 2 does, does not exist. Indeed, in this equilibrium, the type 2 has all incentives to deviate, as it will save the fee, and be considered a high effort patient.

[10] The optimal effort of the other type does not appear in the expression of the physician's expectations given that in this equilibrium the billing strategy signals the type. This is not the case in other equilibria.

effort, his utility is:

$$U^2(0,\not{c}) = 0 \tag{10}$$

Second step, we study *the existence conditions* of the separating equilibrium. If the type 1 individual deviates and chooses the regulated sector $\not{c}$, physicians will believe that he is of the type 2, and, accordingly, that his optimal effort is 0. Formally, $E\left[e|\not{c}\right] = 0$. The utility of the type 1 patient who deviates from his equilibrium strategy is:

$$U^1(e_1,\not{c}) = e_1 E\left[e|\not{c}\right] - k_1\left(e_1\right)^2 = -k_1\left(e_1\right)^2.$$

Obviously, his optimal effort is $e_1 = 0$ and the utility of the type 1 who deviates is $U^1(0,\not{c}) = 0$.

Thus the type 1 patient has no incentive to deviate from the balance billing strategy if:

$$U^1(1,c) = (\gamma - k_1) - c > 0 = U^1(0,\not{c}). \tag{11}$$

This leads to the existence condition:

$$c < c_1 = (\gamma - k_1). \tag{12}$$

If the type 2 patient decides to deviate and pays c (which is the optimal strategy of the type 1), physicians will believe that he is of the type 1 and makes the high effort: $E\left[e|c\right] = 1$. Patient utility would be:

$$\begin{aligned} U^2(e_2,c) &= \gamma e_2 E\left[e|c\right] - k_2\left(e_2\right)^2 - c \\ &= e_2(\gamma - k_2 e_2) - c \end{aligned} \tag{13}$$

He has the choice between making an effort $e_2 = 0$ or $e_2 = 1$. Because $(\gamma - k_2) < 0, U^2(0,c) > U^2(1,c)$. His best choice would be $e_2 = 0$, for an utility $U^2(0,c)$. Yet $U^2(0,c) < U^2(0,\not{c}) = 0$. He has no incentive to deviate.

Thus the only condition required to guarantee the existence of this equilibrium is 12.

$$c < c_1 = (\gamma - k_1). \tag{14}$$

Summary. In this equilibrium, patients with low compliance costs will choose the balance billing system, and pay c. This allows physicians to identify their type, and provide the highest effort in the relationship with these patients. They benefit of the highest production of health care services. Patients with high compliance costs will choose the low effort level (normalized to zero). For these patients it does not worth paying the fee c, and they will attract a low physician attention, in line with their own effort.

For sure, from the perspective of the type 1 patients, a positive but small fee c is to be preferred to a larger fee, as they would obtain the desired separation effect at the lowest cost for them. Physicians would prefer a higher fee, which is maximizing their payoff.

3.2 Pooling equilibrium 1: nobody pays c.

We study the equilibrium in which no patient pays the extra fee c. In this equilibrium patients do no resort to balance billing to signal their type, physicians consider that the likelihood that one patient is of a given type is equal to the frequency of that type in the population of patients. However, should one patient deviate and decide to pay the extra fee c, in line with the insight from the separating equilibrium (in which the type 1 patient pays the fee, and type 2 don't), we assume that physicians will consider that he is of the type 1 (i.e., the patient with the high propensity to adhere to the treatment).[11] Therefore physicians' beliefs are:

$$\begin{cases} \Pr[type\_1|\not{c}] = \mu \\ \Pr[type\_1|c] = 1 \end{cases} . \tag{15}$$

Following the same resolution steps as before, we first determine the equilibrium optimal effort of each type. By contrast with the previous case, because now physician's expectations (Eq. 6) include the optimal effort of both types, this optimal effort of one type might depend on the effort of the other type.

We study first the effort strategy of type 2 patients, taking as given the effort strategy of type 1.

(a) Let us first assume that type 1 patients make the high effort, $e_1 = 1$. Then utility of type

[11] In the opposite case, if physicians assume that a patient who decides to pay the fee is of the noncompliant type, the game presents another pooling equilibrium in which nobody makes the high effort, and nobody pays the fee.

2 patients is:

$$U^2(e_2, \not q | e_1 \;=\; 1) = \gamma e_2 E\left[e | \not q\right] - k_2\left(e_2\right)^2$$
$$= \gamma e_2\left[\mu + (1-\mu)e_2\right] - k_2\left(e_2\right)^2 \tag{16}$$
$$= e_2\left[\gamma\mu + \gamma(1-\mu-k_2)e_2\right] \tag{17}$$

The utility of the type 2 agent contingent on his effort is:

$$U^2(e_2, \not q) = \begin{cases} 0 & \text{if } e_2 = 0 \\ (\gamma - k_2) & \text{if } e_2 = 1 \end{cases}. \tag{18}$$

Because $k_2 > \gamma$, the optimal effort is $e_2 = 0$, leading to optimal utility:

$$U^2(e_2, \not q) = 0. \tag{19}$$

(b) Let us now assume that type 1 patients make the optimal effort $e_1 = 0$. The utility of type 2 patients becomes:

$$U^2(e_2, \not q | e_1 \;=\; 0) = \gamma e_2 E\left[e | \not q\right] - k_2\left(e_2\right)^2$$
$$= \gamma e_2\left[(1-\mu)e_2\right] - k_2\left(e_2\right)^2. \tag{20}$$

Depending on whether he/she implements the high/low effort, the utility is:

$$U^2(e_2, \not q) = \begin{cases} 0 & \text{if } e_2 = 0 \\ \gamma(1-\mu) - k_2 & \text{if } e_2 = 1 \end{cases}. \tag{21}$$

Because $k_2 > \gamma$, the optimal effort is $e_2 = 0$, leading to the largest utility:

$$U^2(e_2, \not q) = 0. \tag{22}$$

From (a) and (b), we infer that $e_2 = 0$ is the best effort strategy of type 2 patient in the $S = \not q$ pooling equilibrium regardless of the effort level of the type 1 patient.

We study now the optimal effort of the type 1 patient. His utility is:

$$U^1(e_1, \not q) \;=\; \gamma e_1 E\left[e | \not q\right] - k_1\left(e_1\right)^2$$
$$= \gamma e_1\left[\mu e_1 + (1-\mu)e_2\right] - k_1 e_1. \tag{23}$$

We have shown that $e_2 = 0$ is the optimal strategy for type 2 patients regardless of e_1, then the type 1 patient's utility is:

$$U^1(e_1, \not{c}) = (e_1)^2 (\mu\gamma - k_1). \tag{24}$$

Two cases can be distinguished depending on the value of the marginal cost of adherence k_1 relative to $\gamma\mu$.

A. The "efficient" case: $k_1 < \mu\gamma$

According to condition 24, if $k_1 < \mu\gamma$ (compliant patients have a relatively high propensity to adhere to the treatment), then the optimal effort of the type 1 patient is $e_1 = 1$.

What are the existence conditions in this case? According to the system of beliefs (15), a patient who deviates from the $S = \not{c}$ strategy and pays c will be perceived by physicians as a type 1 patient, thus: $E[e|c] = e_1 = 1$.

What are the existence conditions in this case? According to the system of beliefs (15), a patient who deviates from the $S = \not{c}$ strategy and pays c will be perceived by physicians as a type 1 patient, thus: $E[e|c] = e_1$.

The utility of a "deviant" type 1 who decides to pay c is:

$$\begin{aligned} U^1(e_1, c) &= \gamma e_1 E[e|c] - k_1 (e_1)^2 - c \\ &= (e_1)^2 (\gamma - k_1) - c. \end{aligned} \tag{25}$$

The comparison of the utilities,

$$U^1(e_1, c) = \begin{cases} -c & \text{if } e_1 = 0 \\ (\gamma - k_1) - c & \text{if } e_1 = 1 \end{cases} \tag{26}$$

reveals that the optimal effort of the "deviating" type 1 patient is: $e_1 = 1$, leading to:

$$U^1(e_1 = 1, c) = (\gamma - k_1) - c. \tag{27}$$

Thus, the type 1 patient has no incentive to deviate from $S = \not{c}$ if:

$$U^1(e, \not{c}) > U^1(e, c) \tag{28}$$

$$(\gamma\mu - k_1) > (\gamma - k_1) - c \tag{29}$$

$$c > c_2 = \gamma(1 - \mu) \tag{30}$$

The utility of a "deviant" type 2 patient (who pays c) is:

$$\begin{aligned} U^2(e_2, c) &= \gamma e_2 E\left[e|c\right] - k_2 \left(e_2\right)^2 - c \\ &= \gamma e_2 e_1 - k_2 \left(e_2\right)^2 - c. \end{aligned} \tag{31}$$

Because the equilibrium effort of the type 1 patient is $e_1 = 1$, the the comparison of utilities indicates:

$$U^2(e_2, c) = \begin{cases} -c & \text{if } e_2 = 0 \\ (\gamma - k_2) - c & \text{if } e_2 = 1 \end{cases}$$

It turns out that the optimum "deviating effort" is $e_2 = 0$, leading to the "deviating utility":

$$U^2(0, c) = -c. \tag{32}$$

We compare this utility with equilibrium utility of the type 2 patient.

$$U^2(e_2 = 0, c) = -c < 0 = U^2(e_2 = 0, \not{c}). \tag{33}$$

Obviously, the type 2 patient has not the incentive to deviate from the $S = \not{c}$ strategy.

Thus, for $k_1 < \gamma\mu$, the necessary condition for this equilibrium $(S = \not{c}; e_1 = 1, e_2 = 0)$ to exist is:

$$c > c_2 = \gamma(1 - \mu). \tag{34}$$

with $c_2 < c_1$.

In this "efficient" pooling equilibrium $(S = \not{c}; e_1 = 1; e_2 = 0)$, no patient pays, yet each type is implementing the same effort as in the perfect information case (see: separating equilibrium). However, physicians form imprecise expectations about the type of patient they are facing and implement an average effort level that penalize the type 1 patients compared to the separating situation.

B. The "inefficient" case: $\mu\gamma < k_1 < \gamma$

According to condition 24, if $\mu\gamma < k_1$, then the optimal effort of the type 1 patient is $e_1 = 0$. In this case, a patient's utility is $U^i(e_i, \not{c}) = 0$ regardless of his type.

Which are the existence conditions of this equilibrium ?

If the type 1 patient deviates from the equilibrium strategy, he will pay c and will be clearly identified as a type 1 patient; physician's expectations are $E[e|c] = e_1$. His utility is:

$$\begin{aligned} U^1(e_1, c) &= \gamma e_1 E[e|c] - k_1 (e_1)^2 - c \\ &= -k_1 (e_1)^2 - c. \end{aligned} \tag{35}$$

His optimal effort when deviating from the no-billing strategy is $e_1 = 0$, leading to utility $U^1(0, c) = -c$. He has no incentive to deviate from the equilibrium utility.

Remark that a type 2 patient has no incentive to deviate. Should he decide to pay c, he will be identified as a type 1 patient and because $E[e|c] = e_1 = 0$, his utility would be:

$$U^2(e_2, c) = \gamma e_2 E[e|c] - k_2 (e_2)^2 - c \tag{36}$$

$$= \begin{cases} -c & \text{if } e_2 = 0 \\ -k_2 - c & \text{if } e_2 = 1 \end{cases} \tag{37}$$

The optimal effort of a type 2 patient who deviates and pays c is $e_2 = 0$. We can check that the deviating utility $U^2(e_2 = 0, c) = -c$ is lower than in the equilibrium utility.

Thus, if the complaint patient are not compliant enough ($\mu\gamma < k_1$), regardless of c, there is an "inefficient" pooling equilibrium in which nobody pays the balance billing fee, and both agents implement the low effort level, $(S = \not{c}; e_1 = 0; e_2 = 0)$.

3.3 Equilibria regioning and discussion

To summarize the former findings, the game between physicians and patients presents two pure strategy equilibria: a separating equilibrium and a pooling equilibrium in which no patient opts for the balance billing sector. The feasibility of the equilibria depends on both k_1 and the balance billing fee c.

Table 1 summarizes the properties of the various pure strategy equilibria, and recalls the range of parameters k_1 and c in which they can exist. Figure 1 provides a graphical representation of the regions of existence of the various equilibria.

As we can see,

- For $k_1 < \mu\gamma$ and $c > c_2$, the efficient pooling equilibrium and the separating equilibrium overlap.

- For $k_1 > \mu\gamma$ and $c < c_2$ the inefficient pooling equilibrium and the separating equilibrium overlap.

The separating equilibrium is the only equilibrium only if $c < c_1 = (\gamma - k_1)$ and $k_1 < \mu\gamma$!! (the complyant patients are really eager to comply)

Equilibrium	Condition on c	Additional condition	Optimal effort
Separating	$c < c_1 = (\gamma - k_1)$	–	$e_1 = 1, e_2 = 0$
Pooling $\not c$ efficient	$c > c_2 = \gamma(1 - \mu)$	$k_1 < \mu\gamma$	$e_1 = 1, e_2 = 0$
Pooling $\not c$ inefficient	$c > 0$	$\mu\gamma < k_1 < \gamma$	$e_1 = 0, e_2 = 0$
Hybrid equilibrium (efficient)	$c_2 < c < c_1$		$e_1 = 1, e_2 = 0$

Table 1: Pure Strategy Equilibria of the Patient-Physician Game

DAMIEN, maintenant le hybride m'a l'air stable (avant il etait instable) - none of the type 2 patients pays the extra charge c, while type 1 patients are indifferent between paying it or not. PUISQU'IL EST STABLE et intuitif (plus c augmente moins de gens veulent payer), je me dis qu'il est interessant !! on peu l'ajouter dans le tableau ??

<< FIGURE 1 >>

Discussion

1. The out-of-pocket fee c was defined as the difference between fees charged by physicians, and the reimbursement by the public and private insurance schemes. Our analysis has revealed that a *small* balance billing fee would suffice to bring the benefit of separating types.

2. Our model points out to the risks associated to an increase in c beyond the critical threshold c_2 that separates the separating from the pooling equilibrium. It also revealed the negative consequences of the pooling situation when the "highly adherent" patients have a relatively low propensity to adhere (k_1 is large), because, in this case, patients' effort is at the lowest level for both types.

3. In our analysis, we assumed that patients pay no out-of-pocket charge in the regulated sector. This might not be the case in real life; for instance, in France, after 2005, the National Health Administration imposed a one-euro "co-payment fee" for patients in the regulated sector,

with goal of setting an incentive on patients to avoid unnecessary consultations, i.e., oppose moral hazard. Co-payments with a more complex structure are also applied in Belgium, another country running a mixed payment system. In our model, such a fee paid by every patient, regardless of the choice of sector, has no consequence on the equilibrium solutions (the critical thresholds). However, as a compulsory out-of-pocket payement, it sets a positive lower bound for the fee c. With the co-payment c_0, the separating equilibrium can exist for $c \in [c_0, c_1]$. A higher co-payment, as required by some experts, would only narrow the range of existence of the separating equilibrium.

4. In 2016, worried by the ongoing increase in the free fees set by physicians in sector 2 and the risk of excluding from health care some of the least wealthy patients, the French National Health Administration imposed new rules on private insurers aiming to cap their reimbursement in balance billing arrangements.[12] If physicians cut their fees by the same amount, then the measure would reach its goal without harming the signaling effect (as c is constant). If sector 2 physicians will not reduce their consultation fees by an identical amount, the reform can actually entail a higher patient out-of-pocket payment for medical services (c), with a higher risk for the medical system to switch from the separating to the pooling equilibrium. Whether the lower reimbursement entails a lower or a higher out-pocket-payment, this would be an interesting question for further empirical research.

4 Conclusion

The use of balance billing in the health care sector is a controversial policy. Some scholars argued that the system allows to price discriminate among patients with different willingness-to-pay, and could be efficient if both the sector's profit and patients' welfare are taken into account. On the other hand, if patients' willingness to pay is private information, other studies have shown that balance billing might deteriorate patients overall welfare. In a more macroeconomic perspective, advocates of balance billing argue that the system would help raising physicians income and attract talents to this profession, without an additional burden on taxpayers. Critics argue that balance billing entails excessive charges for patients without a significant gain in quality. In the

[12] This policy builds on an early attempt to curb the supplementary fee in France, as implemented in 2012 (Calcoen and Van den Ven, 2019).

extreme case, price differentiation would create a dual market for medical services that would exclude poorer patients from important cares. Given this relatively long list of disadvantages, it is somehow puzzling why variants of balance billing have a lasting existence in some rich countries, such as France, Belgium, Canada or the US (Medicare).

This paper contributes to the literature on balance billing by emphasizing the signaling properties of this mechanism, in a model that empathizes differences in patient's propensity to adhere to the treatment. Patients propensity to adhere to the treatment is private information to them. They must chose both the billing strategy (observed by the physician), and the optimal effort (unobserved). We use an original health care production function to introduce the assumption according to which physicians own effort in the patient-physician relationship depends on their beliefs about patients' adherence to the treatment. As main limitations of our analysis, this paper does not provide an analysis of the supply side of the market: physicians' choice of the sector of activity (regulated vs. unregulated), and the extra fee under the balance billing system are exogenously given. Individuals were also assumed to have the same disutility of paying the fee, regardless of their wealth.

The analysis of the equilibria has revealed that a mixed payment system, with and without balance billing, might allow to signal patients' propensity to adhere to the treatment. We have shown that a small positive fee charged under balance billing could help patients to signal their type: patients with low propensity to adhere will chose the regulated, free-of-charge sector, and patients with a high propensity to adhere will opt for the balance billing system. In this separating equilibrium, patients chose their optimal effort level, and physicians can identify the type and effort without error.

We could determine analytically the critical fee above which a pooling equilibrium, in which no patient pays the fee and physicians no longer can identify their type, can emerge. In this equilibrium, physicians are worse-off, but so are patients with a high propensity to follow the treatment, since physicians no longer can identify their type and will provide a weaker investment in the doctor-patient relationship.

It is of course a challenge to infer policy recommendations from such a stylized model, and our conclusions should be taken with a significant degree of prudence; on the other hand, our game theoretic approach allowed us to emphasize the signaling virtues of the price mechanism in health care production. Furthermore, empirical research, which guides policymakers most of the time, predicts in general smooth responses in the output variables to policy changes; our approach can explain why sometime small changes in parameters can trigger sharp changes in expectations and behaviors.

Acknowledgement. *The authors would like to thank Robert Nuscheler, Thomas McGuire, Anastasios Dosis, Nicolas Sirven and participants to the 20th European Health Economics Workshop in Verona, Italy, May 17, 2019 for their suggestions and remarks that helped them to improve the quality of the paper.*

References

Arrow, K. J., 1963. Uncertainty and the welfare economics of medical care. *American Economic Review*, 53, 941–973.

Balsa, A. I. and McGuire T. G., 2003. Prejudice, clinical uncertainty and stereotyping as sources of health disparities, *Journal of Health Economics*, 22: 89-116.

Barros, P. P., and L. Siciliani, 2011. Public and private sector interface, In: *Handbook of Health Economics*. Vol. 2. Elsevier: 927-1001.

Besancenot, D. and Vranceanu R., 2017. An equilibrium search model of the French market for medical services, *ESSEC Working Paper* 1709.

Buller, D. B., and R. L. Street., 1992. Physician-patient relationships, In: R. S. Feldman (Eds.) *Applications of Nonverbal Behavioral Theories and Research*, Taylor and Francis: 119-141.

Benjamin, R. M., 2012, Medication adherence: helping patients take their medicines as directed. *Public Health Reports*, 127, (1): 2-3.

Calcoen, P., and Van de Ven W. P.M.M., 2019. Supplementary physicians' fees: a sustainable system? *Health Economics, Policy and Law,* 14, (1): 40-60.

Charles, C., A. Gafni, and Whelan T., 1997. Shared decision-making in the medical encounter: what does it mean? (or it takes at least two to tango), *Social Science & Medicine,* 44, (5): 681-692.

Coudin E., Pla A., and Samson A. L., 2015. GPs' response to price regulation : evidence from a nationwide French reform. *Health Economics,* 24, (9): 1118-1130.

Clerc, I., L'Haridon, O., Paraponaris, A., Protopopescu, C. and Ventelou B., 2012. Fee-for-service payments and consultation length in general practice: a work–leisure trade-off model for French GPs. *Applied Economics*, 44, (25): 3323-3337.

Davis, M. S. 1968. Variations in patients' compliance with doctors' advice: an empirical analysis of patterns o communication. *American Journal of Public Health and the Nations Health*, 58, (2): 274-288.

Derksen, Frans, Jozien Bensing, and Antoine Lagro-Janssen. 2013. Effectiveness of empathy in general practice: a systematic review. *British Journal of General Practice*, 63: e76-e84.

DiMatteo, M. R., 2004. Variations in patients' adherence to medical recommendations: a quantitative review of 50 years of research, *Medical Care*, 42, (3), 200-209.

DiMatteo, M. R., Giordani, P. J., Lepper, H. S., and Croghan, T. W., 2002. Patient adherence and medical treatment outcomes a meta-analysis. *Medical Care*, 40, (9): 794-811.

Dormont B., and Péron M., 2016. Does health insurance encourage the rise in medical prices? A test on balance billing in France. *Health Economics*, 25: 1073–1089.

DREES, 2018. *Les Dépenses de Santé en 2017 - Edition 2018*, collection Panoramas de la Drees, Paris, 2018.

Eggleston, K., and Bir, A., 2006. Physician dual practice. *Health Policy*, 78, (2-3): 157-166.

Ellis, R. P., and T. G. McGuire, 1986. Provider behavior under prospective reimbursement: Cost sharing and supply, *Journal of Health Economics* 5.2: 129-151.

Epp M., Vining A., Collins-Dodd C., and Love E., 2000. The impact of direct and extra billing for medical services: evidence from a natural experiment in British Columbia. *Social Science & Medicine,* 51, (5): 691–702.

Fehr, E., and Gächter, S., 2000. Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives*, 14, (3): 159-181.

Feldman R., and Sloan F., 1988. Competition among physician, revisited. *Journal of Health Politics, Policy and Law*, 13: 239–261.

Fichera, E., Banks, J., Siciliani, L., and Sutton, M., 2018. Does patient health behaviour respond to doctor effort?. *Journal of Economic Behavior & Organization*, 156: 225-251.

Giuffrida, A., and H. Gravelle, 1998. Paying patients to comply: an economic analysis. *Health Economics* 7, (7): 569-579.

Glazer J., McGuire T.G., 1993. Should physicians be permitted to balance bill patients? *Journal of Health Economics,* 12 (3): 239–258.

Godager, Geir, and Daniel Wiesen, 2013. Profit or patients' health benefit? Exploring the heterogeneity in physician altruism. *Journal of Health Economics* 32, 6: 1105-1116.

Gravelle H., Scott A., Sivey P., and Yong J., 2016. Competition, prices and quality in the market for physician consultations. *Journal of Industrial Economics*, 64, (1): 135-169.

Haynes R.B., 1979. Determinants of compliance: the disease and the mechanics of treatement, In: *Compliance in Health Care*. Eds. Haynes, R. Brian. Taylor, D. Wayne and Sackett, D.L., John Hopkins Univeristy Press: Baltimore, 49-62.

Heritage, and Maynard, D. W., 2006. Problems and prospects in the study of physician-patient interaction: 30 years of research. *Annual Review of Sociology,* 32: 351-374.

Iuga, A. O., & McGuire, M. J., 2014. Adherence and health care costs. *Risk Management and Healthcare Policy*, 7: 35–44.

Jelovac, I., 2015. Physicians' balance billing, supplemental insurance and access to health care, *International Journal of Health Economics and Management,* 15, (2): 269-280.

Kifmann M. and Scheuer F., 2011. Balance billing: the patients' perspective. *Health Economics Review,* 1 (1): 1–14.

Kelm, Z., Womer, J., Walter, J. K., & Feudtner, C., 2014. Interventions to cultivate physician empathy: a systematic review, *BMC Medical Education*, 14, (1): 219.

Kesternich, I., H. Schumacher, and J. Winter, 2015. Professional norms and physician behavior: homo oeconomicus or homo hippocraticus?. *Journal of Public Economics* 131: 1-11.

Kuhn, M. and Nuscheler, R., 2013. Saving the public from the private? Incentives and outcomes in dual practice, Vienna University of Technology Working Papers in Economic Theory and Policy, No. 02/2013

Lamiraud, K., and Geoffard, P.Y., 2007. Therapeutic non-adherence: a rational behavior revealing patient preferences?, *Health Economics*, 16, (11), 1185-1204.

Lecluyse A., Van de Voorde C., De Graeve D., Schokkaert E., and Van Ourti T., 2009. Hospital supplements in Belgium: price variation and regulation. *Health Policy,* 92, (2): 276–287.

McGuire, T. G., 2000. Physician agency, In: *Handbook of Health Economics.*, 1. Elsevier, 2000: 461-536.

McKnight R., 2007. Medicare balance billing restrictions: impacts on physicians and beneficiaries. *Journal of Health Economics,* 26, (2): 326–341.

Mitchell J. B., Cromwell J., 1982. Physician behavior under the Medicare assignment option. *Journal of Health Economics,* 1, 245-264

Mu, C., De Abreu Lourenco, R., van Gool, K., and Hall, J., 2018. Is low-priced primary care bad for quality? Evidence from Australian general practice. *Applied Economics*, 5, (5): 475-491.

Parsons T., 1951. *The Social System.* New York: Free Press.

Savage E. and Jones G., 2004. An analysis of the General Practice Access Scheme on GP incomes, bulk-billing and consumer copayments. *Australian Economic Review,* 37, (1): 31-40

Simpson, S. H., Eurich, D. T., Majumdar, S. R., Padwal, R. S., Tsuyuki, R. T., Varney, J., and Johnson, J. A., 2006. A meta-analysis of the association between adherence to drug therapy and mortality. *British Medical Journal*, 333 (7557), 15.

Spence, M., 1973. Job Market Signaling. *Quarterly Journal of Economics,* 87, (3): 355-374.

Spence, M., 2002. Signaling in retrospect and the informational structure of markets, *American Economic Review*, 92, (3): 434-459.

Vermeire E., Hearnshaw H., Van Royen P., Denekens J., 2001, Patient adherence to treatment: three decades of research. A comprehensive review, *Journal of Clinical Pharmacy and Therapeutics*, 26, (5): 331-342.

Viswanathan M., Golin C.E., Jones C.D., Ashok M., Blalock S.J., Wines R.C., et al., 2012. Interventions to improve adherence to self-administered medications for chronic diseases in the United States: A systematic review. *Annals of Internal Medecine*, 157: 785–795.

Zuckerman S. and Holahan J., 1991. The role of balance billing in medicare physician payment reform. In *Regulating Doctors' Fees: Competition, Benefits and Controls under Medicare*. Edited by: HE Frech, III. AEI Press, Washington, DC: 143–169.

A FINALISER —>>>>>>>>>>>>>>

A Appendix 1. Non-existence of the pooling equilibrium c

DAMIEN JE NE COMPRENDS PLUS RIEN ICI !!!

We prove the non-existence of an equilibrium in which all pay the fee c. Physicians' beliefs are:[13]

$$\begin{cases} \Pr[\text{type } 1|c] = \mu \\ \Pr[\text{type } 1|\not{c}] = 0 \end{cases} \tag{38}$$

Let us start by studying the patients' optimal effort. The utility of type 1 patient is:

$$\begin{aligned} U^1(e_1, c) &= \gamma e_1 E\left[e|c\right] - k_1 \left(e_1\right)^2 - c \\ &= \gamma e_1 \left[\mu e_1 + (1-\mu)e_2\right] - k_1 \left(e_1\right)^2 - c \end{aligned} \tag{A.39}$$

Assume first that the optimal effort of the type 2 patient is $e_2 = 1$ and compare the utility of type 1 for the two feasible level of effort:

$$U^1(e_1, c) = \begin{cases} (\gamma - k_1) - c & \text{if } e_1 = 1 \\ -c & \text{if } e_1 = 0 \end{cases} \tag{40}$$

When type 2 patients choose effort $e_2 = 1$, the optimal effort of a type 1 patient is also $e_1 = 1$.

Let us now assume that the optimal effort of type 2 patient is $e_2 = 0$. The utility of the type 1 patient is then:

$$U^1(e_1, c) = \left(e_1\right)^2 (\mu\gamma - k_1) - c$$

Depending on the relative values of $\mu\gamma$ and k_1, the optimal effort of the type 1 can be either 0 or 1. If $k_1 < \mu\gamma$ (respectively $k_1 > \mu\gamma$) the optimal effort of type 1 patient is $e_1 = 1$ (respectively $e_1 = 0$).

We analyze now the *existence conditions.*

Case 1. $k_1 < \mu\gamma$

[13] Note that with the alternative out-of-equilibrium beliefs, $\Pr[\text{type } 1|\not{c}] = 1$, a type 1 patient would always find optimal to deviate. As strategy $\not{c}$ would reveal his type, playing $\not{c}$ allows not only to avoid the out-of-pocket payment c but also to benefit from the high physician effort.

In this case, the optimal effort strategy for type 1 patient is $e_1 = 1$ whatever e_2. Consider then the type 2 patient; his utility is:

$$\begin{aligned} U^2(e_2, c) &= \gamma e_2 E\left[e|c\right] - k_1 \left(e_2\right)^2 - c \\ &= \gamma e_2 \left[\mu + (1-\mu)e_2\right] - k_2 \left(e_2\right)^2 - c \end{aligned} \tag{A.41}$$

A type 2 patient obtains a utility level $U^2(e_2 = 1, c) = (\gamma - k_2) - c$ if he chooses the high effort level ($e_2 = 1$), and utility $U^2(e_2 = 0, c) = -c$ if $e_2 = 0$. Because $k_2 > \gamma$, in the pooling 1 equilibrium, the optimal effort of type 2 patient is $e_2 = 0$ leading to utility:

$$U^2(e_2 = 0, c) = -c \tag{42}$$

However, if the type 2 patient deviates and does not pay c, physicians acknowledge that he is of the type 2. His utility is:

$$U^2(e_2, \not{c}) = \left(e_2\right)^2 \left(\gamma - k_2\right).$$

Because $(\gamma - k_2) < 0$, his optimal effort is $e_2 = 0$, leading to a "deviation" utility $U^2(0, \not{c}) = 0$. Obviously, *it is always optimal for type 2 to deviate*:

$$U^2(e_2 = 0, c) = -c < 0 < U^2(0, \not{c}) \tag{43}$$

The equilibrium is impossible in the case $k_1 < \mu\gamma$.

Case 2. $k_1 > \mu\gamma$

In this case, when the type 2 patient's effort is $e_2 = 0$, the optimal effort of the type 1 patient is also $e_1 = 0$.

(a) If the optimal effort of the type 1 patient is $e_1 = 0$, the type 2 patient obtains utility $U^2(e_2 = 0, c)$ if he accepts to pay c:

$$\begin{aligned} U^2(e_2, c) &= \gamma e_2 \left[(1-\mu)e_2\right] - k_2 \left(e_2\right)^2 - c \\ &= \left(e_2\right)^2 \left[\gamma(1-\mu) - k_2\right] - c \end{aligned}$$

Because $[\gamma(1-\mu) - k_2] < 0$, the optimal effort is $e_2 = 0$ and therefore the type 1 patient must choose $e_1 = 0$.

JE NE COMPREND PLUS RIEN....

In this case, $e_1 = e_2 = 0$ implies $E[e|c] = 0$ and in equilibrium both type of patients reach the same utility level, $U^i(e_i = 0, c) = -c$ for $i = 1$ or 2.

It is straightforward to see that, like in case 1, the type 2 patient finds optimal to deviate and to refuse the payment c. Given the equilibrium beliefs, if the type 2 patient does not pay the supplementary fee, physicians acknowledge that he is of the type 2. The utility of the type 2 is :

$$U^2(e_2, c) = (e_2)^2 [\gamma(1-\mu) - k_2] \tag{44}$$

and his optimal effort is $e_2 = 0$, leading to utility $U^2(0, \cancel{c}) = 0$ which is higher than the equilibrium utility $U^2(e_2, c) = -c$. The equilibrium is impossible.

(b) If the optimal effort of the type 1 patient is $e_1 = 1$, type 2 patient gets utility $U^2(e_2, c)$ if he accepts to pay c:

$$U^2(e_2, c) = \gamma e_2 [\mu + (1-\mu)e_2] - k_2 (e_2)^2 - c$$

According to his effort level, type 2 patient gets utility:

$$\begin{cases} U^2(e_2, c) = \bar{e}(1-k_2) - c & \text{if } e_2 = \bar{e} \\ U^2(e_2, c) = -c & \text{if } e_2 = 0 \end{cases} \tag{45}$$

The optimal effort is thus $e_2 = 0$ which implies the optimal effort $e_1 = 0$ for type 1 patient and is inconsistent with the initial assumption. In this case, the equilibrium is impossible.

In case 1 and case 2, the equilibrium beliefs are inconsistent with the generalized adoption of the balance billing system, the Pooling 2 equilibrium is impossible.

B Appendix 2. Hybrid equilibrium A

We analyze the equilibrium in which none of the type 2 patients pays the extra charge c, while type 1 patients are indifferent between paying it or not.

Let us denote by $(1-\nu)$ the proportion of patients 1 who pay the extra fee (v do *not* pay it).

Physician's beliefs are:

$$\begin{cases} \Pr[\text{type } 1|c] = 1 \\ \Pr[\text{type } 1|\not{c}] = \frac{\Pr[\not{c}|\text{type } 1]\Pr[\text{type } 1]}{\Pr[\not{c}|\text{type } 1]\Pr[\text{type } 1]+\Pr[\not{c}|\text{type } 2]\Pr[\text{type } 2]} = \frac{\nu\mu}{\nu\mu+(1-\mu)} \end{cases} \tag{46}$$

and:

$$E\left[e|c\right] = e_1 \tag{B.47}$$

$$E\left[e|\not{c}\right] = \left[\frac{\nu\mu}{\nu\mu+(1-\mu)}e_1 + \frac{(1-\mu)}{\nu\mu+(1-\mu)}e_2\right] \tag{B.48}$$

The optimal effort

Consider first the type 2 patient. In equilibrium, this patient does not pay (strategy $\not{c}$) and his utility is:

$$\begin{aligned} U^2(e_2,\not{c}) &= \gamma e_2 E\left[e|\not{c}\right] - k_2\left(e_2\right)^2 & \text{(B.49)} \\ &= \gamma e_2\left[\frac{\nu\mu}{\nu\mu+(1-\mu)}e_1 + \frac{(1-\mu)}{\nu\mu+(1-\mu)}e_2\right] - k_2\left(e_2\right)^2. & \text{(B.50)} \end{aligned}$$

If $e_1 = 0$, the type 2 patient utility is

$$U^2(e_2,\not{c}) = \left(e_2\right)^2\left(\gamma\left[\frac{(1-\mu)}{\nu\mu+(1-\mu)}\right] - k_2\right)$$

and, because $\gamma\frac{(1-\mu)}{\nu\mu+(1-\mu)} < \gamma < k_2$, his optimal effort is $e_2 = 0$.

If $e_1 = 1$, the type 2 patient utility is:

$$U^2(e_2,\not{c}) = \gamma e_2\left[\frac{\nu\mu}{\nu\mu+(1-\mu)} + \frac{(1-\mu)}{\nu\mu+(1-\mu)}e_2\right] - k_2\left(e_2\right)^2$$

taking the effective values:

$$U^2(e_2,\not{c}) = \begin{cases} 0 & \text{if } e_2 = 0 \\ (\gamma - k_2) < 0 & \text{if } e_2 = 1 \end{cases} \tag{51}$$

In this case too, the optimal effort for type 2 patient is $e_2 = 0$.

We conclude that, in this equilibrium, the type 2 patient has a dominant effort strategy, $e_2 = 0$. His equilibrium utility is $U^2(e_2 = 0,\not{c}) = 0$.

We turn not to analyzing the optimal choice of the type 1 patient. In this mixed strategy situation, he/she must be indifferent between paying or not the fee c.

If he/she pays c, his/her type is detected and his/her utility is $U^1(e_1, c) = (\gamma - k_1)(e_1)^2 - c$; the optimal effort is $e_1 = 1$ leading to effective utility $U^1(e_1 = 1, c) = (\gamma - k_1) - c$.

If he/she does not pay ($\not{c}$), his/her utility is (recall that $e_2 = 0$):

$$
\begin{aligned}
U^1(e_1, \not{c}) &= \gamma e_1 E\left[e|\not{c}\right] - k_1 (e_1)^2 \\
&= (e_1)^2 \left(\gamma \left[\frac{\nu\mu}{\nu\mu + (1-\mu)}\right] - k_1\right)
\end{aligned}
\tag{B.52}
$$

The optimal effort is therefore:

$$
\begin{cases}
e_1 = 1 & \text{if } \left(\gamma \left[\frac{\nu\mu}{\nu\mu+(1-\mu)}\right] - k_1\right) > 0 \\
e_1 = 0 & \text{if } \left(\gamma \left[\frac{\nu\mu}{\nu\mu+(1-\mu)}\right] - k_1\right) < 0
\end{cases}
\tag{53}
$$

The resulting effective utilities are:

$$
\begin{cases}
\left(\gamma \left[\frac{\nu\mu}{\nu\mu+(1-\mu)}\right] - k_1\right) > 0 \Leftrightarrow U^1(e_1 = 1, \not{c}) = \left(\gamma \left[\frac{\nu\mu}{\nu\mu+(1-\mu)}\right] - k_1\right) \\
\left(\gamma \left[\frac{\nu\mu}{\nu\mu+(1-\mu)}\right] - k_1\right) < 0 \Leftrightarrow U^1(e_1 = 0, \not{c}) = 0
\end{cases}
\tag{54}
$$

The existence conditions

- Case $\left(\gamma \left[\frac{\nu\mu}{\nu\mu+(1-\mu)}\right] - k_1\right) < 0$ ou $\gamma \left[\frac{\nu\mu}{\nu\mu+(1-\mu)}\right] < k_1 < \gamma$.

In this case, $e_1 = 0$. The type 1 patient must be indifferent between paying or not.

ON AVAIT DIT "Because $U^1(e_1 = 0, \not{c}) = 0$ and $U^1(0, c) = -c$, this equilibrium is impossible."

DAMIEN ?? ICI CA CHANGE ?? $U^1(0, \not{c}) = 0$ and $U^1(e_1 = 1, c) = (\gamma - k_1) - c$. C'est bien POSSIBLE POUR $c = (\gamma - k_1)$?? (une simple droite) // mais c'est quoi v ??

ON TROUVAIT AVANT QUE C'EST IMPOSSIBLE.

- Case $\left(\gamma \left[\frac{\nu\mu}{\nu\mu+(1-\mu)}\right] - k_1\right) > 0$ ou $k_1 < \gamma \left[\frac{\nu\mu}{\nu\mu+(1-\mu)}\right]$.

In this case, the optimal effort is $e_1 = 1$.

Firstly we notice that the type 2 patient has no incentive to deviate. If he pays c and is taken for type 1 doing effort 1, his utility is :

$$
U^2(e_2, c) = \gamma e_2 - k_2 (e_2)^2 - c. \tag{55}
$$

This utility is equal to $-c$ if $e_2 = 0$ and takes value $\gamma - k_2 - c < -c$ if $e_2 = 1$. The optimal effort is then $e_2 = 0$ and the utility $U^2(e_2 = 0, c) = -c < 0 = U^2(e_2 = 0, \not{c})$. Paying c is suboptimal.

As already mentioned, in this equilibrium, the type 1 patient must be indifferent between paying or not:

$$U^1(e_1 = 1, \not{c}) = U^1(e_1 = 1, c) \tag{B.56}$$

$$\left(\gamma\left[\frac{\nu\mu}{\nu\mu + (1-\mu)}\right] - k_1\right) = (\gamma - k_1) - c \tag{B.57}$$

This condition implicitly defines ν depending on the parameters.

$$\left(\frac{\nu\mu}{\nu\mu + (1-\mu)}\right) = \frac{\gamma - k_1 - c}{\gamma} \tag{B.58}$$

$$\frac{(1-\mu)}{\nu\mu} = \frac{\gamma}{\gamma - k_1 - c} - 1 \tag{B.59}$$

$$\frac{(1-\mu)}{\nu\mu} = \frac{k_1 + c}{\gamma - k_1 - c} \tag{B.60}$$

$$\nu = \frac{(1-\mu)}{\mu}\frac{\gamma - k_1 - c}{k_1 + c} \tag{B.61}$$

which is a monontonic decreasing function in c. We check the conditions for which $\nu \in [0, 1]$.

$$\frac{(1-\mu)}{\mu}\frac{\gamma - k_1 - c}{k_1 + c} > 0 \Leftrightarrow c < c_1 = (\gamma - k_1) \tag{62}$$

and

$$\frac{(1-\mu)}{\mu}\frac{\gamma - k_1 - c}{k_1 + c} < 1 \Leftrightarrow c > (1-\mu)\,\gamma - k_1 \tag{63}$$

QUID DE $k_1 < \gamma\left[\frac{\nu\mu}{\nu\mu+(1-\mu)}\right]$????

The equilibrium is feasible for an out-of-pocket payment c in the range $[c_2, c_1]$.

In Figure 1, the hybrid equilibrium is situated WHERE ??? Une bande diagonale ?

Recall that ν is the frequency of the patients who do *not* pay c. Eq. (61) defines a decreasing relationship between ν and c. In other words, for a high out-of-pocket payment (c), the equilibrium number of type 1 patients who refuse to pay the fee decreases.

[DAMIEN, je me trompe ??? On disait l'envers avant, qu'il est instable !!! Mais si, il est stable ! Faut-il le mettre dans le texte ??? Puisqu'il est stable !]

C Appendix 3. Non-existence of the hybrid equilibrium B

We can show that a hybrid equilibrium in which all of the type 1 patients pay c, while the type 2 patients are indifferent between paying or not, is impossible.

Let us denote by with ρ the proportion of type 2 patients who decide to pay. The beliefs of the physicians are:

$$\begin{cases} \Pr[\text{type } 1|\not c] = 0 \\ \Pr[\text{type } 1|c] = \frac{\Pr[c|\text{type } 1]\Pr[\text{type } 1]}{\Pr[c|\text{type } 1]\Pr[\text{type } 1]+\Pr[c|\text{type } 2]P[\text{type } 2]} = \frac{\mu}{\mu+(1-\mu)\rho} \end{cases} \tag{64}$$

Let us consider type 2 patients' behavior. When such a patient plays $\not c$, he reveals his type and his utility is:

$$U^2(e_2, \not c) = \gamma e_2 E\left[e|\not c\right] - k_2 \left(e_2\right)^2 = \left(e_2\right)^2 \left(\gamma - k_2\right)$$

Obviously the optimal effort is $e_2 = 0$ (recall that $k_2 > \gamma$), and the effective utility is $U^2(e_2 = 0, \not c) = 0$

If the type 2 plays c, his utility is:

$$U^2(e_2, c) = \gamma e_2 \left[\frac{\mu}{\mu + (1-\mu)\rho} e_1 + \frac{(1-\mu)\rho}{\mu + (1-\mu)\rho} e_2\right] - k_2 \left(e_2\right)^2 - c \tag{C.65}$$

$$= \begin{cases} -c & \text{if } e_2 = 0 \\ \gamma \left[\frac{\mu}{\mu+(1-\mu)\rho} e_1 + \frac{(1-\mu)\rho}{\mu+(1-\mu)\rho}\right] - k_2 - c & \text{if } e_2 = 1 \end{cases}. \tag{C.66}$$

Because $\gamma \left[\frac{\mu}{\mu+(1-\mu)\rho} e_1 + \frac{(1-\mu)\rho}{\mu+(1-\mu)\rho}\right] - k_2 < 0$ whatever e_1, the optimal strategy for type 2 patients is again $e_2 = 0$. However, in this hybrid equilibrium, type 2 must be indifferent between the two billing strategies. As $e_2 = 0$ is optimal whatever e_1, $U^2(e_2 = 0, c) = -c$ while $U^2(e_2 = 0, \not c) = 0$. The indifference condition cannot be true; the equilibrium is impossible.