Should Physicians Team Up to Treat Chronic Diseases?*

Malte Griebenow†

October 11, 2022

Abstract

This paper studies referral strategy and effort provision of a primary care physician (PCP) and a specialist who are responsible for the treatment of chronically ill patients. Two organizational settings are compared, a team in which physicians cooperate and solo practices in which they do not. If the difference in expected treatment costs between disease severities is relatively larger for the PCP, an efficient flow of patients can be achieved in the physician team. In this case, effort is incentivized by a markup on PCP treatment. Otherwise, care may be delivered second-best efficiently in solo practices with a gatekeeping PCP.

Keywords: referral, chronic care, PCP, specialist, moral hazard, team

JEL Codes: I11, D61, D82, M52

*This study was supported by a research grant from the Federal Ministry of Education and Research in Germany (grant number BMBF 01EH1601A).

†Universität Hamburg, Department of Socioeconomics and Hamburg Center for Health Economics, Esplanade 36, 20354 Hamburg, Germany. Email: Malte.Griebenow@uni-hamburg.de.

1. Introduction

Chronic diseases are a costly burden on health care systems. In the US, for example, 60% of the population suffer from at least one chronic disease. Further, chronically ill patients account for 90% of health care expenditures (Buttorff, 2017). According to Bodenheimer et al. (2009), one barrier to the efficient provision of chronic care is that health care systems are built around the treatment of acute problems rather than the long-term health of the patient. One of the measures they propose to improve chronic care is to provide it in teams rather than solo practices. In a meta analysis, Pascucci et al. (2020) find that inter-professional collaboration improves a number of health related outcomes for chronically ill patients indeed. Lemieux-Charles and W. L. McGuire (2006) analyze studies relating to the effectiveness of health care teams compared to usual care. They also find improvements in patient treatment for some interventions. However, they also find that team care may increase costs. This begs the question whether team care is more cost-efficient than care delivered in solo practices. This paper's aim is to provide theoretical guidance regarding this question.

In this paper, I consider two important aspects of chronic care. Firstly, chronically ill patients should receive treatment from an appropriate physician. Which physician should treat a patient crucially depends on the patient's disease severity. Whereas a primary care physician (PCP) is able to cost-efficiently treat a patient in mild condition, a specialist's services are required for a more severe case. This aspect is especially important for chronic diseases as the disease severity of chronically ill patients may change over time. Secondly, treatment efforts exerted today impact health outcomes and costs in the future. The preventive effort of a PCP can decrease the need for future treatment and, thus, decrease costs for the health care system (Dusheiko et al., 2011; Bruin et al., 2001; Li et al., 2010). Similarly, high-quality specialist treatment can lead to quicker recovery of the patient.

In the main part of the paper I consider two profit-oriented physicians, a PCP and a specialist, who do not internalize patient benefit[1]. A profit-oriented physician may provide too little effort because she suffers the costs of effort provision but does not internalize the future health losses of patients. Furthermore, she ignores patient health losses when considering whether to refer a patient[2]. As an extension I consider partially altruistic physicians, who care about both their own profits and the patient's wellbeing.

[1]For simplicity, I refer to the physicians with female pronouns and the payer and patient with male pronouns.

[2]Empirical evidence suggests that financial incentives indeed influence physicians' patient selection (Sarma et al., 2018; Iversen and Lurås, 2000), which may result in both over- and under-referrals (Mehrotra et al., 2011).

If physicians work in solo practices, they do not consider the other physician's profit in their treatment and effort decisions. This can lead to cost-inefficiencies if patients who could be treated less expensively by the other physician are not referred. Furthermore, too little treatment effort may be provided because cost savings generated by the effort accrues to both physicians. If there is no asymmetric information between the physicians, this *coordination problem* can be solved by delivering chronic care in health care teams that are reimbursed by (risk-adjusted) per patient payments as Bodenheimer et al., 2009 propose[3]. In this case, cost savings that accrue because of a physician's decisions can be transferred to that physician with the help of internal profit sharing rules. However, solving the coordination problem is not necessarily socially efficient if physicians do not fully internalize the patient's health losses. In this case, they may choose a treatment path that minimizes costs at the expense of the patient's health. Furthermore, organizing physicians in a team provides them an opportunity to collude in order to earn larger profits. For example, if specialists are paid larger treatment fees than PCPs, there is an incentive for the PCP to over-refer patients to the specialist. Conversely, assuming that kickback payments between physicians are not allowed, the PCP faces no such incentive in the solo practices. Due to the reasons outlined above, it is not clear which organizational form is superior froma social perspective[4].

The aim of this paper is to answer the following question. Under which conditions should a chronically ill patient receive care from a physician team (PCP and specialist) or from independent physicians who work in solo practices? In order to answer this question I derive optimal treatment fees for both physicians in each setting and compare the second-best optimal outcomes between both organizational forms. As an extension, I consider the restriction that the team is paid by a flat treatment fee which does not differ between physicians. The main difference between organizational forms is that in a team, physicians coordinate their referral and effort decisions, whereas in solo practices they do not.

To answer the research question, I develop a model with a PCP and a specialist who treat a fixed number of chronic patients for an indefinite time frame. Patients can either be in a mild condition, which is inexpensive to treat, or in a severe condition, which is costly to treat. The severe condition could, for example, correspond to a diabetes patient who is hyperglycemic or suffers from neuropathic or retinopathic complications. Physicians can exert tertiary preventive effort (time spent on patient, self-help support, appropriate medication, support personnel...) in order to lower the probability that a patient's condition deteriorates or they can exert curative effort in order to increase the probability that

[3]For example, US accountable care organizations (ACOs) use shared budgetary responsibility between physicians in order to make them financially responsible for their own referrals (Song et al., 2014).

[4]Interestingly, the way physicians are organized varies drastically from country to country. For example, in Germany 60% of ambulatory care physicians work in solo practices (Blümel and Busse, 2015). In the US only 18% of physicians work in solo practices (Kash and Tan, 2016).

a patient's condition improves. Further, physicians refer patients between each other and can accept or reject each other's referrals.

The main innovation of the paper is to analyze physicians' agency problems relating to effort and referral efficiency in a model that captures the dynamic nature of the chronic care market. Patients' severity in each period is determined by a competitive Markov decision process (see Filar and Vrieze, 1996, for the theoretical background for this type of game). A patient's probability distribution over the disease severity in the next period is determined by the patient's current severity, the type of physician treating him, and the exerted effort of the treating physician. Consequently, treatment decisions made by physicians in one period affect the expected costs of care and expected patient health losses in all periods to follow.

Both the severity of the patient (hidden information) and the effort exerted by the physicians (hidden action) are unknown to the payer contracting with the physicians. In order to achieve efficient outcomes, both physicians should be incentivized to exert effort. Further, patients in the mild condition should receive care from the PCP and patients in the severe condition should receive specialist care. I derive conditions under which physicians in each organizational form exert more effort and/or more adequately refer patients.

There are several advantages of using a dynamic model rather than a static model. First, it captures the provision of non-contractible effort without reference to altruism or pay-for-performance mechanisms that require the payer to have information on outcomes ex-post. Instead, physicians provide effort in order to reduce their own (or their team's) future costs of care. Second, a dynamic model allows for the study of the complete treatment path of the patient. This includes a back-referral to the PCP after successful specialist treatment.

I find that if profit-maximizing physicians work in solo practices, it is not possible for the payer to implement optimal referral patterns, though it may be possible in the team. Whether or not this is possible depends on the cost structure and the effectiveness of the physicians' treatments. In particular, the expected treatment cost differences between the patient types need to be relatively large for the PCP and small for the specialist, i.e. the PCP must have the relative cost-advantage when treating patients in mild condition. Markups should be used for PCP treatment, whereas the specialist should be paid below-cost. In the converse case, in which the specialist has the relative cost-advantage for mildly ill patients, it can be optimal to organize the physicians in solo practices. This allows the PCP to act as a gatekeeper for the specialist. Mildly ill patients are initially received by the PCP and only referred to specialist care when their state deteriorates. This treatment path cannot be implemented in the team because the PCP would always refer mildly ill patients if doing so increases the team's profits.

The remainder of the paper is structured as follows. Section 2 reviews the theoretical literature on the topic. Section 3 describes the model used in this paper. Section 4 defines the first-best benchmark. In Section 5, treatment fees for both the team and the solo practice are derived to implement potentially second-best optimal outcomes under the assumption that the payer cannot verify effort provision or the type of the patient. Subsequently, the second-best optimal outcomes for team and solo practice cases are compared. Conditions are derived under which either organizational form is superior. In Section B of the appendix, the case that teams are paid with flat fees and the case that physicians are partially altruistic are considered as extensions. Section 6 concludes.

2. Literature Review

Hey and Patel (1983) are the first to develop a Markov model in order to analyze prevention and cure investments of an individual. Hennessy (2008) extends the analysis of Hey and Patel and finds that prevention efforts and cure efforts can be both complements and substitutes. In particular, a subsidy for curative effort may increase the prevalence of an adverse health state as individuals exert less prevention effort. A limitation of these papers is that they do not consider physician agency issues relating to the provision of chronic care.

This paper is also related to the literature on organizational design in expert markets (Jelovac and Macho-Stadler, 2002; Grassi and Ma, 2016). Grassi and Ma (2016) study a referral market between experts who each provide cost-efficient treatment for one type of client. They find that forming an organization is beneficial for referral efficiency. However, this reduces incentives for cost-control. In their model, cost is the only factor that determines which expert should optimally serve a client. By contrast, in my model the physicians not only differ in their costs but also in their ability to treat patient types. Consequently, cost-minimizing treatment may not be socially efficient. Furthermore, they focus on information asymmetry between physicians, which is not a factor in my model. Jelovac and Macho-Stadler (2002) finds that a payer delegating a hospital to contract with its physician may be superior to contracting with both parties simultaneously. This is the case if the hospital's investment decisions matter sufficiently much for the quality of care. Instead of considering subcontracting between hospital and physician, I consider cooperation between two physicians.

Garcia-Mariñoso and Jelovac (2003), Shumsky and Pinker (2003), Allard et al. (2011), and Griebenow and Kifmann (2022) study referrals in health care markets. The focus of these papers is on setting up efficient payment mechanisms to incentivize appropriate referrals from a PCP to a specialist in a static context. Garcia-Mariñoso and Jelovac (2003) study PCP diagnostic effort and referral decision. They find that gatekeeping is superior whenever physician incentives matter. Shumsky and Pinker (2003) consider a PCP gatekeeper, who not only has an information advantage with regard to the optimal treatment decision but also his own ability. A bonus for patient volume in addition to bonuses based on referral rates may be necessary for first-best performance in their setting. Allard et al. (2011) study PCPs with heterogeneous altruism and ability. They find that FFS and fundholding provide similar referral incentives for PCPs. Griebenow and Kifmann (2022) consider referral patterns between altruistic physicians (PCP and specialist). They find that bonus payments either for immediate PCP treatment or for specialist back-referral can improve the flow of patients. Informational requirements of the optimal contracts can be large. For example, information on the ex-post benefit of a treatment needs to be available in some cases in Garcia-Mariñoso and Jelovac (2003) and Allard et al. (2011). In contrast, I consider simple treatment fees that may not always be able to incentivize first-best solutions but are easy to implement by the payer.

Physician services are classic examples of credence goods (see Dulleck and Kerschbamer, 2006, for a survey). In this paper the payer of health care services can verify which type of treatment a patient received but not whether it was appropriate. This is also not revealed ex-post. Thus, both over- and under-provision of care can be issues. Furthermore, the payer cannot observe the physicians' effort choices leading to a potential under-provision of effort. Over-charging is not a problem as the payer pays fixed treatment fees per physician.

Malcomson (2005) studies the optimal treatment choice of a physician that can provide one of two treatments to patients that differ in their disease severity and treatment costs. The payer is unable to verify the severity of the patients, though he is able to verify which treatment has been provided. It is only possible to implement the optimal treatment choice of the physician if the costs of the treatment that is appropriate for less severely ill patients rise more strongly in patient severity than the costs of the treatment that is appropriate for the more severely ill patients. The model setup in this paper is similar to Malcomson's model adapted to a dynamic framework with two physicians, in which each physician offers one of the treatment options.

3. Model

The payer contracts with two physicians, a PCP (P) and a specialist (S), to treat a fixed number of chronic patients for an indefinite number of periods. At the end of each period, there is a probability of $0 < \beta < 1$ that the game will continue. Furthermore, there exists an outside provider (in the following: hospital) (H) who treats all patients who are not treated by either physician. I assume that the disease is an ambulatory care sensitive condition. Because hospital treatment tends to be expensive, I assume that the

Variable	Definition
i_k	Patient state, consisting of type $i \in \{l, h\}$ and physician $k \in \{P, S\}$
H	Hospital state
p_k^{ij}	Base transition probability from type i to j when treated by physician k
c_k^i	Cost of treatment of patient type i by physician k
Δ_{c_k}	Difference in the costs between patient types for physician k
e_k^i	Effort level of physician k for patient i
Δ_p	Improvement of transition probability when exerting effort
c_e	Cost of effort
γ_k	Treatment fee for physician k
F_k	Fixed payment for physician k
β	Discount rate
L	Health cost of the patient when in severe condition
i_k^0	Initial value for state i_k
D_k^i	Decision of physician k to treat or refer a patient of type i
s_k	Strategy of physician k

Table 1: Definition of variables

payer always prefers ambulatory treatment to hospital treatment. To save on notation, I do not model the output of the hospital in detail.

In each period, a patient receives treatment from one physician only. There are two types of patients, patients in mild condition (l-types) and patients in severe condition (h-types). For every period that a patient remains in severe condition, he suffers a health loss $L \geq 0$. The payer cannot observe the patients' types.

The type of each patient changes probabilistically over time. The base transition probabilities between types i and j when receiving treatment from physician k are denoted by $p_k^{ij} \in (0, 1)$. It holds $p_k^{ij} + p_k^{ii} = 1, i, j \in \{l, h\}$. Physicians $k \in \{P, S\}$ can exert effort $e_k^i \in \{0, 1\}, i \in \{l, h\}$ in order to reduce the probability that a patient's condition deteriorates from condition l to h (tertiary prevention effort) or increase the probability that a patient's condition improves from condition h to l by Δ_p (treatment effort) in a given period:

$$Pr(l \to l | e_k^l = 1) = p_k^{ll} + \Delta_p, \qquad Pr(l \to h | e_k^l = 1) = p_k^{lh} - \Delta_p$$
$$Pr(h \to h | e_k^h = 1) = p_k^{hh} - \Delta_p, \qquad Pr(h \to l | e_k^h = 1) = p_k^{hl} + \Delta_p$$

The variables used in this paper are summarized in Table 1. In order to ensure that all probabilities stay within zero and one, I assume $\Delta_p < p_k^{ij} \; \forall i, j \in \{l, h\}, k \in \{P, S\}$.

Physician treatment comes at costs $c_k^i + c_e$, where $c_k^i \geq 0$ are the treatment costs for physician k to treat a patient in condition i and $c_e \geq 0$ is the cost of exerting effort[5]. For the sake of simplicity I assume that both the costs and benefits of providing effort are identical between the physicians. However, both the treatment costs and base transition probabilities differ between them.

I assume that for mildly ill patients PCP treatment is cheaper than specialist treatment ($c_P^l < c_S^l$). This is often the case due to the more intensive specialist treatment. For severely ill patients either provider can be cheaper. Furthermore, I assume that severely ill patients are more expensive to treat than mildly ill patients ($\Delta c_k := c_k^h - c_k^l > 0$) because patients in severe condition tend to require more intensive treatment and are at greater risk for complications[6].

Specialists are more effective at treating severely ill patients. This means that with equal physician effort a patient in severe condition will be more likely to improve his condition if he is treated by the specialist rather than the PCP ($p_S^{hh} < p_P^{hh}$). For mildly ill patients, either provider may be more effective. However, if the specialist has a treatment advantage for mildly ill patients, it is smaller than the treatment advantage for severely ill patients ($p_P^{hh} - p_S^{hh} > p_S^{ll} - p_P^{ll}$).

Both physicians can observe a patient's type when taking a decision. At the beginning of a period, the physician who treated a patient in the last period is responsible for treating the patient in this period. Physicians decide on whether to treat or refer patients who they are currently responsible for. If the physician decides to refer the patient, the referral only completes if the receiving physician accepts it. In this case, responsibility for the patient's treatment shifts towards the receiving physician. Otherwise, if the referral is rejected, the patient receives treatment from the hospital from then on forward. Figure 1 shows the possible states a patient can be in and the transition probabilities between the states. PCP behavior is at the top and specialist behavior is at the bottom.

Physicians receive a treatment fee $\gamma_k \geq 0$ in each period per patient whom they treat in addition to a fixed payment of F_k which they receive after signing the contract. Periodical treatment fees are common components of real world payment systems. The fixed payment represents subsidies that physicians may receive upon opening a practice.

Physicians are assumed to be profit maximizers and to not be capacity constrained. In order for them to accept their contracts, they need to earn at least a minimum expected

[5]A part of the physicians costs are time costs. The physician's valuation of his time may depend on his opportunity costs. For example, a physician who treats both privately and publicly insured patients may have higher opportunity costs than a physician who only treats publicly insured patients.

[6]This is a common assumption in health economic models (Grassi and Ma, 2016; Hafsteinsdottir and Siciliani, 2010; Malcomson, 2005; Eggleston, 2000).

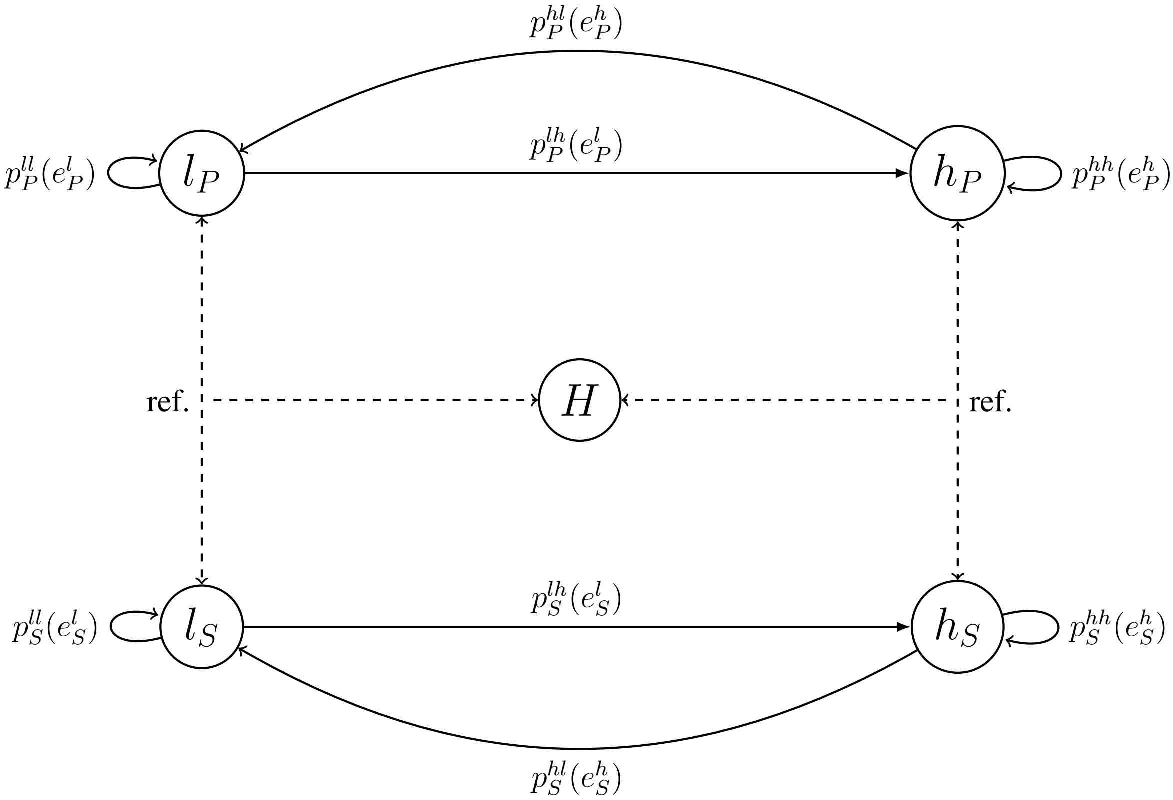


Figure 1: Patient states and transition probabilities

discounted profit with the common discount factor β[7]. For simplicity, I assume that any rent can be extracted by lowering the fixed payments. Patients are fully insured and follow their physicians' recommendations. Because the costs, benefits, and treatment fees per patient are independent of the number of treated patients, I analyze a single representative patient.

Figure 2 depicts the sequence of events. Before the treatments start, the payment system is created and physicians accept or reject their contracts and receive their fixed payments if they accepted the contract. If either physician rejects, the game ends, physicians receive no payoff, and patients are treated in the hospital. After accepting the contracts, the payer decides which provider is initially responsible for the treatment of the patient. Then the initial disease severity of the patient is drawn by nature.

The following describes the periodical patient treatment. First, the physicians observe the patient's type. The second step depends on the organizational form of the health

[7]Physicians discount the possibility that the game ends. For simplicity, I abstract from other reasons to discount the future.

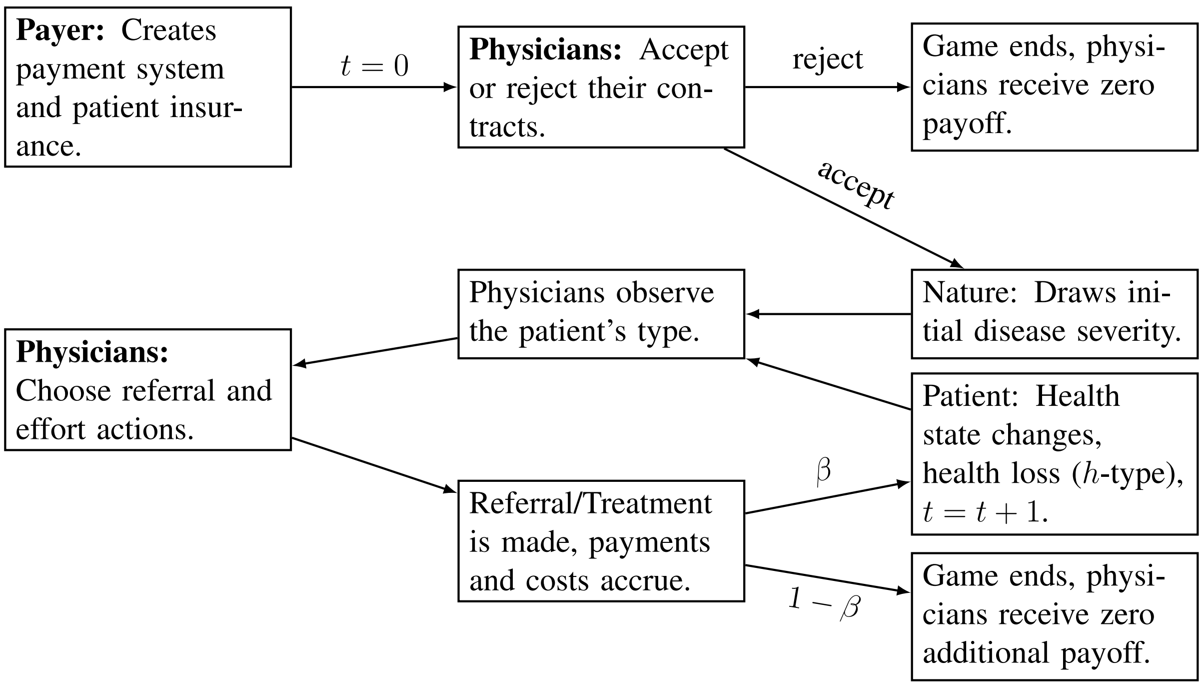


Figure 2: Sequence of Events

care market.

Solo Practices: If the physicians work in solo practices, the physicians simultaneously decide non-cooperatively on patient type-dependent referral and effort actions in this period. If a physician chooses to refer the patient, the referral will only be completed if the other physician accepts it. If the receiving physician rejects the referral, the patient is treated in the hospital.

Team: If the physicians form a team, they commit to a cooperative strategy in the second step instead. I assume that this strategy maximizes the discounted expected joint profit of the team. Further, physicians can freely transfer money between each other, provided that both physicians' participation constraints are fulfilled.

After referrals have been made, the patient is treated, costs accrue, and payments are made. With probability β the game continues, with probability $1 - \beta$ it ends. If the game ends, the physicians receive no additional payoff. If the game continues, the state of the patient changes and the patient suffers a health loss if he is in severe condition.

4. First-Best Benchmark

In this section, I describe the optimal outcome under the condition that the payer can implement any outcome without leaving profits to the physicians. Excluding the degenerate state H, there are four patient state variables $x \in \mathcal{X} := \{l_P, h_P, l_S, h_S\}$ in the game depicted in Figure 1, namely a patient in mild (l-types)/severe (h-types) condition who receives treatment in this period from the PCP/specialist respectively. Note that the patient *type* is the patient's disease severity, whereas the patient *state* also includes information about which physician is responsible for the treatment of the patient.

Both physicians know the state of the game whenever they take an action. I restrict the analysis of the game to Markov strategies, i.e. strategies that are conditioned only on the state of the game x and not on the history of the game in general[8]. Thus, each physician's strategy is fully described by their referral and effort decisions for each of the four states. The payer minimizes the discounted sum of expected health losses and physician payments (including fixed payments). Rents are optimally set to equal zero, thus the payments (γ, F) just cover the physicians' costs.

The game proposed in Section 3 fulfills the definition of a discounted stochastic game[9]. Let us consider first the *continuation welfare* $W(s, x)$. It describes the expected welfare generated by the physicians' strategies $s = (s_P, s_S)$ for a patient in a given state x:

$$W(s,x) = \underbrace{-\mathcal{L}(x) - \mathcal{C}(s,x)}_{\text{per period health/treatment/effort costs}} \underbrace{+\beta \sum_{\tilde{x} \in \mathcal{X}} Pr(x \to \tilde{x}|s) W(s,\tilde{x})}_{\text{discounted expected future welfare}}, x \in \mathcal{X}. \tag{1}$$

Here, a physician strategy $s_k, k \in \{P, S\}$ assigns an action from the space $\{\text{treat, refer}\} \times \{\text{effort, no effort}\}$ to every state $x \in \mathcal{X}$. Note that continuation welfare is independent of the period t. Thus, from the four equations corresponding to the four states, ex-ante expected welfare can be calculated as

$$\mathbb{E}W(s) = l_P^0 W(s, l_P) + l_S^0 W(s, l_S) + h_P^0 W(s, h_P) + h_S^0 W(s, h_S), \tag{2}$$

where $l_P^0, l_S^0, h_P^0, h_S^0 \in [0, 1)$ are the initial probabilities for the patient to be in a given state[10].

[8] Considering non-Markovian equilibrium strategies would give rise to Folk Theorem type strategies. This would create a multiplicity of equilibria. Because the main purpose of this paper is to compare the performance of solo practice vs. team care, obtaining unique behavioral predictions is essential. Therefore, I do not consider non-Markovian strategies.

[9] I follow the solution method for this type of game in Filar and Vrieze (1996).

[10] Let the probability for the patient to be in severe condition be given by $p^0 \in (0, 1)$. The payer decides which physician is responsible for the patient initially. Thus, $h_P^0 + h_S^0 = p^0$, $l_P^0 + l_S^0 = 1 - p^0$. I assume that the patient does not have prior information on his type, thus the payer cannot improve

Let the patient have a fixed disease severity in some period. The type of physician who treated the patient in the last period is irrelevant to the question of which physician should treat the patient in this period and which effort level he should receive. Therefore, the patient should receive the same treatment and effort for a given type. I denote strategies by

$$s_k = (D_k^l, D_k^h)^{e_k^l e_k^h} \text{ with } D_k^l, D_k^h \in \{T, R\}, \tag{3}$$

where D_k^i is the decision of physician k to treat or refer a patient of type i and where e_k^i is the effort decision of the physician. The joint treatment and referral decisions of the physicians define the *treatment paths* for all patients. Table 2 provides an overview over all possible treatment paths. A treatment path is defined by the physician(s) who treat(s) a patient of type l and h.

If the PCP chooses to treat a patient type i and the specialist chooses to refer ($D_P^i = T$, $D_S^i = R$), that patient type will be treated by the PCP. This is indicated by the letter P. Conversely, if the specialist chooses to treat a patient type and the PCP chooses to refer ($D_S^i = T$, $D_P^i = R$), the specialist treats this type. This is indicated by the letter S. If both physicians choose to treat a patient type ($D_P^i = D_S^i = T$), whichever physician is responsible for treating the patient continues to treat him. This is indicated by the letter M. If both physicians choose to refer a patient type ($D_P^i = D_S^i = R$), the patient receives hospital treatment. This is indicated by the letter H. Once a patient receives treatment from the hospital, he does not return to the physicians. I assume that hospital costs are large enough such that implementing a treatment path including H is never optimal from the payer's perspective. To summarize,

P: Treatment by the PCP.

S: Treatment by the specialist.

M: Treatment by the physician who treated the patient in the last period.

H: Treatment by the hospital.

Excluding the paths in which the patient receives hospital treatment, nine different treatment paths emerge.

PS: l-types get referred to the PCP, h-types get referred to the specialist,.

PP: All patients get treated by the PCP.

SS: All patients get treated by the specialist,.

by letting the patient choose his own initial physician. In all cases discussed in this paper, the payer (weakly) prefers the PCP to receive patients initially. All main results are robust as long as some patients visit the PCP initially.

		Specialist			
		(T,T)	(T,R)	(R,T)	(R,R)
PCP	(T,T)	MM	MP	PM	PP
	(T,R)	MS	MH	PS	PH
	(R,T)	SM	SP	HM	HP
	(R,R)	SS	SH	HS	HH

Table 2: All possible treatment paths.

PM: l-types get referred to the PCP, h-types stay at their resp. provider.

MS: h-types get referred to the specialist, l-types stay at their resp. provider.

MM: All patients stay at their resp. provider.

MP: h-types get referred to the PCP, l-types stay at their resp. provider.

SM: l-types get referred to the specialist, h-types stay at their resp. provider.

SP: h-types get referred to the PCP, l-types get referred to the specialist.

The strategies of the physicians determine the *outcome* of the game. Outcomes are defined by the treatment path that the patient takes and the physicians' efforts. They will be denoted by $\{\text{treatment path}\}^{e_P e_S}$ with $e_k = e_k^l = e_k^h$. In principle, physicians could provide effort for one patient type but not the other. However, as will be shown in Subsection 5.2, this is never a second-best equilibrium outcome. For proofs I use the following extended notation for outcomes whenever necessary: $\{\text{treatment path}\}^{e_P^l e_P^h e_S^l e_S^h}$. If a physician k does not treat a patient type i, her chosen effort level for this patient type does not impact expected welfare. In this case, effort is denoted by "e_k^i".

Contracting with the physicians is always in the payer's interest. I assume, consistently with chronic care recommendations (Gask, 2005), that specialist treatment is only socially efficient for patients in severe condition. Furthermore, the cost of effort provision is low enough, such that effort provision for both physicians is always first-best optimal. Therefore, high-effort PCP treatment for low types and high-effort specialist treatment for high types is first-best in all periods. This outcome is denoted by PS^{11}. For this outcome to emerge, the PCP needs to play the strategy $s_P = (T,R)^{1e_P^h}$ and the specialist needs to play the strategy $s_S = (R,T)^{e_S^l 1}$. Formally, the conditions for PS^{11} to be optimal are (for any initial probabilities x^0):

$$\mathbb{E}W(PS^{11}) \geq \mathbb{E}W(PS^{00}) \iff c_e \leq \frac{\Delta_p \beta(L + c_S^h - c_P^l)}{1 + \beta(1 - p_P^{ll} - p_S^{hh})} \tag{4}$$

$$\mathbb{E}W(PS^{11}) \geq \mathbb{E}W(PP^{1e_S}) \iff c_S^h - c_P^h \leq \frac{(p_P^{hh} - p_S^{hh})\beta(L + c_P^h - c_P^l)}{1 + \beta(1 - p_P^{ll} - p_P^{hh})} \tag{5}$$

$$\mathbb{E}W(PS^{11}) \geq \mathbb{E}W(SS^{e_P1}) \iff c_S^l - c_P^l \geq \frac{(p_S^{ll} - p_P^{ll})\beta(L + c_S^h - c_P^l)}{1 + \beta(1 - p_P^{ll} - p_S^{hh})} \tag{6}$$

Condition (4) implies that effort provision is optimal given that the PCP treats l-types and the specialist treats h-types. Intuitively, the costs of effort may not exceed its discounted benefits.

Condition (5) implies that it is first-best optimal to let the specialist treat h-types given that the PCP treats l-types. Here, the cost difference between the specialist and PCP treatment of h-types may not be larger than the additional benefits of specialist treatment.

Condition (6) implies that it is first-best optimal to let the PCP treat l-types given that the specialist treats h-types. Note that if $p_P^{ll} \geq p_S^{ll}$, Condition (6) is always fulfilled. Otherwise, Condition (6) delivers a lower bound on L, whereas Condition (5) delivers an upper bound on L. In this case Condition (7) needs to hold in order for feasible parameters to exist.

$$\frac{c_S^h - c_P^h}{p_P^{hh} - p_S^{hh}} \leq \frac{c_S^l - c_P^l}{p_S^{ll} - p_P^{ll}} \tag{7}$$

Intuitively, Condition (7) states that potential cost savings from PCP care relative to specialist care may not be too large for h-types relative to l-types.

An outcome is *superior* (denoted by $\geq$) to another outcome in terms of welfare if and only if it delivers at least the same continuation welfare for any state. It is *strictly superior* if it is superior and delivers strictly greater welfare for some state. For the second-best welfare comparison between organizational forms, I assume the strict superiority of PS^{11} over all other strategies. It follows Lemma 1 (proof in Appendix A.1).

Lemma 1. *If Conditions (4) to (6) hold (strictly), PS^{11} is (strictly) superior to all other outcomes.*

5. Second-Best Analysis

5.1. Physician Behaviour

From now on I assume that the payer is not able to observe the type of a patient; however, he is still able to observe which physician provides treatment. The *continuation profit* of physician k for state x, given both physicians strategies s, is

$$u_k^x(s) = \underbrace{\Gamma_k(s,x) - \mathcal{C}_k(s,x)}_{\text{per period profit}} + \underbrace{\beta \sum_{\tilde{x}\in\mathcal{X}} Pr(x \to \tilde{x}|s) u_k^{\tilde{x}}(s)}_{\text{discounted expected future profit}}, x \in \mathcal{X} \tag{8}$$

with $\Gamma_k(s,x)$ being the physicians' per period treatment fees. The discounted expected profit in period $t = 0$ amounts to

$$U_k(s) = F_k + l_P^0 u_k^{l_P}(s) + l_S^0 u_k^{l_S}(s) + h_P^0 u_k^{h_P}(s) + h_S^0 u_k^{h_S}(s). \tag{9}$$

In order to participate, a physician/the physician team needs to earn minimum profits. For simplicity, they are normalized to zero. Thus, an equilibrium strategy s^* must satisfy the *ex-ante participation constraints*:

$$U_k(s^*) \geq 0, k \in \{P,S\} \text{ (PC)} \tag{10}$$

In the team, participation depends on the sum of the physicians' profits because the team can always ensure that both physicians earn non-negative profits with the help of internal transfers.

$$U_P(s^*) + U_S(s^*) \geq 0 \text{ (PC - T)} \tag{11}$$

Condition (8) defines a system of linear equations with four equations that can be used to determine the four unknown continuation profits. They are made up of two parts; the per-period profit of the physician and her discounted expected profit over the remaining periods. In the solo practices, physicians non-cooperatively aim to maximize their discounted expected profit (9), whereas the physician team aims to maximize the sum over both physicians' profits.

In order to ensure that the hospital does not receive patients, solo practice physicians need to receive non-negative continuation profit for each patient type, whereas in the team it is sufficient that the team's continuation profit is non-negative.

$$u_k^x(s^*) \geq 0, x \in \mathcal{X}, k \in \{P,S\} \tag{12}$$

$$u_P^x(s^*) + u_S^x(s^*) \geq 0, x \in \mathcal{X} \tag{13}$$

The payer's goal is to determine the treatment fees and a gatekeeping rule (i.e. a rule determining which physician is initially responsible for treating the patient) that maximize expected welfare (2) subject to the constraint that physicians maximize their expected profits (9), given the participation constraints (10) respectively (11) and interim participation constraints (12) respectively (13). I derive the contract that fulfills the above condition with minimal treatment fees. The fixed payments F_k do not impact incentives. Therefore, F_k is set to exactly fulfill the ex-ante participation constraints:

$$F_k = -[l_P^0 u_k^{l_P}(s) + l_S^0 u_k^{l_S}(s) + h_P^0 u_k^{h_P}(s) + h_S^0 u_k^{h_S}(s)]$$

Consequently, no rents accrue to the physicians. For the solo practices I derive Markov-Perfect-Equilibria (MPE) in pure strategies, i.e. both physician's strategies need to be best responses to each other in each state:

$$u_P^x(s_P^*|s_S^*) \geq u_P^x(s_P|s_S^*) \forall s_P, \forall x \in \mathcal{X},$$
$$u_S^x(s_S^*|s_P^*) \geq u_S^x(s_S|s_P^*) \forall s_S, \forall x \in \mathcal{X}$$

A strategy is *superior* (denoted by $\geq$) to another strategy in terms of utility if and only if it delivers at least the same continuation utility for each state. It is *strictly superior* (denoted by $>$) if it is superior and it delivers strictly greater continuation utility for some state. For the team I derive outcomes that maximize the teams joint profit. For the comparison between organizational forms, I only consider strategies that can be implemented as a unique equilibrium, in order to ensure that they can be reliably implemented by the payer. In order to state propositions succinctly, I assume that if physicians (or the physician team) receive the same continuation utility from two outcomes in each state, they strictly prefer the outcome that delivers superior welfare.

As assumed, a patient of a given type has the same costs and transition probabilities if he receives the same treatment, regardless of which physician treated him last period. Thus, a physician can never improve by treating the patient differently based on the previously treating physician. Consequently, all equilibrium strategies assign the same treatment and effort decision to the same patient type (see Condition (3)).

If possible, the payer aims to implement both the first-best treatment path PS as well as effort provision for both physicians $e_P^l = e_S^h = 1$. However, both implementing effort provision and implementing the desired treatment path PS are not always possible. Thus, a second-best outcome needs to be implemented in this case. Lemma 2 is useful in the second-best analysis. It follows from Conditions (5) and (6), see Appendix A.5.

Lemma 2. *It holds that*

$$\begin{aligned} &\mathbb{E}W(PS^{e_P e_S}) \geq \mathbb{E}W(PM^{e_P e_S}) \geq \mathbb{E}W(PP^{e_P e_S}) \text{ and} \\ &\mathbb{E}W(PS^{e_P e_S}) \geq \mathbb{E}W(MS^{e_P e_S}) \geq \mathbb{E}W(SS^{e_P e_S}) \end{aligned} \tag{14}$$

for $e_P = e_S$. If Conditions (4) to (6) hold strictly, so does Condition (14).

Furthermore, paths MM, SP, MP and SM are weakly dominated by PP and SS for fixed e_P, e_S.

Intuitively, Lemma 2 states that given equal levels of effort, an outcome is superior to another if its treatment path deviates less from the first-best treatment path PS.

In Subsection 5.2 and 5.3, I derive the set of potentially second-best outcomes which can be implemented by the payer for the solo practices and the team, respectively. Furthermore, I derive conditions on the treatment fees that allow the payer to implement his desired outcome. In Subsection 5.4, I compare the second-best optimal outcomes from both organizational forms and derive conditions under which either one is optimal.

5.2. Solo Practices

In this subsection, I derive the set of potentially second-best outcomes, i.e., the set of outcomes that are not dominated in terms of welfare and which can be implemented by the payer. Let us consider first which treatment paths cannot be implemented in the solo practices. This is shown in Proposition 1.

Proposition 1. *Strategies $s_k = (R, T), k \in \{P, S\}$ are never a part of an MPE in the solo-practices. Consequently, the first-best outcome can not be implemented.*

Proposition 1 is true because l-types are cheaper to treat than h-types. Therefore, treating an l-type now is always more profitable than treating an h-type now or later. Hence, physicians never refer l-types when they are willing to treat h-types. The affected treatment paths have been double crossed out in Table 3. Not being able to incentivize $s_S = (R, T)$ is problematic because the first-best calls for the specialist to refer back l-types and to treat h-types.

In contrast to the first-best treatment path, the "blind" treatment paths in which only one physician treats all patients (PP and SS), can trivially be implemented by the payer by only paying one physician sufficient treatment fees for treating both patient types. If a physician treats all patients, she is willing to exert effort for both patient types if the costs of effort are lower than the discounted future cost savings from keeping the patient in the mild condition. For the PCP and specialist, the conditions are respectively,

$$c_e \leq \tilde{c}_e^{\,P} := \frac{\Delta_p \beta \Delta_{c_P}}{1 + \beta(1 - p_P^{ll} - p_P^{hh})},$$
$$c_e \leq \tilde{c}_e^{\,S} := \frac{\Delta_p \beta \Delta_{c_S}}{1 + \beta(1 - p_S^{ll} - p_S^{hh})}.$$

		Specialist			
		(T,T)	(T,R)	~~(R,T)~~	(R,R)
PCP	(T,T)	~~MM~~	~~MP~~	~~PM~~	PP
	(T,R)	MS	~~MH~~	~~$\mathbf{PS}$~~	~~PH~~
	~~(R,T)~~	~~SM~~	~~SP~~	~~HM~~	~~MP~~
	(R,R)	SS	~~SH~~	~~HS~~	~~HH~~

Table 3: All possible treatment paths (solo practices), bold: first-best, single crossed-out: not second-best optimal, double crossed-out: impossible to implement as a unique MPE.

Thus, if there is a large difference in the treatment costs between l- and h-types (i.e large Δ_{c_P} or Δ_{c_S}), physicians are more willing to exert effort. For both treatment paths, the exerted effort is independent from the treatment fees. It follows Proposition 2 (proof in Appendix A.2).

Proposition 2. *Treatment paths PP and SS can always be implemented by the payer. Effort for both patient types is exerted by the PCP (specialist) if and only if $c_e \leq \tilde{c_e}^P(\tilde{c_e}^S)$. These two treatment paths together weakly dominate paths MM and MP in terms of social welfare.*

As PP and SS can always be implemented, any treatment path that is dominated by them in terms of social welfare, is never second-best optimal. Lemma 2 has already shown that MM and MP are weakly dominated for fixed levels of effort. Whenever effort can be incentivized for MM and MP, it can also be incentivized for PP and SS. Thus, the latter paths weakly dominate the former. MM and MP have been crossed out once in Table 3. Furthermore, as assumed, hospital treatment is never efficient from the payer's perspective. All remaining paths containing hospital treatment have also been crossed out once.

Proposition 2 shows that despite the fact that physicians are profit oriented in this model, they may still exert non-contractible effort in order to reduce their future treatment costs. Thus, the dynamic model of physician treatment offers an alternative explanation to altruism for effort in a credence goods market. Note that in the blind treatment paths physicians fully internalize the cost savings from high effort provision, though they do not internalize the patient's health losses. Therefore, only under-provision of effort is a potential issue but not over-provision.

It may seem surprising that the threshold for providing effort is independent of patient severity. As Hennessy (2008) point out, prevention effort and curative effort can be both complements and substitutes. If the probability that a patient remains healthy increases, so does the future gain from increasing curative effort. The reason for this is that patients stay healthy for longer once they are cured. If the probability that a patient improves his condition increases, the future gain from increasing prevention effort decreases. This is so because patients return to the healthy state more quickly. In this model, both effects just cancel each other out because both effort costs and the effect of effort coincide between disease severities.

There remains only a single treatment path from Table 3 which has not been analyzed yet, namely MS. In this path, the PCP treats l-type patients until their condition deteriorates to h and then refers them to the specialist (as in the first-best). The specialist then continuously treats the patient (as in the blind specialist treatment). This treatment path differs from the first-best path PS insofar that patients who have been cured by the specialist are not referred back to the PCP. In order to implement this path, the PCP plays strategy $(T,R)^{e_P^l e_P^h}$ and the specialist plays strategy $(T,T)^{e_S^l e_S^h}$. Even though the specialist would prefer to treat l-types, the PCP does not refer them. Thus, the PCP acts as a gatekeeper for the patient. In order to maximally utilize this gatekeeping effect, all patients should initially visit the PCP.

The specialist can be easily incentivized to treat all referred patients by setting her treatment fee higher or equal to her expected costs. Her effort provision is independent of the payments. Effort will be provided for both patient types if and only if $c_e \leq \tilde{c_e}^S$ (see Proposition 2).

Let the specialist play $(T,T)^{e_S^l e_S^h}$. Incentivizing the PCP to play $(T,R)^{e_P^l e_P^h}$ requires that the treatment fee is larger than her costs for l-types but not too large so that her continuation profits for h-types are not positive.

With PCP effort:

$$U_P[(T,R)^{1e_P^h}] \geq U_P[(R,R)^{e_P^l e_P^h}] \iff \gamma_P \geq c_P^l + c_e \tag{15}$$

$$U_P[(T,R)^{1e_P^h}] \geq U_P[(T,T)^{11}] \iff \gamma_P \leq c_P^h + c_e - \frac{\beta\Delta_{c_P}(1-p_P^{hh}+\Delta_p)}{1+\beta(1-p_P^{ll}-p_P^{hh})} \tag{16}$$

Without PCP effort:

$$U_P[(T,R)^{0e_P^h}] \geq U_P[(R,R)^{e_P^l e_P^h}] \iff \gamma_P \geq c_P^l \tag{17}$$

$$U_P[(T,R)^{0e_P^h}] \geq U_P[(T,T)^{00}] \iff \gamma_P \leq c_P^h - \frac{\beta\Delta_{c_P}(1-p_P^{hh})}{1+\beta(1-p_P^{ll}-p_P^{hh})} \tag{18}$$

Both pairs of conditions can be fulfilled at the same time respectively by a contract that just covers the PCP's treatment (and in the first case effort) costs for the treatment of l-types.

In order to incentivize effort, the PCP needs to make profits on l-types. This ensures that she has an incentive to keep patients in the mild state.

$$U_P[(T,R)^{1e_P^h}] \geq U_P[(T,R)^{0e_P^h}] \iff \gamma_P \geq c_P^l + \frac{c_e(1-\beta p_P^{ll})}{\beta \Delta_p} \tag{19}$$

Condition (19) shows that in order to incentivize PCP effort, it is not sufficient to only pay her treatment and effort costs. Instead, a markup on these costs needs to be paid. Continuation profits for treating l-types with effort rise more strongly in the treatment fee γ_P than the continuation profits for treating l-types without effort. The reason for this is that effort increases the expected amount of periods that a patient will be treated by the PCP. However, it is not possible to increase γ_P indefinitely because this would incentivize the PCP to not refer h-types (see Condition (16)). The larger the cost differences between types are for the PCP (Δ_{c_P}), the more room exists for increasing γ_P and, thus, for incentivizing effort. This effect is only present due to the multi-period nature of the model. In a one-shot interaction, the PCP would not have to consider the impact of her effort decision on future costs, and thus, paying a markup on her costs would be ineffective for incentivizing effort. Proposition 3 (proof in Appendix A.3) describes when and how effort incentivization for MS is possible.

Proposition 3. *In the solo practices, the treatment path in which both physicians treat l-types and the specialist treats h-types (MS) can always be implemented as a unique MPE. If and only if $c_e \leq \tilde{c_e}^P$, the PCP can be incentivized to provide effort and if and only if $c_e \leq \tilde{c_e}^S$, the specialist provides effort.*

In order for the PCP to provide effort, she needs to be paid a markup on both her treatment and effort costs of l-types and the specialist needs to be paid at least her expected treatment and effort costs for the treatment of h-types:

$$\gamma_P^* = c_P^l + \frac{c_e(1-\beta p_P^{ll})}{\beta \Delta_p},$$
$$\gamma_S^* \geq c_S^h + c_e - \frac{\beta \Delta_{c_S}(1-p_S^{hh}+\Delta_p)}{1+\beta(1-p_S^{ll}-p_S^{hh})}.$$

If effort cannot be incentivized ($c_e > \tilde{c_e}^S, \tilde{c_e}^P$):

$$\gamma_P^* = c_P^l,$$
$$\gamma_S^* \geq c_S^h - \frac{\beta \Delta_{c_S}(1-p_S^{hh})}{1+\beta(1-p_S^{ll}-p_S^{hh})}.$$

Figure 3 summarizes the results from Propositions 2 and 3. It shows, depending on the difference in costs between types for the PCP Δ_{c_P} and the costs of effort c_e, the

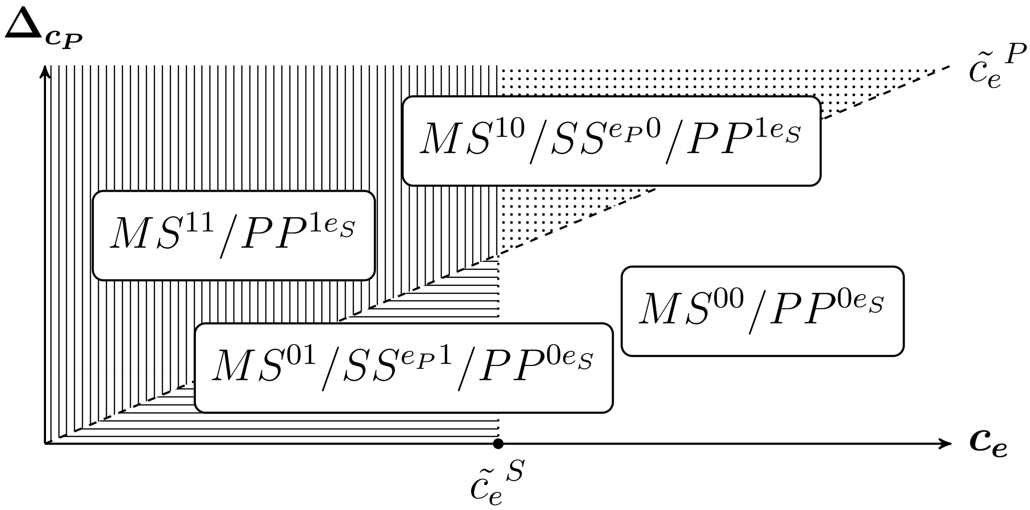


Figure 3: Potentially second-best outcomes in the solo practices.

set of potentially second-best outcomes in the four differently shaded areas. These sets consist of PP, SS, and MS with PCP effort provision for $c_e \leq \tilde{c_e}^P$ and specialist effort provision for $c_e \leq \tilde{c_e}^S$. Either of the three treatment paths can be second-best optimal with two exceptions. Given $c_e \leq \min(\tilde{c_e}^S, \tilde{c_e}^P)$ or $c_e \geq \max(\tilde{c_e}^S, \tilde{c_e}^P)$, blind specialist treatment is not second-best optimal as MS dominates it (see Lemma 2). In Section 5.4 and Appendix A.9, I take a closer look at the cases in which MS is the second-best optimal equilibrium. Corollary 1 summarizes the results for the solo practices.

Corollary 1. *The first-best treatment path cannot be implemented in the solo practices.*

In addition to blind PCP or specialist treatment PP and SS, there is another potentially second-best treatment path MS that can always be implemented. In this path the PCP acts as a gatekeeper for patients in mild condition and only refers severely ill patients to the specialist who continues treatment until the game ends.

5.3. Team Care

Physicians who work in a team coordinate their treatment and effort decisions via internal profit-sharing rules. I assume that there is no asymmetric information between the physicians and that they can freely transfer profits between them. Thus, they maximize their joint profit $U_T = U_P + U_S$ and split it in some manner[11]. For each patient of the same type, the continuation profit for the team (i.e., the sum of the continuation profits

[11] For example, as predicted by Nash Bargaining.

		Specialist			
		(T,T)	(T,R)	(R,T)	(R,R)
PCP	(T,T)	$\xcancel{MM}$	$\xcancel{MP}$	$\xcancel{PM}$	PP
	(T,R)	$\xcancel{MS}$	$\xcancel{MH}$	$\boldsymbol{PS}$	$\sout{PH}$
	(R,T)	$\xcancel{SM}$	SP	$\xcancel{HM}$	$\xcancel{MP}$
	(R,R)	SS	$\sout{SH}$	$\sout{HS}$	$\sout{HH}$

Table 4: All possible treatment paths (team), bold: first-best, single crossed-out: not second-best optimal, double crossed-out: impossible to implement as a unique equilibrium.

of both physicians) is identical. Therefore, the physician team never strictly prefers a mixed treatment path (i.e. a treatment path for which at least one patient type receives treatment from both physicians) to a treatment path in which only one physician treats a certain patient type. Furthermore, according to Condition (13), as long as the team makes a non-negative continuation profit from each type, the team will never refer a patient to the hospital. This can be ensured by setting treatment fees sufficiently high. Consequently, the possible treatment paths from Table 2 can be reduced to those in Table 4. These paths have in common that for each patient type there is exactly one physician who provides treatment.

This subsection proceeds as follows. First, I derive under which conditions PS^{11} can be implemented. Second, I derive second-best candidates given that PS^{11} cannot be implemented. Finally, I derive upper boundaries on L such that PS^{11} is first-best optimal and can be implemented in the team.

Let us consider first the first-best outcome PS^{11}. In order to implement it, after eliminating all inactive incentive constraints (see Appendix A.4), three constraints need to be fulfilled

$$PS^{11} \geq SS^{e_P 1} : (\gamma_S - c_S^h) - (\gamma_P - c_P^l) \leq -\frac{[1+\beta(1-p_P^{ll}-p_S^{hh})]\Delta_{c_S}}{1+\beta(1-p_S^{ll}-p_S^{hh})} \tag{20}$$

$$PS^{11} \geq PS^{00} : (\gamma_S - c_S^h) - (\gamma_P - c_P^l) \leq -\frac{c_e[1+\beta(1-p_P^{ll}-p_S^{hh})]}{\beta\Delta_p} \tag{21}$$

$$PS^{11} \geq PP^{1e_S} : (\gamma_S - c_S^h) - (\gamma_P - c_P^l) \geq -\frac{[1+\beta(1-p_P^{ll}-p_S^{hh})]\Delta_{c_P}}{1+\beta(1-p_P^{ll}-p_P^{hh})} \tag{22}$$

Condition (20) ensures that l-types are treated by the PCP. This is a lower bound on the team's profits when treating l-types. Condition (21) demonstrates that in order for the team to provide effort, it is necessary that the team makes greater profits when the PCP treats l-types rather than when the specialist treats h-types. Earning a markup for PCP treatment relative to specialist treatment increases the continuation profits of the team when treating l-types. Thus, they are incentivized to exert effort in order to increase the expected number of l-types in future periods. Which of these two condition is more strict depends on the effort costs c_e and the difference in treatment costs between types for the specialist (Δ_{c_s}). If $c_e \leq \tilde{c_e}^S$, Condition (20) is stricter than (21). Otherwise, the opposite is true.

However, markups for PCP treatment may not be too large relative to the specialist's markups. Otherwise, the PCP treats all patients. Condition (22) ensures that this does not happen. All three constraints cannot always be fulfilled together. Proposition 4 (proof in Appendix A.4) describes how, and under which conditions, the first-best outcome PS^{11} can be implemented in the team. Because Condition (22) is the only lower boundary on γ_S, the payer can simply fulfill it exactly in order to implement the outcome whenever this is possible.

Proposition 4. *The first-best treatment path (PS) can be implemented as a unique equilibrium in the team if and only if the difference in the expected treatment costs between mild and severe cases is larger for the PCP than the specialists:*

$$\Delta_{c_P} \geq \tilde{\Delta}_{c_P} := \frac{\Delta_{c_S}[1 + \beta(1 - p_P^{ll} - p_P^{hh})]}{1 + \beta(1 - p_S^{ll} - p_S^{hh})}. \tag{23}$$

Then, if and only if effort costs are low enough, i.e.

$$c_e \leq \tilde{c_e}^P, \tag{24}$$

the following contract implements PS^{11}:

$$\gamma_P^* = c_P^l + c_e + \frac{\beta\Delta_{c_P}(1 - p_P^{ll} - \Delta_p)}{1 + \beta(1 - p_P^{hh} - p_P^{ll})}$$
$$\gamma_S^* = c_S^h + c_e - \frac{\beta\Delta_{c_P}(1 - p_S^{hh} + \Delta_p)}{1 + \beta(1 - p_P^{hh} - p_P^{ll})}$$

According to Proposition 4, given that Conditions (23) and (24) hold, the first-best outcome PS^{11} can be implemented by paying the PCP a markup on her costs and paying below-cost fees to the specialist. This ensures that the team does not shift care towards the more expensive specialist and that the team exerts effort because they benefit from a patient in mild condition. Furthermore, the team's continuation profits for h-types are set to zero.

Condition (23) requires that the difference in expected costs when continuously treating an l-type patient rather than an h-type patient is larger for the PCP than the specialist. In this case, the PCP has the *(expected) relative cost-advantage* when treating l-type

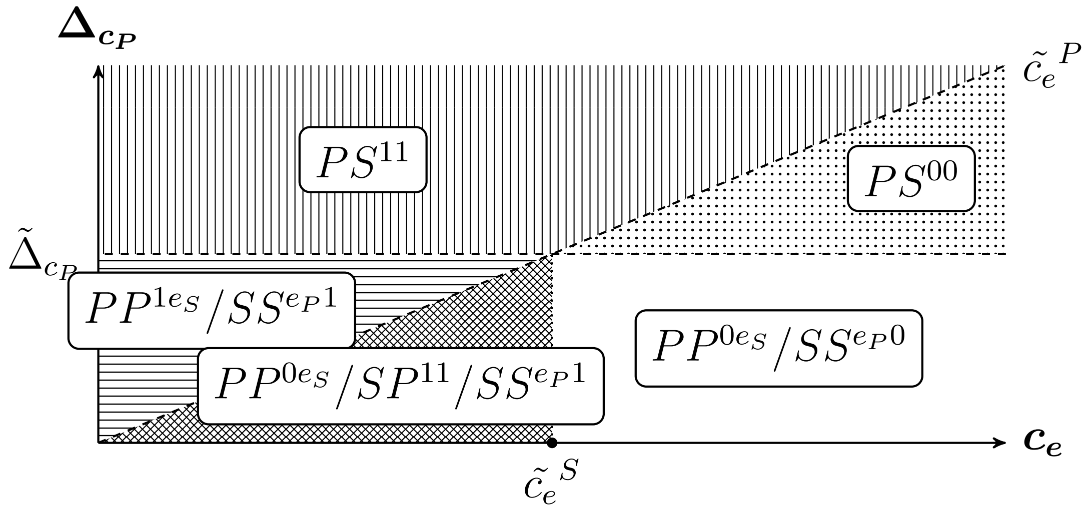


Figure 4: Second-best implementation in the team.

patients, and the specialist has the relative cost-advantage when treating h-type patients. If, conversely, the specialist has the relative cost advantage for treating l-types, the team is only willing to provide specialist care to h-types if the payment for specialist care is large. In this case, however, l-type patients are also treated by the specialist and, thus, the first-best outcome is not implemented. Condition (24) requires that effort costs are lower than the expected cost differences between the patient types for the PCP. This is necessary because effort is incentivized by a markup on PCP care. The larger the effort costs are, the larger is the markup required to incentivize effort. However, if the expected cost differences between types are small for the PCP, the team shifts care of h-types towards the PCP rather than exerting effort.

The second-best outcomes in the team are depicted in Figure 4. If the PCP has the relative cost-advantage for l-types ($\Delta_{c_P} > \tilde{\Delta}_{c_P}$), treatment path PS can always be implemented. For low enough effort costs, the first-best outcome PS^{11} is implemented. If PS^{11} cannot be implemented because effort costs are too large, PS^{00} can be implemented instead. If the specialist has the relative cost-advantage for l-types ($\Delta_{c_P} < \tilde{\Delta}_{c_P}$), either blind treatment or SP^{11} can be second-best. Details are discussed in Appendix A.6.

Consider the role of the discount factor β. If the future was discounted almost completely ($\beta \to 0$), Condition (24) would converge to $\Delta_{c_P} \geq \Delta_{c_S}$. Furthermore, there would be no effort provision. Because, per assumption, the specialist's treatment advantage is larger for h-types ($p_P^{hh} - p_S^{hh} \geq p_S^{ll} - p_P^{ll}$), increasing β increases the range of parameters for which the first-best can be implemented in the team. Intuitively, if the future is more important, it becomes more important for the team to provide effort and to match patient types with those physicians who can most effectively manage their disease.

So far I have considered whether PS^{11} can be implemented by the payer. However, the first-best cannot be implemented whenever this is desirable, i.e. whenever Conditions (4) to (6) are fulfilled. Here, a large health loss L is problematic. If $L = 0$, the cost-minimizing outcome is first-best optimal. A simple flat fee $\gamma_P = \gamma_S$ implements it in the team. However, with growing L the first-best benchmark is influenced more by concern for the patient's health and less by cost considerations. Thus, the strong focus on cost reduction in the team turns into a disadvantage. Differing markups for PCP and specialist can alleviate but not necessarily solve this problem by partly internalizing patient health losses. Proposition 7 in Appendix A.7 shows a sufficient condition under which PS^{11} can be implemented in the team for all feasible parameters. Corollary 2 summarizes the results for the team.

Corollary 2. *Providing chronic care in a team can incentivize appropriate referrals between physicians if the PCP has the relative cost-advantage for treating patients in mild condition.*

If effort costs are low compared to the difference in the PCP's expected cost of care between severely ill and mildly ill patients, effort can be incentivized as well. In this case, markups for the PCP's treatment of mildly ill patients and below-cost fees for the specialist incentivizes the team to play the first-best strategies.

If and only if the health losses from the severe conditions are sufficiently small, the first-best can always be implemented in the team.

5.4. Optimal Organization of Care

Comparing the second-best outcomes from solo practice care and team care yields the results depicted in Figure 5. If the PCP has the strict relative cost-advantage when treating l-types ($\Delta_{c_P} > \tilde{\Delta}_{c_P}$) the first-best treatment path can always be implemented in the team but not in the solo practices. Furthermore, the physicians always provide effort if they provide effort under solo practice care. Thus, team care strictly dominates solo practice care.

Conversely, if the specialist has the strict relative cost-advantage when treating l-types ($\Delta_{c_P} < \tilde{\Delta}_{c_P}$), there is an adverse incentive for the specialist to treat the mildly ill patients instead of, or in addition to, the patients in severe condition. In the solo practices this incentive is isolated on the specialist, whereas in the team this incentive also affects the PCP. Thus, only the blind treatment paths PP and SS or, if $\tilde{c}_e{}^P < c_e < \tilde{c}_e{}^S$,

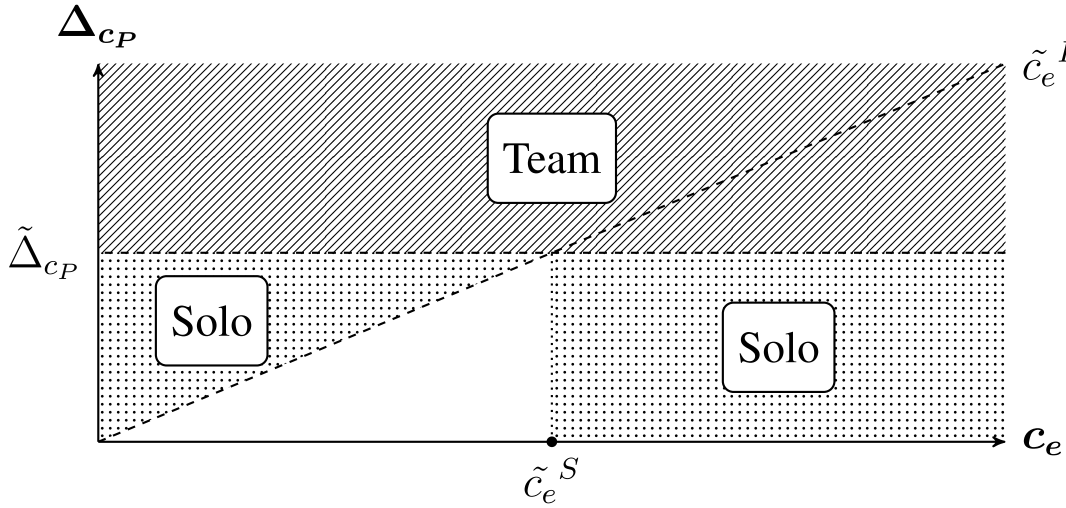


Figure 5: Second-best optimal organization of care (dots: weak dominance for solo practices, lines: strict dominance for the team, no pattern: no general dominance relation)

the reversed first-best path SP will be implemented in the team. Not being able to implement PS^{11} is problematic if this outcome is first-best optimal given $\Delta_{c_P} < \tilde{\Delta}_{c_P}$, which is the case only if the health loss in the severe condition L is sufficiently large (see Appendix A.7).

In the solo practices, the payer can potentially improve on the team's outcomes. In addition to implementing the blind outcomes it is possible to incentivize the PCP to treat only patients in mild condition and to incentivize the specialist to treat all referred patients (treatment path MS). In the team, it is not possible to implement this path. Therefore, if $c_e < \tilde{c}_e{}^P$ or $c_e > \tilde{c}_e{}^S$, the set of implementable outcomes is larger in the solo practices than in the team. Thus, solo practice care weakly dominates team care in this case. If we add the reasonable assumption that SS^{e_P1} weakly dominates SP^{11} in terms of welfare, solo practice care weakly dominates team care for $\tilde{c}_e{}^P < c_e < \tilde{c}_e{}^S$ also.

Strict superiority of solo practice care can also be demonstrated for some parameters. Let $\Delta_{c_P} < \tilde{\Delta}_{c_P}$. If $c_e < \tilde{c}_e{}^P$, either MS^{11} or PP^{1e_S} are second-best optimal. If $c_e > \tilde{c}_e{}^S$, either MS^{00} or PP^{0e_S} are second-best optimal. Whenever MS is strictly optimal, the solo practices are strictly superior to the team. As shown in Appendix A.9, MS is strictly superior to PP whenever the PCP's treatment costs for treating h-types are sufficiently large. Thus, cooperation between physicians is not desirable for the payer in this case. If $\tilde{c}_e{}^P < c_e < \tilde{c}_e{}^S$, no dominance relationship between organizational forms can be established for all parameters because outcome SP^{11} cannot be implemented in the solo practices. However, as shown via simulation in Appendix A.9, there also exist parameters in this case such that solo practice care is strictly superior to team care.

To illustrate the case in which solo practice care strictly dominates team care, consider the following example. Let specialist care of high-severity patients be both necessary (L very large, $p_S^{hh} << p_P^{hh}$) and expensive (c_S^h large). Note that in this case, it is likely that the specialist has the relative cost-advantage for l-types. If so, and if additionally the PCP only has a slight advantage when treating l-types ($c_P^l \approx c_S^l, p_S^{ll} \approx p_P^{ll}$), MS will be strictly superior to PP for levels of c_P^h that are sufficiently large but not large enough to shift the relative cost-advantage for l-types to the PCP. Figures 6 and 7 in Appendix A.9 showcase this example by means of simulation. Proposition 5 summarizes the main results of this paper.

Proposition 5. *Team care is strictly superior to solo practice care if the PCP has the relative cost advantage for treating patients in mild condition, i.e.* $\Delta_{c_P} > \tilde{\Delta}_{c_P}$. *Otherwise, if* $\Delta_{c_P} < \tilde{\Delta}_{c_P}$ *and* $c_e < \tilde{c}_e^{\,P}$ *or* $c_e > \tilde{c}_e^{\,S}$, *solo practice care is weakly superior to team care. In this case, if the PCP's costs for treating* h*-types is sufficiently large, solo practice care is strictly superior to team care. A necessary condition for solo practice superiority is that health losses in severe condition are sufficiently large.*

In Appendix B.1, I consider the additional problems that emerge when fees have to be flat. Appendix B.2 considers physician altruism. Under flat fees, team care is weakly dominant whenever the specialist has the relative cost advantage for low-types. Physician altruism increases the region in which team care is weakly dominant.

6. Conclusion

This paper studies the referral and effort provision of two physicians, a PCP and a specialist, who are responsible for treatment of chronically ill patients. I compare two organizational forms, solo practices and a physician team. In both organizational forms non-contractible effort can be incentivized for profit-maximizing physicians if effort costs are low enough. Effort can be incentivized because in this paper's dynamic setting effort provision today lowers physicians' future costs. If physicians are profit-maximizers, an optimal flow of patients between physicians (patients in mild conditions receive treatment from the PCP and patients in severe condition receive treatment from the specialist) can only be achieved in the team. However, it is necessary that the expected treatment cost differences between the patient types are relatively large for the PCP and small for the specialist, i.e. the PCP must have the relative cost-advantage when treating patients in mild condition. In this case, the first-best outcome is implemented by above-cost treatment fees for PCP treatment and below-cost fees for specialist treatment.

If the specialist has the relative cost advantage, however, social welfare in the solo practice care is weakly superior to team care under reasonable assumptions. Organizing physicians in solo practices allows the PCP to act as a gatekeeper, who treats all patients in mild condition until they are severely ill. If the PCP's treatment costs for severely ill patients are sufficiently large, this gatekeeping outcome is strictly superior to all outcomes in the team.

Two extensions of the model are analyzed in the appendix. If treatment fees in the team are flat, solo practice care is weakly superior to team care if the first-best treatment path is not cost-minimizing. Furthermore, if the first-best treatment path is cost-minimizing, the team does not dominate the solo practices for medium levels of effort because effort incentives are stronger in the solo practices. Altruistic physicians were considered as a second extension. If physicians are partially altruistic, team care weakly dominates solo practice care for a greater parameter region. If they are perfectly altruistic, the first-best can always be implemented in the team by a flat fee.

The main policy implication is that either team care or care in solo practices can be optimal. Team care is more likely to be superior to solo practice care if

- there is a greater variance in cost between patient types for the PCP than the specialist.
- physicians are altruistic.
- physicians discount the future only weakly.
- PCPs provide strongly superior treatment for mildly ill patients and specialists for severely ill patients.
- the health costs of being in the severe state are small.

It is difficult to make specific recommendations on the basis of the model presented in this paper as to which types of specialists should or should not form a team. The main issue is that little is currently known about the effect of the interaction between specialty and patient type on treatment costs. Future research in this area is warranted.

Another policy implication is that, regardless of organizational form, markups on PCP treatment can enhance effort provision. If, conversely, a team specialist were to generate most of the team's profits, this could have an adverse effect on effort provision because the team profits more from sick rather than healthy patients.

An important agency problem in the solo practices identified in this paper is a lack of back-referral of patients, who have been successfully treated by the specialist, to the PCP. This is a highly relevant practical problem. Take, for example, patients suffering from cancer. Cheung et al. (2009) find that there is considerable disagreement between PCPs and oncologists as to who should provide aftercare as both consider themselves to be the appropriate main provider for aftercare. However, current evidence suggests that PCPs are able to provide aftercare to the same quality standard as oncologist, albeit at lower costs (Meiklejohn et al., 2016). Consequently, not back-referring the patient to PCP care may create unnecessary costs. In the Singaporean health care system, the concept of "right siting" describes the idea that stable chronically ill patients should be treated by a PCP rather than a specialist in order to free up scarce specialist resources. In the case of care for diabetic patients, for example, evidence from Singapore suggests that patients receive similar treatment quality from the PCP and the endocrinologist (Wee et al., 2008; Ho et al., 2013), while PCP care could save on costs (Lim et al., 2008). An empirical investigation into the extend to which specialists hold on to patients that could be treated more efficiently by their PCPs would be a fruitful subject for future research.

The main innovation of this paper is that it considers agency problems regarding referrals and effort provision in a dynamic framework. Because this paper focuses on the efficiency of different organizational forms of care, a simplified payment system is used. Future work could allow for more complexity. For example, the length of stay of the patient at a physician provides information on the state of the patient. Thus, the payer could use it as a basis for the size of treatment fees or mandate a back-referral to PCP care after a certain time frame. Alternatively, the payer could consider a budgeted system. These modifications could help to improve the second-best outcomes in the solo practices. Furthermore, capacity constraints of physicians, diagnostic uncertainty, and informational asymmetries between the physicians regarding the type of the patient could be studied. In this paper, only Markovian equilibrium strategies are considered for solo practices to generate unique predictions for comparison with the team case. Another potential avenue for future research is to consider non-Markovian strategies.

Appendix

A general note: When calculating expected welfare or profit between different strategies s, it is sufficient to only consider states whose continuation welfare/profit is different between strategies. For the sake of brevity I may omit states for which there is no such difference.

A. Mathematical Appendix

A.1. Proof for Lemma 1 (First-best conditions)

Proof. I consider all possible outcomes of the game and show that, given Conditions (4) to (6), PS^{11} yields greater or equal expected welfare for any initial distribution of patients.

Optimal treatment effort:

First, I consider the condition under which providing effort is optimal given that PCPs treat l-types and specialists treat h-types. The system of equations for the continuation welfare of $PS^{e_P e_S}$ in all states $x \in \mathcal{X}$ are given by:

$$\begin{aligned}
&W^{l_P}_{PS^{e_P e_S}} = -c^l_P - e_P c_e + \beta[(p^{ll}_P + e_P \Delta_p) W^{l_P}_{PS^{e_P e_S}} + (1 - p^{ll}_P - e_P \Delta_p) W^{h_S}_{PS^{e_P e_S}}] \\
&W^{l_S}_{PS^{e_P e_S}} = W^{l_P}_{PS^{e_P e_S}} \\
&W^{h_P}_{PS^{e_P e_S}} = W^{h_S}_{PS^{e_P e_S}} \\
&W^{h_S}_{PS^{e_P e_S}} = -L - c^h_S - e_S c_e + \beta[(p^{hh}_S - e_S \Delta_p) W^{h_S}_{PS^{e_P e_S}} + (1 - p^{hh}_S + e_S \Delta_p) W^{l_P}_{PS^{e_P e_S}}]
\end{aligned}$$

Solving this system leads to the following continuation welfare:

$$W^{l_P}_{PS^{11}} = W^{l_S}_{PS^{11}} = \frac{\frac{-\beta(1-p^{ll}_P-\Delta_p)(L+c^h_S-c^l_P)}{1+\beta(1-p^{ll}_P-p^{hh}_S)} - c^l_P - c_e}{1-\beta}$$

$$W^{h_P}_{PS^{11}} = W^{h_S}_{PS^{11}} = \frac{\frac{-(1-\beta p^{ll}_P-\beta\Delta_p)(L+c^h_S-c^l_P)}{1+\beta(1-p^{ll}_P-p^{hh}_S)} - c^l_P - c_e}{1-\beta}$$

The systems for the other combinations of effort levels can be solved analogously. It follows

$$\begin{aligned}
&\mathbb{E}W(PS^{11}) \geq \mathbb{E}W(PS^{10}) \iff W^{l_P}_{PS^{11}} \geq W^{l_P}_{PS^{10}} \iff W^{h_S}_{PS^{11}} \geq W^{h_S}_{PS^{10}} \iff \\
&\mathbb{E}W(PS^{11}) \geq \mathbb{E}W(PS^{01}) \iff W^{l_P}_{PS^{11}} \geq W^{l_P}_{PS^{01}} \iff W^{h_S}_{PS^{11}} \geq W^{h_S}_{PS^{01}} \iff \\
&\mathbb{E}W(PS^{11}) \geq \mathbb{E}W(PS^{00}) \iff W^{l_P}_{PS^{11}} \geq W^{l_P}_{PS^{00}} \iff W^{h_S}_{PS^{11}} \geq W^{h_S}_{PS^{00}} \iff \\
&\qquad c_e \leq \frac{\Delta_p \beta(L + c^h_S - c^l_P)}{1 + \beta(1 - p^{ll}_P - p^{hh}_S)}.
\end{aligned}$$

Clearly, varying a physician's effort level for a patient type that she does not treat has no impact on expected welfare. These cases are omitted here.

For the remaining outcomes the procedure is analogous. I will present them in an abridged manner.

PS vs. PP:

If effort provision is efficient for one patient type, it is efficient for the other type as well:

$$
\begin{aligned}
&\mathbb{E}W(PP^{11e_S^l e_S^h}) \geq \mathbb{E}W(PP^{10e_S^l e_S^h}) \iff \mathbb{E}W(PP^{11e_S^l e_S^h}) \geq \mathbb{E}W(PP^{01e_S^l e_S^h}) \iff \\
&\mathbb{E}W(PP^{11e_S^l e_S^h}) \geq \mathbb{E}W(PP^{00e_S^l e_S^h}) \iff c_e \leq \frac{\Delta_p \beta(L + c_P^h - c_P^l)}{1 + \beta(1 - p_P^{ll} - p_P^{hh})} \\
&\mathbb{E}W(PP^{00e_S^l e_S^h}) \geq \mathbb{E}W(PP^{10e_S^l e_S^h}) \iff \mathbb{E}W(PP^{00e_L^l e_S^h}) \geq \mathbb{E}W(PP^{01e_S^l e_S^h}) \iff \\
&\mathbb{E}W(PP^{00e_S^l e_S^h}) \geq \mathbb{E}W(PP^{11e_S^l e_S^h}) \iff c_e \geq \frac{\Delta_p \beta(L + c_P^h - c_P^l)}{1 + \beta(1 - p_P^{ll} - p_P^{hh})}
\end{aligned}
$$

Therefore, only PP^{1e_S} and PP^{0e_S} need consideration.

$$
\begin{aligned}
\mathbb{E}W(PS^{11}) \geq \mathbb{E}W(PP^{1e_S}) &\iff \mathbb{E}W(PS^{00}) \geq \mathbb{E}W(PP^{0e_S}) \\
&\iff c_S^h - c_P^h \leq \frac{(p_P^{hh} - p_S^{hh})\beta(L + c_P^h - c_P^l)}{1 + \beta(1 - p_P^{ll} - p_P^{hh})}
\end{aligned}
$$

PS vs. SS:

If effort provision is efficient for one patient type, it is efficient for the other type as well:

$$
\begin{aligned}
&\mathbb{E}W(SS^{e_P^l e_P^h 11}) \geq \mathbb{E}W(SS^{e_P^l e_P^h 10}) \iff \mathbb{E}W(SS^{e_P^l e_P^h 11}) \geq \mathbb{E}W(SS^{e_P^l e_P^h 01}) \iff \\
&\mathbb{E}W(SS^{e_P^l e_P^h 11}) \geq \mathbb{E}W(SS^{e_P^l e_P^h 00}) \iff c_e \leq \frac{\Delta_p \beta(L + c_S^h - c_S^l)}{1 + \beta(1 - p_S^{ll} - p_S^{hh})} \\
&\mathbb{E}W(SS^{e_P^l e_P^h 00}) \geq \mathbb{E}W(SS^{e_P^l e_P^h 10}) \iff \mathbb{E}W(SS^{e_P^l e_P^h 00}) \geq \mathbb{E}W(SS^{e_P^l e_P^h 01}) \iff \\
&\mathbb{E}W(SS^{e_P^l e_P^h 00}) \geq \mathbb{E}W(SS^{e_P^l e_P^h 11}) \iff c_e \geq \frac{\Delta_p \beta(L + c_S^h - c_S^l)}{1 + \beta(1 - p_S^{ll} - p_S^{hh})}
\end{aligned}
$$

Therefore, only $SS^{e_P 1}$ and $SS^{e_P 0}$ need consideration.

$$
\begin{aligned}
\mathbb{E}W(PS^{11}) \geq \mathbb{E}W(SS^{e_P 1}) &\iff \mathbb{E}W(PS^{00}) \geq \mathbb{E}W(SS^{e_P 0}) \\
&\iff c_S^l - c_P^l \geq \frac{(p_S^{ll} - p_P^{ll})\beta(L + c_S^h - c_P^l)}{1 + \beta(1 - p_P^{ll} - p_S^{hh})}
\end{aligned}
$$

PS vs. PM: First note that if PCP effort provision in PM is efficient for one of the types, it must be efficient for the other as well. This is so because the continuation welfare of PM is identical to that of PP for any patient who is treated by the PCP.

Therefore, it is sufficient to consider only cases in which physicians either provide effort to all patient types they receive or none.

$$\begin{aligned}\mathbb{E}W(PS^{11}) \geq \mathbb{E}W(PM^{11}) &\iff \mathbb{E}W(PS^{00}) \geq \mathbb{E}W(PM^{00})\\ &\iff c_S^h - c_P^h \leq \frac{(p_P^{hh} - p_S^{hh})\beta(L + c_P^h - c_P^l)}{1 + \beta(1 - p_P^{ll} - p_P^{hh})}\end{aligned}$$

If $\mathbb{E}W(PM^{11}) \geq \mathbb{E}W(PM^{10})$, $\mathbb{E}W(PS^{11}) \geq \mathbb{E}W(PM^{10})$ follows immediately. Let instead $\mathbb{E}W(PM^{10}) \geq \mathbb{E}W(PM^{11})$. This implies a lower boundary $c_e \geq \hat{c}_e$ with

$$\hat{c}_e := \frac{\Delta_p\beta\{L - c_P^l + c_S^h - \beta(1 - p_P^{ll} - \Delta_p)(L + c_P^h - c_P^l)(1 + \beta[1 - p_p^{ll} - p_P^{hh}])\}}{1 - \beta(p_S^{hh} - \Delta_p)}.$$

Further,

$$\begin{aligned}&\mathbb{E}W(PS^{10}) - \mathbb{E}W(PM^{10}) =\\ &\frac{\beta^2(1 - p_S^{hh})(1 - p_P^{ll} - \Delta_p)(L + c_P^h - c_P^l)(1 + \beta[1 - p_P^{ll} - p_P^{hh}])}{(1 - \beta)(1 - \beta p_S^{hh})} - \\ &\frac{\beta^2(1 - p_S^{hh})(1 - p_P^{ll} - \Delta_p)(L - c_P^l + c_S^h - c_e)(1 + \beta[1 - p_P^{ll} - p_S^{hh} - \Delta_p])}{(1 - \beta)(1 - \beta p_S^{hh})}\\ &\implies \frac{\partial(\mathbb{E}W(PS^{10}) - \mathbb{E}W(PM^{10}))}{\partial c_e} > 0.\end{aligned}$$

Thus, if PS^{10} is superior to PM^{10} for some level of c_e, it is also superior for any larger level. Inserting $\hat{c}_e$ delivers

$$\mathbb{E}W(PS^{11}) \geq \mathbb{E}W(PM^{11}) \implies \mathbb{E}W(PS^{10}) \geq \mathbb{E}W(PM^{10}).$$

An analogue argument shows that

$$\mathbb{E}W(PS^{11}) \geq \mathbb{E}W(PM^{11}) \implies \mathbb{E}W(PS^{01}) \geq \mathbb{E}W(PM^{01}).$$

PS vs. MS: An analogue argument applies as in the previous case.

PS vs. MM: MM is weakly dominated by the set of all other outcomes due to the linearity of costs and benefits.

PS vs. SP:

First note that SP^{11} and SP^{00} are superior to SP^{10} and SP^{01}:

$$\begin{aligned}\mathbb{E}W(SP^{11}) \geq \mathbb{E}W(SP^{10}) &\iff \mathbb{E}W(SP^{11}) \geq \mathbb{E}W(SP^{01})\\ \iff \mathbb{E}W(SP^{11}) \geq \mathbb{E}W(SP^{00}) &\iff c_e \leq \frac{\Delta_p\beta(L + c_P^h - c_S^l)}{1 + \beta(1 - p_S^{ll} - p_P^{hh})}\\ \mathbb{E}W(SP^{00}) \geq \mathbb{E}W(SP^{10}) &\iff \mathbb{E}W(SP^{00}) \geq \mathbb{E}W(SP^{01})\\ \iff \mathbb{E}W(SP^{00}) \geq \mathbb{E}W(SP^{11}) &\iff c_e \geq \frac{\Delta_p\beta(L + c_P^h - c_S^l)}{1 + \beta(1 - p_S^{ll} - p_P^{hh})}\end{aligned}$$

Thus, it is sufficient to show the superiority of PS^{11} over SP^{11} and SP^{00}. $SP^{e_P e_S}$ is dominated by the blind outcomes:

$$\mathbb{E}W(PP^{1e_S}) \geq \mathbb{E}W(SP^{11}) \iff \mathbb{E}W(PP^{0e_S}) \geq \mathbb{E}W(SP^{00}) \iff$$
$$\frac{L + c_P^h - c_P^l}{1 + \beta(1 - p_P^{ll} - p_P^{hh})} \geq \frac{L + c_P^h - c_S^l}{1 + \beta(1 - p_S^{ll} - p_P^{hh})}$$

If $p_P^{ll} \geq p_S^{ll}$, this condition is always fulfilled. Otherwise, it bounds L from above. Further,

$$\mathbb{E}W(SS^{e_P 1}) \geq \mathbb{E}W(SP^{11}) \iff \mathbb{E}W(SS^{e_P 0}) \geq \mathbb{E}W(SP^{00}) \iff$$
$$L \geq c_S^l - c_P^h - \frac{(c_P^h - c_S^h)[1 + \beta(1 - p_S^{ll} - p_P^{hh})]}{\beta(p_P^{hh} - p_S^{hh})}.$$

One of the conditions is fulfilled if

$$\frac{c_S^h - c_P^h}{p_P^{hh} - p_S^{hh}} \leq \frac{c_S^l - c_P^l}{p_S^{ll} - p_P^{ll}}.$$

This is just Condition (7). As PS^{11} is superior to both $PP^{e_P e_S}$ and $SS^{e_P e_S}$, it follows $PS^{11} \geq SP^{e_P e_S}$.

PS vs. SM:

First note that if specialist effort provision in SM is efficient for one of the types, it must be efficient for the other as well. This is so because the continuation welfare of SM is identical to that of SS for any patient who is treated by the specialist. Therefore, it is sufficient to consider only cases in which physicians either provide effort to all patient types they receive or none.

$SM^{e_P e_S}$ is dominated:

$$\begin{aligned}
\mathbb{E}W(SS^{e_P 1}) \geq \mathbb{E}W(SM^{11}) &\iff c_S^h - c_P^h \leq \frac{(p_P^{hh} - p_S^{hh})\beta(L + c_P^h - c_S^l)}{1 + \beta(1 - p_S^{ll} - p_P^{hh})} \\
\mathbb{E}W(SP^{11}) \geq \mathbb{E}W(SM^{11}) &\iff c_S^h - c_P^h \geq \frac{(p_P^{hh} - p_S^{hh})\beta(L + c_P^h - c_S^l)}{1 + \beta(1 - p_S^{ll} - p_P^{hh})} \\
\mathbb{E}W(SS^{e_P 0}) \geq \mathbb{E}W(SM^{10}) &\iff c_S^h - c_P^h - c_e \leq \frac{(p_P^{hh} - p_S^{hh} - \Delta_p)\beta(L + c_S^h - c_S^l)}{1 + \beta(1 - p_S^{ll} - p_S^{hh})} \\
\mathbb{E}W(SP^{10}) \geq \mathbb{E}W(SM^{10}) &\iff c_S^h - c_P^h - c_e \geq \frac{(p_P^{hh} - p_S^{hh} - \Delta_p)\beta(L + c_S^h - c_S^l)}{1 + \beta(1 - p_S^{ll} - p_S^{hh})} \\
\mathbb{E}W(SS^{e_P 1}) \geq \mathbb{E}W(SM^{01}) &\iff c_S^h - c_P^h + c_e \leq \frac{(p_P^{hh} - p_S^{hh} + \Delta_p)\beta(L + c_S^h - c_S^l)}{1 + \beta(1 - p_S^{ll} - p_S^{hh})} \\
\mathbb{E}W(SP^{01}) \geq \mathbb{E}W(SM^{01}) &\iff c_S^h - c_P^h + c_e \geq \frac{(p_P^{hh} - p_S^{hh} + \Delta_p)\beta(L + c_S^h - c_S^l)}{1 + \beta(1 - p_S^{ll} - p_S^{hh})} \\
\mathbb{E}W(SS^{e_P 0}) \geq \mathbb{E}W(SM^{00}) &\iff c_S^h - c_P^h \leq \frac{(p_P^{hh} - p_S^{hh})\beta(L + c_S^h - c_S^l)}{1 + \beta(1 - p_S^{ll} - p_S^{hh})} \\
\mathbb{E}W(SP^{00}) \geq \mathbb{E}W(SM^{00}) &\iff c_S^h - c_P^h \geq \frac{(p_P^{hh} - p_S^{hh})\beta(L + c_S^h - c_S^l)}{1 + \beta(1 - p_S^{ll} - p_S^{hh})}
\end{aligned}$$

As PS^{11} is superior to $PP^{e_P e_S}$ and $SS^{e_P e_S}$, and $SP^{e_P e_S}$, it follows $PS^{11} \geq SM^{e_P e_S}$.

PS vs. MP: An analogue argument applies as in the previous case.

The proof for the strict superiority of PS^{11} follows by replacing inequalities with strict inequalities. □

A.2. Proof for Proposition 2 (implementation of blind treatment paths)

Proof. The system of equations for the continuation profit for the PCP of $PP^{e_P^l e_P^h e_S^l e_S^h}$ in all states $x \in \mathcal{X}$ are given by:

$$\begin{aligned}
u_P^{l_P} &= -c_P^l - e_P^l c_e + \beta[(p_P^{ll} + e_P^l \Delta_p)u_P^{l_P} + (1 - p_P^{ll} - e_P^l \Delta_p)u_P^{h_P}] \\
u_P^{l_S} &= u_P^{l_P} \\
u_P^{h_P} &= -c_P^h - e_P^h c_e + \beta[(p_P^{hh} - e_P^h \Delta_p)u_P^{h_P} + (1 - p_P^{hh} + e_P^h \Delta_p)u_P^{l_P}] \\
u_P^{h_S} &= u_P^{h_P}
\end{aligned}$$

Solving this system leads to continuation profits:

$$u_P^{l_P}(PP^{11e_S^l e_S^h}) = u_P^{l_S} = \frac{\gamma_P - c_e - c_P^l}{1-\beta} - \frac{\beta(c_P^h - c_P^l)(1 - p_P^{ll} - \Delta_p)}{(1-\beta)[1+\beta(1-p_P^{ll}-p_P^{hh})]}$$
$$u_P^{h_P}(PP^{11e_S^l e_S^h}) = u_P^{h_S} = \frac{\gamma_P - c_e - c_P^l}{1-\beta} - \frac{(c_P^h - c_P^l)(1 - \beta p_P^{ll} - \beta\Delta_p)}{(1-\beta)[1+\beta(1-p_P^{ll}-p_P^{hh})]}$$

The remaining effort combination follow analogously. The specialist always earns zero profit in this outcome. In treatment path SS there exists an analogue set of continuation profits for the specialist.

If providing effort is preferred for one patient type, it is also preferred for the other type:

$$\begin{aligned}
&U_P(PP^{1e_S}) \geq U_P(PP^{0e_S}), U_P(PP^{10e_S^l e_S^h}), U_P(PP^{01e_S^l e_S^h}) \iff \\
&u_P^x(PP^{1e_S}) \geq u_P^x(PP^{0e_S}), u_P^x(PP^{10e_S^l e_S^h}), u_P^x(PP^{01e_S^l e_S^h}) \forall x \in \mathcal{X} \iff \\
&\quad c_e \leq \tilde{c}_e^{\,P} \\
&U_P(PP^{0e_S}) \geq U_P(PP^{1e_S}), U_P(PP^{10e_S^l e_S^h}), U_P(PP^{01e_S^l e_S^h}) \iff \\
&u_P^x(PP^{0e_S}) \geq u_P^x(PP^{1e_S}), u_P^x(PP^{10e_S^l e_S^h}), u_P^x(PP^{01e_S^l e_S^h}) \forall x \in \mathcal{X} \iff \\
&\quad c_e \geq \tilde{c}_e^{\,P}.
\end{aligned}$$

An analogous result holds for the specialist in treatment path SS:

$$\begin{aligned}
&U_S(SS^{e_P1}) \geq U_S(SS^{e_P0}), U_S(SS^{e_P^l e_P^h 01}), U_S(SS^{e_P^l e_P^h 10}) \iff c_e \leq \tilde{c}_e^{\,S} \\
&U_S(SS^{e_P0}) \geq U_S(SS^{e_P1}), U_S(SS^{e_P^l e_P^h 01}), U_S(SS^{e_P^l e_P^h 10}) \iff c_e \geq \tilde{c}_e^{\,S}.
\end{aligned}$$

This completes the proof of the first part of Proposition 2.

The weak dominance of PP and SS over MM has already been argued in Appendix A.5. Effort provision functions the same in path MM as it does in PP for the PCP and in SS for the specialist.

The following two step argument can be made to prove that the blind treatment paths weakly dominate treatment path MP.

Step 1: PP and SS weakly dominate MP for the same effort levels (see Lemma 2).

Step 2: Whenever effort can be incentivized for MP, it can be incentivized in the blind treatment paths.

Step 2 is trivial for the PCP as her incentives are identical to the blind treatment paths. Let the PCP play (T, T). For the specialist, two conditions need to hold in order to implement MP^{11}:

$$U_S[(T,R)^{1e_S^h}] \geq U_S[(T,R)^{0e_S^h}] \iff \gamma_S \geq c_S^l + \frac{c_e(1-\beta p_S^{ll})}{\beta\Delta_p}$$
$$U_S[(T,R)^{1e_S^h}] \geq U_S[(T,T)^{11}] \iff \gamma_S \leq c_S^l + c_e + \frac{(c_S^h - c_S^l)(1-\beta(p_S^{ll}-\Delta_p))}{1+\beta(1-p_S^{ll}-p_S^{hh})}$$

These conditions can be fulfilled together if and only if $c_e \leq \tilde{c_e}^S$. Thus, the specialist is always willing to provide effort in SS if she is providing effort in MP. □

A.3. Proof for Proposition 3 (gatekeeping treatment path MS)

Proof. I assume that the payer only implements $MS^{e_Pe_S}$ if this is second-best optimal. Thus, per assumption, it is sufficient to show that physicians are indifferent between $MS^{e_Pe_S}$ and any other outcome in terms of profit, in order to show a strict preference for $MS^{e_Pe_S}$.

Let the specialist play $s_S = (T,T)^{11}$. The following describes the conditions for $s_P = (T,R)^{1e_P^h}$ being a best response for the PCP for all states.

Optimal effort choice:

Continuation profits for $s_P = (T,R)^{e_P^l e_P^h}$ are given by:

$$u_P^{l_P} = \frac{\gamma_P - c_P^l - e_P^l c_e}{1-\beta(p_P^{ll}+e_P^l\Delta_p)}$$
$$u_P^{h_P} = u_P^{l_S} = u_P^{h_S} = 0$$

It follows

$$U_P[(T,R)^{1e_P^h}] \geq U_P[(T,R)^{0e_P^h}] \iff u_P^{l_P}[(T,R)^{1e_P^h}] \geq u_P^{l_P}[(T,R)^{0e_P^h}]$$
$$\iff \gamma_P \geq c_P^l + \frac{c_e(1-\beta p_P^{ll})}{\beta\Delta_p}. \tag{25}$$

$s_P = (T,R)^{e_P^l e_P^h}$ vs. $s_P = (T,T)^{e_P^l e_P^h}$: For $s_P = (T,T)^{e_P^l e_P^h}$, continuation profits, and thus incentives for effort provision, are identical to outcome $PP^{e_P^l e_P^h 11}$ (see Appendix A.2).

$$U_P[(T,R)^{1e_P^h}] \geq U_P[(T,T)^{11}] \iff \gamma_P \leq c_P^l + c_e + \frac{\Delta_{c_P}(1-\beta(p_P^{ll}+\Delta_p))}{1+\beta(1-p_P^{ll}-p_P^{hh})} \tag{26}$$

$$U_P[(T,T)^{11}] \geq U_P[(T,T)^{00}], U_P[(T,T)^{10}], U_P[(T,T)^{01}] \iff c_e \leq \tilde{c_e}^P$$

Conditions (25) and (26) can only be simultaneously fulfilled if and only if $c_e \leq \tilde{c_e}^P$.

$s_P = (T,R)^{e_P^l e_P^h}$ vs. $s_P = (R,R)^{e_P^l e_P^h}$:

For $s_P = (R,R)^{e_P^l e_P^h}$ continuation profits for the PCP are 0.

$$U_P[(T,R)^{1e_P^h}] \geq U_P[(R,R)^{e_P^l e_P^h}] \iff \gamma_P \geq c_P^l + c_e$$

This is implied by Condition (25).

Now, let the PCP play $s_P = (T,R)^{1e_P^h}$. The following describes the conditions for $s_S = (T,T)^{11}$ being a best response for the specialist for all states. If the specialist is willing to treat h-types, then she is willing to treat l-types. Thus, it is sufficient to show that the specialist is willing to accept h-types and that she is willing to provide effort.

Optimal effort choice:

For any patient who is treated by the specialist, continuation profits are identical so outcome $SS^{e_P^l e_P^h e_S^l e_S^h}$. Continuation profits for h_P are identical to those for h_S. Continuation profits for l_P only include an additional discount factor.

$$\begin{aligned}
&U_S[(T,T)^{11}] \geq U_S[(T,T)^{00}], U_S[(T,T)^{10}], U_S[(T,T)^{01}] \iff \\
&\quad \iff c_e \leq \tilde{c_e}^S \\
&U_S[(T,T)^{00}] \geq U_S[(T,T)^{11}], U_S[(T,T)^{10}], U_S[(T,T)^{01}] \iff \\
&\quad \iff c_e \geq \tilde{c_e}^S
\end{aligned}$$

$s_S = (T,T)^{11}$ vs. $s_S = (T,R)^{e_S^l e_S^h}$:

The only difference between $s_S = (T,T)^{e_S^l e_S^h}$ and $s_S = (T,R)^{e_S^l e_S^h}$ is whether the specialist accepts the treatment of h-types referred by the PCP. Thus, $s_S = (T,T)^{e_S^l e_S^h}$ is preferred whenever the continuous treatment of initial h-types is profitable for the specialist.

$$\begin{aligned}
&U_S[(T,T)^{11}] \geq U_S[(T,R)^{11}] \iff \gamma_S \geq c_S^h + c_e - \frac{\beta \Delta_{c_S}(1 - p_S^{hh} + \Delta_p)}{1 + \beta(1 - p_S^{ll} - p_S^{hh})} =: \gamma_S^{min} \\
&U_S[(T,R)^{11}] \geq U_S[(T,R)^{00}], U_S[(T,R)^{10}], U_S[(T,R)^{01}] \iff c_e \leq \tilde{c_e}^S
\end{aligned}$$

Let us now turn to the proof for the uniqueness of the equilibrium. Consider the case with effort provision for both physicians. I will show that $s_S = (T,T)^{11}$ is the best response to $s_P = (T,T)^{e_l e_h}$ and $s_P = (R,R)^{e_l e_h}$ given $c_e \leq \tilde{c_e}^S$. Because $s_P = (T,R)^{1e_P^h}$ is the best response to $s_S = (T,T)^{11}$, the uniqueness result follows.

Consider in turn the alternatives $s_S = (T, R)^{e^l_S e^h_S}$, $s_S = (R, R)^{e^l_S e^h_S}$, and $s_S = (R, T)^{e^l_S e^h_S}$. Let $s_P = (T, T)^{e_l e_h}$, then $s_S = (T, T)^{11}$ is the best response:

$$
\begin{aligned}
U_S[(T,T)^{11}] \geq U_S[(T,R)^{1e^h_S}] \iff \gamma_S \geq \gamma_S^{min} \underset{c_e \leq \tilde{c_e}^S}{\Longrightarrow} \\
\gamma_S \geq c^h_S + c_e + \frac{c_e(1-\beta)}{\beta(1-p^{ll}_S)} - \frac{\Delta_p(c^h_S - c^l_S)(1-\beta)}{(1-p^{ll}_S)(1-\beta p^{hh}_S)} \\
- \frac{\beta(c^h_S - c^l_S)(1-p^{hh}_S)[1+\beta(\Delta_p - p^{hh}_S)]}{(1-\beta p^{hh}_S)[1+\beta(1-p^{ll}_S - p^{hh}_S)]} \iff \\
U_S[(T,T)^{11}] \geq U_S[(T,R)^{0e^h_S}]
\end{aligned}
$$

Furthermore,

$$\gamma_S \geq \gamma_S^{min} \implies U_S[(T,T)^{11}] \geq U_S[(R,R)^{e^l_S e^h_S}]$$

since

$$u^{l_S}_S[(T,T)^{11}] \geq u^{h_S}_S[(T,T)^{11}] \geq 0.$$

$s_S = (R, T)^{e^l_S e^h_S}$ is never a best response as it is dominated by $s_S = (T, T)^{e_l e_h}$ and $s_S = (R, R)^{e^l_S e^h_S}$. I have already proved that $s_P = (T, R)^{1e^h_P}$ is the best response to $s_S = (T, T)^{11}$. The same holds true in the case without effort. Thus, there exists no equilibrium with $s_P = (T, T)^{e^l_S e^h_S}$.

Let $s_P = (R, R)^{e^l_P e^h_P}$, then $s_S = (T, T)^{e^l_S e^h_S}$ is the only response that does not lead to hospital treatment. Otherwise, both physicians receive a payoff of zero. As the payoff of $s_S = (T, T)^{11}$ is non-negative for $c_e < \tilde{c_e}^S$, it is the best response. Again, $s_P = (T, R)^{1e^h_P}$ is the best response to $s_S = (T, T)^{11}$. Thus, there exists no equilibrium with $s_P = (R, R)^{e^l_P e^h_P}$.

$s_P = (R, T)^{e_l e_h}$ is never an equilibrium as it is dominated by $s_P = (T, T)^{e^l_P e^h_P}$ and $s_P = (R, R)^{e^l_P e^h_P}$.

The proof for the outcome without effort follows along the same lines. □

A.4. Proof for Proposition 4 (first-best in the team)

Proof. To show: $U_T(PS^{11}) := U_P[PS^{11}] + U_S[PS^{11}] \geq U_T(s)$
$\forall s \in \{PS^{e_P e_S}, PP^{e_P e_S}, SS^{e_P e_S}, SP^{e_P e_S}\}$

Optimal effort choice:

First, I consider the condition under which effort is provided in the team given that PCPs treat l-types and specialists treat h-types. The system of equations for the continuation profit of $PS^{e_P e_S}$ in all states $x \in \mathcal{X}$ are given by:

$$\begin{aligned}
u_T^{l_P} &= \gamma_P - c_P^l - e_P c_e + \beta[(p_P^{ll} + e_P\Delta_p)u_T^{l_P} + (1 - p_P^{ll} - e_P\Delta_p)u_T^{h_S}] \\
u_T^{l_S} &= u_T^{l_P} \\
u_T^{h_P} &= u_T^{h_S} \\
u_T^{h_S} &= \gamma_S - c_S^h - e_S c_e + \beta[(p_S^{hh} - e_S\Delta_p)u_T^{h_S} + (1 - p_S^{hh} + e_S\Delta_p)u_T^{l_P}]
\end{aligned}$$

Continuation profits are given by:

$$\begin{aligned}
u_T^l[PS^{11}] &= \frac{\gamma_P - c_P^l - c_e}{1-\beta} + \frac{\beta(1 - p_P^{ll} - \Delta_p)(\gamma_S - \gamma_P - c_S^h + c_P^l)}{(1-\beta)(1 + \beta[1 - p_P^{ll} - p_S^{hh}])} \\
u_T^h[PS^{11}] &= \frac{\gamma_P - c_P^l - c_e}{1-\beta} + \frac{(1 - \beta[p_P^{ll} + \Delta_p])(\gamma_S - \gamma_P - c_S^h + c_P^l)}{(1-\beta)(1 + \beta[1 - p_P^{ll} - p_S^{hh}])}
\end{aligned}$$

The remaining effort combination follow analogously. In the team, one physician always treats all patients of the same type. Thus, continuation profits for l^P and l^S as well as h^P and h^S are identical.

$$\begin{aligned}
U_T(PS^{11}) \geq U_T(PS^{10}), U_T(PS^{01}), U_T(PS^{00}) &\iff \\
u_T^l[PS^{11}] \geq u_T^l[PS^{10}], u_T^l[PS^{01}], u_T^l[PS^{00}] &\iff \\
u_T^h[PS^{11}] \geq u_T^h[PS^{10}], u_T^h[PS^{01}], u_T^h[PS^{00}] &\iff \\
(\gamma_S - c_S^h) - (\gamma_P - c_P^l) \leq -\frac{c_e[1 + \beta(1 - p_P^{ll} - p_S^{hh})]}{\beta\Delta_p}
\end{aligned}$$

For all compared strategies, differences in continuation profits are identical for a fixed state. Therefore, I omit continuation profits in the following.

PS vs PP:

$$\begin{aligned}
U_T(PP^{11e_S^l e_S^h}) \geq U_T(PP^{10e_S^l e_S^h}), U_T(PP^{01e_S^l e_S^h}), U_T(PP^{00e_S^l e_S^h}) &\iff \\
c_e \leq \tilde{c}_e^{\,P} \\
U_T(PP^{00e_S^l e_S^h}) \geq U_T(PP^{10e_S^l e_S^h}), U_T(PP^{01e_S^l e_S^h}), U_T(PP^{11e_S^l e_S^h}) &\iff \\
c_e \geq \tilde{c}_e^{\,P} \\
U_T(PS^{11e_S^l e_S^h}) \geq U_T(PP^{11e_S^l e_S^h}) &\iff \\
U_T(PS^{00e_S^l e_S^h}) \geq U_T(PP^{00e_S^l e_S^h}) &\iff \\
(\gamma_S - c_S^h) - (\gamma_P - c_P^l) \geq \frac{-[1 + \beta(1 - p_P^{ll} - p_S^{hh})]\Delta_{c_P}}{1 + \beta(1 - p_P^{ll} - p_P^{hh})}
\end{aligned}$$

PS vs SS:

For SS the continuation profits of the team are identical to continuation profits for PP with all indices P switched so S.

PS vs SP:

For SP the continuation profits of the team are identical to continuation profits for PS with all indices P switched so S and indices S switched so P. Thus, SP is dominated if

$$\Delta_{c_P} \geq \frac{\Delta_{c_S}[1+\beta(1-p_P^{ll}-p_P^{hh})]}{1+\beta(1-p_S^{ll}-p_S^{hh})}$$

For outcome PS^{11} to be implemented with smallest possible payment fees, Condition (27) must hold in order for the team to make non-negative profits for both patient types.

$$u_T^l(PS^{11}) \geq u_T^h(PS^{11}) = 0 \iff$$

$$\frac{\gamma_P - c_P^l - c_e}{1-\beta} + \frac{[1-\beta(p_P^{ll}+\Delta_p)(\gamma_S-\gamma_P+c_P^l-c_S^h)]}{(1-\beta)[1+\beta(1-p_P^{ll}-p_S^{hh})]} = 0 \tag{27}$$

Solving for γ_P and inserting the binding Condition (27) into the binding Incentive Constraint (22), delivers

$$\gamma_S^* = c_S^h + c_e - \frac{\beta\Delta_{c_P}(1-p_S^{hh}+\Delta_p)}{1+\beta(1-p_P^{hh}-p_P^{ll})}.$$

Inserting γ_S^* into Condition (27) delivers

$$\gamma_P^* = c_P^l + c_e + \frac{\beta\Delta_{c_P}(1-p_P^{ll}-\Delta_p)}{1+\beta(1-p_P^{hh}-p_P^{ll})}.$$

Because Condition (22) is the only lower boundary on γ_S, whereas Conditions (20) and (21) are upper boundaries, this contract implements PS^{11} whenever this is possible.

Necessary conditions:

Condition (20) and (21) are both upper boundaries on γ_S. If $c_e \leq \tilde{c}_e{}^S$, Condition (21) is implied by Condition (20), otherwise Condition (20) is implied by Condition (21).

Let $c_e \leq \tilde{c}_e{}^S$. In this case FB^{11} can be implemented if and only if Condition (20) and Condition (22) can be fulfilled together, i.e.

$$-\frac{[1+\beta(1-p_P^{ll}-p_S^{hh})]\Delta_{c_S}}{1+\beta(1-p_S^{ll}-p_S^{hh})} \geq \frac{-[1+\beta(1-p_P^{ll}-p_S^{hh})]\Delta_{c_P}}{1+\beta(1-p_P^{ll}-p_P^{hh})} \iff$$

$$\Delta_{c_P} \geq \frac{\Delta_{c_S}[1+\beta(1-p_P^{ll}-p_P^{hh})]}{1+\beta(1-p_S^{ll}-p_S^{hh})}.$$

Let $c_e \geq \tilde{c}_e{}^S$. In this case FB^{11} can be implemented if and only if Condition (21) and Condition (22) can be fulfilled together, i.e.

$$-\frac{c_e[1+\beta(1-p_P^{ll}-p_S^{hh})]}{\beta\Delta_p} \geq \frac{-[1+\beta(1-p_P^{ll}-p_S^{hh})]\Delta_{c_P}}{1+\beta(1-p_P^{ll}-p_P^{hh})} \iff$$
$$c_e \leq \tilde{c}_e{}^P.$$

As per assumption the team strictly prefers the first-best treatment paths over the other paths given that it receives the same profit.

□

A.5. Proof for Lemma 2 (welfare ordering)

Proof. Part 1:

$$\begin{aligned}
\mathbb{E}W(PS^{11}) \geq \mathbb{E}W(PM^{11}) &\iff \mathbb{E}W(PS^{00}) \geq \mathbb{E}W(PM^{00}) \\
\mathbb{E}W(PM^{11}) \geq \mathbb{E}W(PP^{1e_S}) &\iff \mathbb{E}W(PM^{00}) \geq \mathbb{E}W(PP^{0e_S}) \\
&\iff c_S^h - c_P^h \leq \frac{(p_P^{hh}-p_S^{hh})\beta(L+c_P^h-c_P^l)}{1+\beta(1-p_P^{ll}-p_P^{hh})} \\
\mathbb{E}W(PS^{11}) \geq \mathbb{E}W(MS^{11}) &\iff \mathbb{E}W(PS^{00}) \geq \mathbb{E}W(MS^{00}) \\
\mathbb{E}W(MS^{11}) \geq \mathbb{E}W(SS^{e_P1}) &\iff \mathbb{E}W(MS^{00}) \geq \mathbb{E}W(SS^{e_P0}) \\
&\iff c_S^l - c_P^l \geq \frac{(p_S^{ll}-p_P^{ll})\beta(L+c_S^h-c_P^l)}{1+\beta(1-p_P^{ll}-p_S^{hh})}
\end{aligned}$$

The part of the Proposition concerning strict inequality follows analogously.

Part 2: The dominance of PP and SS over MP and SM has been demonstrated in Appendix A.1 already. In treatment path MM one part of the patients receives continuous PCP treatment and the other part receives continuous specialist treatment. Clearly, either PP or SS is weakly preferred by the payer. □

A.6. Proposition 6 (second-best in the team)

Proposition 6. *Let $\Delta_{c_P} < \tilde{\Delta}_{c_P}$. For the team, the following statements hold.*

If $c_e \leq \tilde{c}_e{}^P$, PP^{1e_S} or SS^{e_P1} is second-best.

If $c_e > \tilde{c}_e{}^S$, PP^{0e_S} or SS^{e_P0} is second-best.

If $\tilde{c}_e{}^P < c_e \leq \tilde{c}_e{}^S$, PP^{0e_S}, SS^{e_P1}, or SP^{11} is second-best.

Proof. If the specialist has the relative cost-advantage for l-types ($\Delta_{c_P} < \tilde{\Delta}_{c_P}$), the treatment path PS cannot be implemented in the team. In this case, PP^{1e_S} can be implemented for $c_e \leq \tilde{c_e}^P$, whereas SS^{e_P1} and SP^{11} can be implemented for $c_e < \tilde{c_e}^S$. If the conditions hold with flipped inequality, the respective outcomes can be implemented without effort provision. The proofs for these statements regarding the blind paths are identical to the solo practice case. The proof for path SP follows by exchanging indices S and P for the proof for PS.

If $c_e \leq \tilde{c_e}^P$ or $c_e > \tilde{c_e}^S$, one of the blind treatment paths PP or SS is second-best optimal as the remaining path SP is weakly dominated by them for equal effort levels. However, if $\tilde{c_e}^P < c_e \leq \tilde{c_e}^S$, physicians provide more effort in path SP than in path PP. Note that the dominance of PS^{11} over PP^{11} does not imply the dominance of SS^{e_P1} over SP^{11}. The reason for this is that in PS^{11} l-types are treated by the PCP and in SS^{e_P1} l-types are treated by the specialist. Therefore, because PCP treatments is generally more efficient than specialist treatment for l-types, it is more important to increase the transition probability from h to l in the first case. Consequently, SP^{11} can be second-best optimal in the team. In Appendix A.9 I show by means of simulation that feasible parameters do indeed exist such that SP^{11} is second-best optimal. □

A.7. Proposition 7 (sufficient first-best condition in team)

Proposition 7. *Let λ^* be the set of parameter combinations satisfying Conditions (4) to (6). The first-best outcome PS^{11} can be implemented in the team for all $\lambda \in \lambda^*$ that satisfy*

$$L \leq \min(L^1, L^2) \text{ with}$$
$$L^1 := [1 + \beta(1 - p_P^{ll} - p_P^{hh})] \left(\frac{c_S^h - c_P^h}{\beta(p_P^{hh} - p_S^{hh})} - \frac{\Delta_{c_S}}{1 + \beta(1 - p_S^{ll} - p_S^{hh})} \right),$$
$$L^2 := c_P^l - c_S^h + \frac{\Delta_{c_P}(1 + \beta(1 - p_P^{ll} - p_S^{hh}))}{1 + \beta(1 - p_P^{ll} - p_P^{hh}))}.$$

Conversely, if $L > \min(L^1, L^2)$, there exist $\lambda \in \lambda^*$ such that PS^{11} cannot be implemented in the team.

Proof. If PS^{11} can be implemented, Condition (23) provides a lower bound for Δ_{c_P}. If PS^{11} is first-best, Condition (5) also provides a lower bound on Δ_{c_P}. The latter condition is more strict if and only if

$$L \leq L^1.$$

Thus, if $L \leq L^1$, Condition (5) implies Condition (23). Conversely, if $L > L^1$, a λ that fulfills Condition (5) with equality as well as $c_e = 0$ and $p_P^{ll} = p_S^{ll}$, fulfills the first-best conditions but PS^{11} cannot be implemented in the team because Condition (23) is not fulfilled.

Let us now consider the optimal effort. Effort can be incentivized in the team if Condition (24) is fulfilled. In terms of effort provision, PS^{11} is first-best if Condition (4) is fulfilled. Both conditions are upper boundaries on the cost of effort provision c_e. The latter condition is more strict if and only if

$$L \leq L^2.$$

Thus, if $L \leq L^2$, Condition (4) implies Condition (24). Conversely, if $L > L^2$, a λ that fulfills Condition (4) and (5) with equality as well as $p_P^{ll} = p_S^{ll}$, fulfills the first-best conditions but PS^{11} cannot be implemented in the team because Condition (24) is not fulfilled.

□

A.8. Condition for dominated SP

To show: If Condition (34) does not hold, treatment path PP dominates SP.

Proof.

$$\begin{aligned} U_T(PP^{1es}) \geq U_T(SP^{11}) \iff U_T(PP^{0es}) \geq U_T(SP^{00}) \iff \\ c_S^l \geq c_P^l - \frac{\beta(c_P^h - c_P^l)(p_P^{ll} - p_S^{ll})}{1 + \beta(1 - p_P^{ll} - p_P^{hh})} \end{aligned} \tag{28}$$

Clearly, this condition holds for $p_P^{ll} \geq p_S^{ll}$. Assume $p_P^{ll} < p_S^{ll}$. Then Condition (35) implies Condition (28) if and only if Condition (34) does not hold:

$$\begin{aligned} c_P^l - \frac{\beta(c_P^h - c_P^l)(p_P^{ll} - p_S^{ll})}{1 + \beta(1 - p_P^{ll} - p_P^{hh})} \leq c_P^l - \frac{\beta(c_S^h - c_P^l)(p_P^{ll} - p_S^{ll})}{1 + \beta(1 - p_P^{ll} - p_S^{hh})} \iff \\ c_S^h - c_P^h \geq \frac{(p_P^{hh} - p_S^{hh})\beta(c_P^h - c_P^l)}{1 + \beta(1 - p_P^{ll} - p_P^{hh})} \iff \\ U_T(PS^{11}) \leq U_T(PP^{1es}) \end{aligned}$$

Because Condition (35) is already implied by first-best Condition (6), PP dominates SP. □

A.9. Proposition 8 (strict superiority of MS)

Proposition 8. *Let* $\Delta_{c_P} < \tilde{\Delta}_{c_P}$.

(I) Let $c_e \leq \min(\tilde{c}_e{}^S, \tilde{c}_e{}^P)$. *There exist parameters such that* MS^{11} *is strictly superior to all other outcomes in both solo practices and the team.*

(II) Let $c_e > \max(\tilde{c}_e{}^S, \tilde{c}_e{}^P)$. *There exist parameters such that* MS^{00} *is strictly superior to all other outcomes in both solo practices and the team.*

(III) Let $\tilde{c}_e{}^P < c_e \leq \tilde{c}_e{}^S$. *There exist parameters such that* MS^{01} *is strictly superior to all other outcomes in both solo practices and the team.*

In order for either statement to be true, c_P^h *must be sufficiently large.*

(I) For $c_e \leq \min(\tilde{c}_e{}^S, \tilde{c}_e{}^P)$, there are, considering only unique equilibria, three possible second-best outcomes: MS^{11} (only solo practices), $SS^{e_P 1}$, and PP^{1e_S}. Given fixed efforts, MS always yields larger welfare than SS (see Lemma 2). Hence, for $c_e \leq \min(\tilde{c}_e{}^S, \tilde{c}_e{}^P)$ implementing MS^{11} is always preferred to $SS^{e_P 1}$.

Conditions for which MS^{11} is superior to PP^{1e_S}:

$$\begin{aligned} W^{l_P}_{MS^{11}} \geq W^{l_P}_{PP^{1e_S}} \iff W^{h_P}_{MS^{11}} \geq W^{h_P}_{PP^{1e_S}} \iff W^{h_S}_{MS^{11}} \geq W^{h_S}_{PP^{1e_S}} \iff \\ c_P^h \geq \frac{\beta(1-p_P^{hh}+\Delta_p)([L+c_S^h][-1+\beta(p_S^{ll}+\Delta_p)]+c^l_P[1+\beta(1-p_S^{hh}-p_S^{ll})]+c^l_S r[-1+p_S^{hh}-\Delta_p])}{[-1+\beta(p_P^{ll}+\Delta_p)][1+\beta(1-p_S^{hh}-p_S^{ll})]} \\ - \frac{(L+c^l_S)\beta(1-p_S^{hh}+\Delta_p)+c_S^h(-1+\beta(p_S^{ll}+\Delta_p))}{[1+\beta(1-p_S^{hh}-p_S^{ll})]} \end{aligned} \tag{29}$$

$$\begin{aligned} W^{l_S}_{MS^{11}} \geq W^{l_S}_{PP^{1e_S}} \iff \\ c_P^h \geq \frac{L\beta[1-p_P^{ll}-\Delta_p]+c^l_P[1+\beta(\Delta_p-p_P^{hh})]}{\beta[-1+p_P^{ll}+\Delta_p]} \\ - \frac{\{1+\beta(1-p_P^{hh}-p_P^{ll})\}\{(L+c_S^h)\beta(-1+p_S^{ll}+\Delta_p)+c^l_S[-1+\beta(p_S^{hh}-\Delta_p)]\}}{[\beta(-1+p_P^{ll}+\Delta_p)][1+\beta(1-p_S^{ll}-p_S^{hh})]} \end{aligned} \tag{30}$$

Given Condition (5), Condition (30) is stricter than Condition (29). Thus, if Condition (30) is fulfilled, MS^{11} is strictly superior to PP^{1e_S}. Let us consider now whether parameters exist, such that this is true.

Inserting the supremum c_P^h implied by $\Delta_{c_P} < \tilde{\Delta}_{c_P}$ and the maximum c_P^l implied by Condition (6) delivers

$$\begin{aligned} W^{l_S}_{MS^{11}} \geq W^{l_S}_{PP^{1e_S}} \iff \\ p_P^{hh} - p_S^{hh} \geq p_S^{ll} - p_P^{ll}, \end{aligned} \tag{31}$$

which is strictly true by assumption. Thus, there exists a set of parameters such that MS^{11} is strictly superior to PP^{1e_S} and $SS^{e_P 1}$.

(II) For MS^{00}, an analogue argument applies. The condition for MS^{00} to be strictly superior to PP^{0e_S} is:

$$
\begin{aligned}
W^{l_S}_{MS^{00}} \geq W^{l_S}_{PP^{0e_S}} &\iff \\
c^h_P \geq & \frac{L\beta[1-p^{ll}_P] + c^l_P[1+\beta(-p^{hh}_P)]}{\beta[-1+p^{ll}_P]} \\
& - \frac{\{1+\beta(1-p^{hh}_P - p^{ll}_P)\}\{(L+c^h_S)\beta(-1+p^{ll}_S+) + c^l_S[-1+\beta(p^{hh}_S)]\}}{[\beta(-1+p^{ll}_P)][1+\beta(1-p^{ll}_S-p^{hh}_S)]}
\end{aligned}
$$

To illustrate the result, Figure 6 depicts a simulation that shows which outcome is strictly superior to the other outcome for $c_e \leq \tilde{c_e}^P$ and $\tilde{c_e}^S < c_e$ depending on feasible parameters c_e and c^h_P. Clearly, there exist feasible parameters such that either outcome can be optimal.

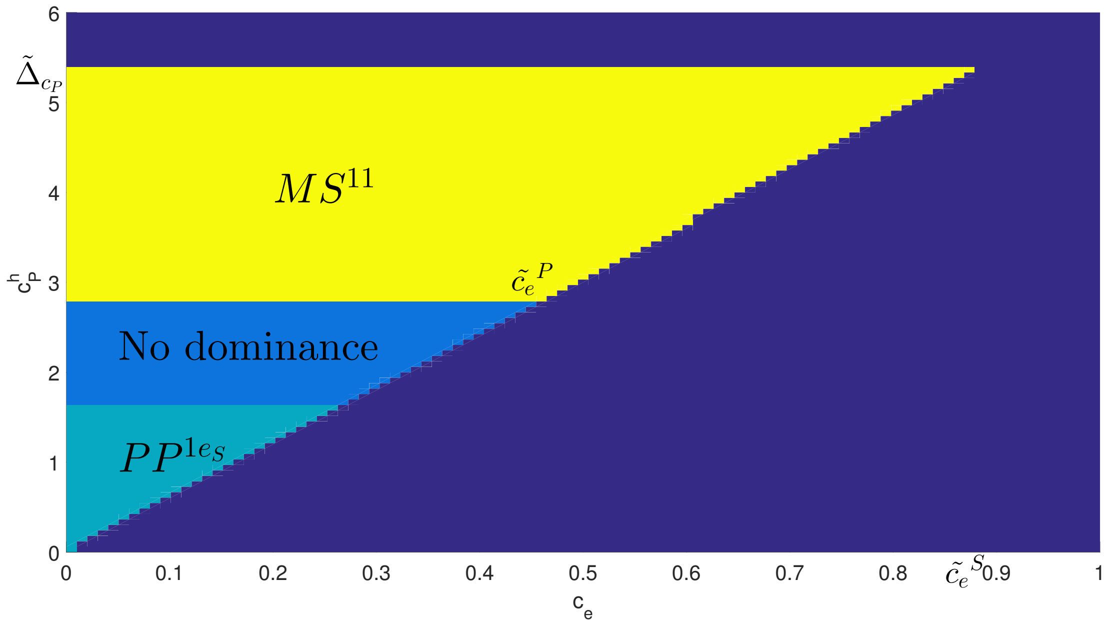
(III) Let us now turn to the case $\tilde{c_e}^P < c_e \leq \tilde{c_e}^S$. In this case, there are, four possible second-best outcomes: MS^{01} (only solo practices), SS^{e_P1}, SP^{11} (only team) and PP^{1e_S}

There is a trade-off between always receiving specialist effort in outcome SS^{e_P1} and better referral efficiency in outcome MS^{01}. The payer prefers MS^{01} over SS^{e_P1} if and only if the additional costs of specialist treatment for l-types is larger than its expected benefit:

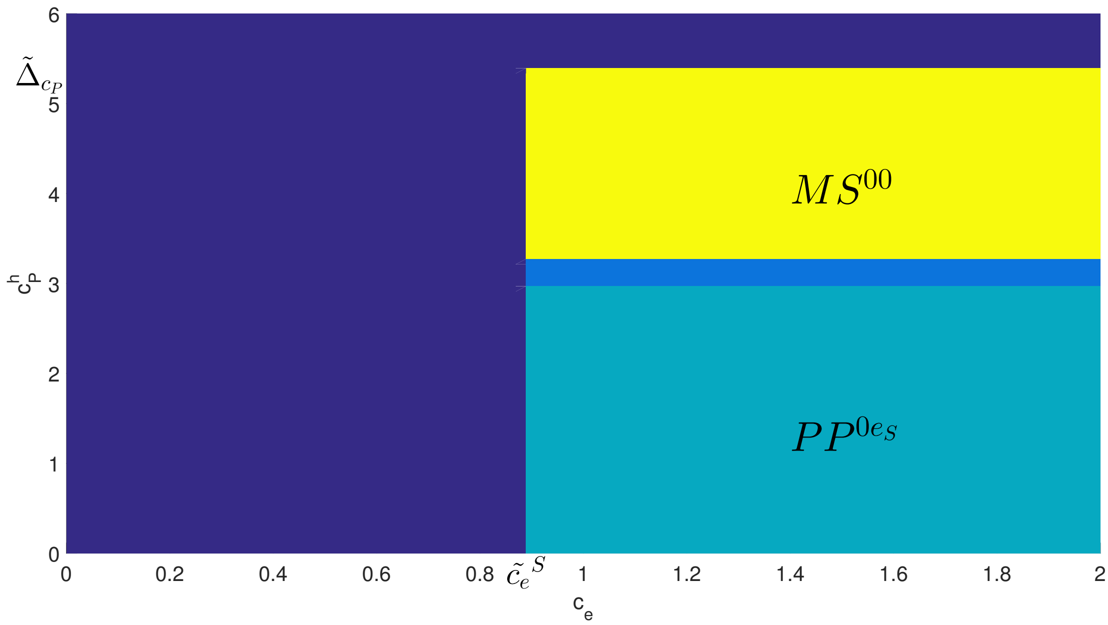
$$
c_e - c^l_P + c^l_S \geq \frac{\beta(L+c^h_S-c^l_S)(\Delta_p + p^{ll}_S - p^{ll}_P)}{1+r(1-p^{ll}_S - p^{hh}_S)} \tag{32}
$$

Let us now turn to the comparison of MS^{01} and PP^{0e_S}.

$$
\begin{aligned}
W^{l_S}_{MS^{01}} \geq W^{l_S}_{PP^{0e_S}} &\iff \\
c^h_P \geq & \frac{[1+\beta(1-p^{ll}_P-p^{hh}_P)](L+c^h_S)\beta[-1+\Delta_p+p^{ll}_S] + c_e[-1+\beta(-1+p^{ll}_S+p^{hh}_S)] + c^l_S[-1+r(-\Delta_p+p^{hh}_S)]}{[1+\beta(1-p^{ll}_S-p^{hh}_S)]\beta(p^{ll}_P-1)} \\
& + \frac{L\beta(1-p^{ll}_P) + c^l_P(1-\beta p^{hh}_P)}{\beta(p^{ll}_P-1)} \\
W^{l_P}_{MS^{01}} \geq W^{l_P}_{PP^{0e_S}} &\iff W^{h_P}_{MS^{01}} \geq W^{h_P}_{PP^{0e_S}} \iff W^{h_S}_{MS^{01}} \geq W^{h_S}_{PP^{0e_S}} \iff \\
c^h_P \geq & -\frac{(L-c^l_S)\beta[1+\Delta_p-p^{hh}_S] + c_e[-1+\beta(-1+p^{hh}_S+p^{ll}_S)] + c^h_S[-1+\beta(\Delta_p+p^{ll}_S)]}{1+\beta(1-p^{ll}_S-p^{hh}_S)} \\
& + \frac{\beta(p^{hh}_P-1)\{(L+c^h_S)[-1+\beta(\Delta_p+p^{ll}_S)] + c^l_P[1+\beta(1-p^{ll}_S-p^{hh}_S)] + c_e[-1+\beta(-1+p^{ll}_S+p^{hh}_S)]\}}{\beta(1-p^{ll}_P)[1+\beta(1-p^{ll}_S-p^{hh}_S)]}
\end{aligned}
$$

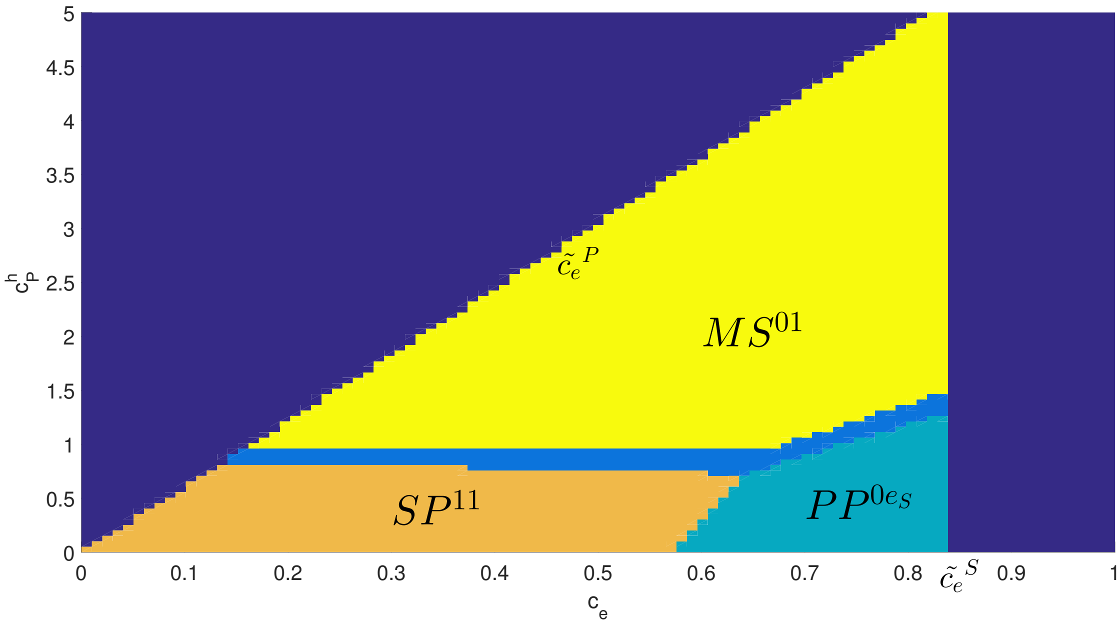


(a) $c_e < \tilde{c_e}^P, L = 45$

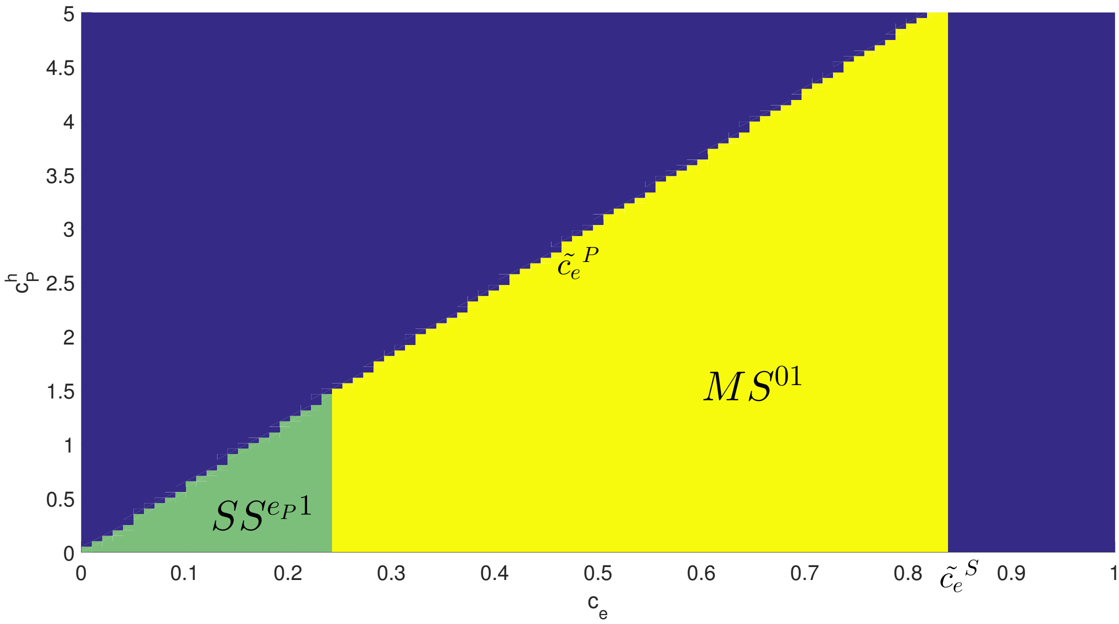


(b) $\tilde{c_e}^S < c_e, L = 25$. There exists no dominant outcome in the blue boundary region in the middle between optimal outcomes.

Figure 6: Simulation of the second-best outcome with $c_P^l = 0, p_P^{ll} = 0.8, p_P^{hh} = 0.6, c_S^l = 2, c_S^h = 10, p_S^{ll} = 0.8, p_S^{hh} = 0.3, \Delta_p = 0.1, r = 0.99$.



(a) $L = 21$. There exists no dominant outcome in the blue boundary region in the middle between optimal outcomes.



(b) $L = 35$

Figure 7: Simulation of the second-best outcome for $\tilde{c}_e{}^P < c_e < \tilde{c}_e{}^S$ with $c_P^l = 0, p_P^{ll} = 0.8, p_P^{hh} = 0.6, c_S^l = 2, c_S^h = 10, p_S^{ll} = 0.75, p_S^{hh} = 0.3, \Delta_p = 0.1, r = 0.99$.

Let us now turn to the comparison of MS^{01} and SP^{11}.

$$
\begin{aligned}
&W^{l_P}_{MS^{01}} \geq W^{l_P}_{SP^{11}} \iff \\
&\quad c^h_P \geq \frac{(L - c^l_P + c_e)(1-\beta) + c^h_S \beta(-1 + p^{ll}_P)}{p^{ll}_P - 1} \\
&\qquad + \frac{(\beta-1)[1+\beta(\Delta_p - p^{hh}_P)]\{(L + c^h_S)\beta(-1+p^{ll}_P) + c^l_S[1+\beta(1+\Delta_p - p^{ll}_P - p^{hh}_S)] + (c_e - c^l_P)[1+\beta(\Delta_p - p^{hh}_S)]\}}{\beta(p^{ll}_P - 1)(\Delta_p + p^{ll}_S - 1)[1+\beta(\Delta_p - p^{hh}_S)]} \\
&\qquad + \frac{\beta^3(p^{hh}_P - p^{hh}_S)(p^{ll}_P - 1)(\Delta_p - p^{hh}_S + 1)(L + c^h_S - c^l_S)}{(1-\beta p^{ll}_P)[1+\beta(\Delta_p - p^{hh}_S)][1+\beta(1-p^{ll}_S - p^{hh}_S)]} \\
&W^{l_S}_{MS^{01}} \geq W^{l_S}_{SP^{11}} \iff W^{h_P}_{MS^{01}} \geq W^{h_P}_{SP^{11}} \iff W^{h_S}_{MS^{01}} \geq W^{h_S}_{SP^{11}} \iff \\
&\quad c^h_P \geq c^h_S - \frac{\beta(p^{hh}_P - p^{hh}_S)(L + c^h_S - c^l_S)}{1+\beta(1-p^{ll}_S - p^{hh}_S)}
\end{aligned}
$$

Figure 7 depicts a simulation that shows which outcome is strictly superior to all other outcomes for $\tilde{c_e}^P < c_e < \tilde{c_e}^S$ depending on feasible parameters c_e and c^h_P. Clearly, there exist feasible parameters such that either outcome can be optimal.

B. Extensions

B.1. Flat Fees in the Team

In this Subsection, I consider the effects of paying a flat periodical treatment fee to the physician team, i.e. a fee that is not differentiated by the type of physician who provided treatment during the period ($\gamma_P = \gamma_S$). This is relevant if practices are paid in bundled, periodical payments for the provision of both primary and specialty care. Bundled payments may be used because the payer cannot differentiate which team physician has supplied a treatment. Naturally, we can expect to implement the first-best outcome in a smaller parameter region than before because the optimization problem includes an additional condition. In fact, the size of the flat fee does not impact incentives in the team because the team's behavior is influenced only by the difference in treatment fees between physicians (see Conditions (20) to (22)). The solo practices are still paid by treatment fees that may differ. Consequently, solo practices perform relatively better against the team than before.

Hospital treatment can be prevented by paying sufficiently large flat fees. After eliminating all non-binding incentive constraints, the following constraints remain in order for PS^{11} to be played:

$$U_T(PS^{11}) \geq U_T(PS^{00}) \iff c_e \leq \frac{\Delta_p \beta(c_S^h - c_P^l)}{1+\beta(1-p_P^{ll}-p_S^{hh})} \quad (33)$$

$$U_T(PS^{11}) \geq U_T(PP^{1e_S}) \iff c_S^h - c_P^h \leq \frac{(p_P^{hh}-p_S^{hh})\beta(c_P^h - c_P^l)}{1+\beta(1-p_P^{ll}-p_P^{hh})} \quad (34)$$

$$U_T(PS^{11}) \geq U_T(SS^{e_P 1}) \iff c_S^l - c_P^l \geq \frac{(p_S^{ll}-p_P^{ll})\beta(c_S^h - c_P^l)}{1+\beta(1-p_P^{ll}-p_S^{hh})} \quad (35)$$

Note that these conditions are identical to the first-best Conditions (4) to (6) given that $L = 0$. Thus, the team minimizes expected treatment costs without internalizing the patient's health loss. Inefficient treatment is thus provided if and only if providing PS^{11} is not cost-minimizing but still socially efficient due to patient health losses. PS^{00} is played if Condition (33) holds with flipped inequality. Providing effort to one type but not the other is always weakly dominated by providing effort for both types or not providing effort for either type.

Condition (35) states that the additional costs of specialist treatment are greater than any cost savings from specialist treatment that accrue in the future. This is already implied by first-best Condition (6). Because fees are flat, the team would only use expensive specialists treatment for l-types, if specialist treatment is much superior to PCP treatment. In this case, however, specialist treatment for l-types would be preferred by the payer because the payer also considers the health benefits of the patient. As shown in Appendix A.8, SP is also never played in the team with flat fees. Thus, specialist over-treatment is not an issue.

Let the minimum c_e implied by Condition (33) be defined as $\bar{c}_e$ and the maximum Δ_{c_P} implied by Condition (34) as $\bar{\Delta}_{c_P}$, i.e.

$$\bar{c}_e := \frac{\Delta_p \beta(c_S^h - c_P^l)}{1+\beta(1-p_P^{ll}-p_S^{hh})}$$
$$\bar{\Delta}_{c_P} := \frac{(c_S^h - c_P^h)[1+\beta(1-p_P^{ll}-p_P^{hh})]}{\beta(p_P^{hh}-p_S^{hh})}.$$

Figure 8 depicts the outcomes in the team given flat fees. We can make the following observations in Lemma 3 (follows directly from Conditions (6) & (5)).

Lemma 3. *Compared to the optimal treatment fees, under flat fees, a larger difference in expected treatment costs for the PCP is necessary in order to implement the first-best outcome* PS^{11}*:*

$$\bar{\Delta}_{c_P} \geq \tilde{\Delta}_{c_P}$$

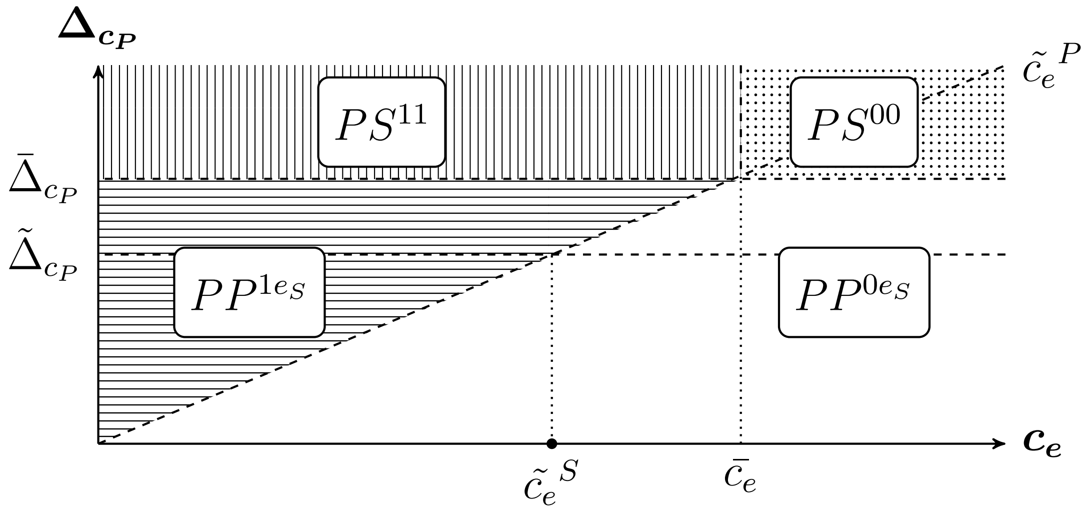


Figure 8: Outcomes in the team with flat fees ($\gamma_P = \gamma_S$).

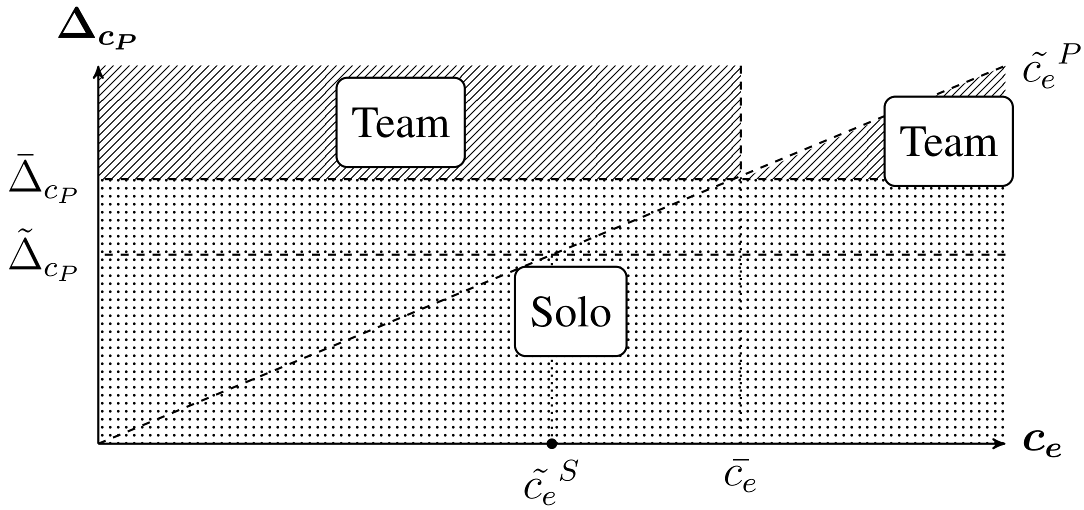


Figure 9: Second-best optimal organization of care given $\gamma_P = \gamma_S$: weak dominance for solo practices (dots), strict dominance for the team (lines), no general dominance relation (no pattern)

Given that implementing treatment path PS is cost-efficient ($\Delta_{c_P} \geq \bar{\Delta}_{c_P}$), effort is implemented only for lower effort costs:

$$\bar{c}_e \leq \tilde{c}_e{}^P$$

with

$$\bar{c_e} = \tilde{c_e}^P \text{ if } \Delta_{cP} = \bar{\Delta}_{cP}.$$

Furthermore, if $\bar{\Delta}_{cP} < \tilde{\Delta}_{cP}$, only blind PCP treatment is provided.

Lemma 3 and Figure 8 highlight the three problems that exist when fees are flat. Firstly, for $\Delta_{cP} \in [\tilde{\Delta}_{cP}, \bar{\Delta}_{cP}]$, the treatment fee for the PCP is too large compared to the specialist's fee. The result is too little specialist treatment. This could have been prevented by a relative increase in the specialist's fee. Secondly, SP^{11} cannot be implemented anymore. Consequently, as Figure 9 shows, the area for which solo practices weakly dominate team care is increased. Thirdly, for $\Delta_{cP} \geq \bar{\Delta}_{cP}$ and $c_e \in [\bar{c_e}, \tilde{c_e}^P]$, the treatment fee for the PCP is too small compared to the specialist's fee. The result of this is an undersupply of effort, which could have been prevented by a relative increase in the PCP's fee.

As Figure 9 shows, team care does not strictly dominate solo practices for all parameters given $\Delta_{cP} > \bar{\Delta}_{cP}$. Instead, there exists a parameter region $c_e \in [\bar{c_e}, \tilde{c_e}^P]$, in which the team plays PS^{00} and the solo practices play either PP^{1es} or MS^{10}. Depending on the exact parameters, either organizational form may perform better. Here, the team has the advantage for the referral efficiency, while the solo practices can provide larger PCP effort. Proposition 9 summarizes the results.

Proposition 9. *Let teams be paid by flat fees and solo practices by differing fees. Then the following observations can be made.*

1. *Solo practices weakly dominate team care whenever treatment path PS is not cost-efficient ($\Delta_{cP} < \bar{\Delta}_{cP}$).*
2. *Given that PS is cost-efficient, team care (strictly) dominates solo practice care (given that Conditions (4) to (6) hold strictly), except when effort costs are moderate ($c_e \in [\bar{c_e}, \tilde{c_e}^P]$). In this case, neither organizational form is dominant for all parameters.*

B.2. Physician Altruism

Let us now consider how the main result of Section 5 changes if physicians are partially altruistic. As in the seminal model of Ellis and T. G. McGuire (1986), physicians are assumed to consider the patient's benefit in addition to their own profits. In order to ensure a fair comparison between the organizational settings, I assume that both physicians and the physician team internalize patient benefit to the same degree α which is commonly known, i.e. the solo practice physicians' continuation utility is

$$u_k^{x\alpha}(s) = -\alpha\mathcal{L}(s,x) + \Gamma_k(s,x) - \mathcal{C}_k(s,x) + \beta \sum_{\tilde{x}\in\mathcal{X}} p(\tilde{x}|s,x) u_k^{\tilde{x}\alpha}(s), x \in \mathcal{X} \tag{36}$$

and the discounted expected utility in period $t = 0$ is

$$U_k^{\alpha}(s) = F_k + l_P^0 u_k^{l_P\alpha}(s) + l_S^0 u_k^{l_S\alpha}(s) + h_P^0 u_k^{h_P\alpha}(s) + h_S^0 u_k^{h_S\alpha}(s). \tag{37}$$

For the team, continuation utility is

$$\begin{aligned} u_T^{x\alpha}(s) =& -\alpha\mathcal{L}(s,x) + \Gamma_P(s,x) + \Gamma_S(s,x) - \mathcal{C}(s,x) + \\ & \beta \sum_{\tilde{x}\in\mathcal{X}} p(\tilde{x}|s,x) u_T^{\tilde{x}\alpha}(s), x \in \mathcal{X} \end{aligned} \tag{38}$$

and the discounted expected utility in period $t = 0$ is

$$U_T^{\alpha}(s) = F_T + l_P^0 u_T^{l_P\alpha}(s) + l_S^0 u_T^{l_S\alpha}(s) + h_P^0 u_T^{h_P\alpha}(s) + h_S^0 u_T^{h_S\alpha}(s). \tag{39}$$

I assume, for simplicity, that solo practice physicians are indifferent between the other physician or the hospital treating a patient. Using the following two-step argument I will show that physician altruism increases the set of parameters for which the team weakly dominates the solo practices in terms of social welfare.

1. If the team acts in an altruistic manner, the parameter region in which PS can be implemented grows.

2. Given that PS can be implemented in the team, effort is always as least as large in the team as it is in the solo practices.

Consider step 1 first. Including altruism changes Conditions (20) to (22) for the implementation of PS^{11} to:

$$PS^{11} \geq SS^{e_P 1} : (\gamma_S - c_S^l) - (\gamma_P - c_P^l) \leq \frac{(p_P^{ll} - p_S^{ll})\beta(\alpha L + \Delta_{c_S})}{1 + \beta(1 - p_S^{ll} - p_S^{hh})} \tag{40}$$

$$PS^{11} \geq PS^{00} : (\gamma_S - c_S^h) - (\gamma_P - c_P^l) \leq \alpha L - \frac{c_e[1 + \beta(1 - p_P^{ll} - p_S^{hh})]}{\beta\Delta_p} \tag{41}$$

$$PS^{11} \geq PP^{1es} : (\gamma_S - c_S^h) - (\gamma_P - c_P^h) \geq -\frac{(p_P^{hh} - p_S^{hh})\beta(\alpha L + \Delta_{c_P})}{1 + \beta(1 - p_P^{ll} - p_P^{hh})} \tag{42}$$

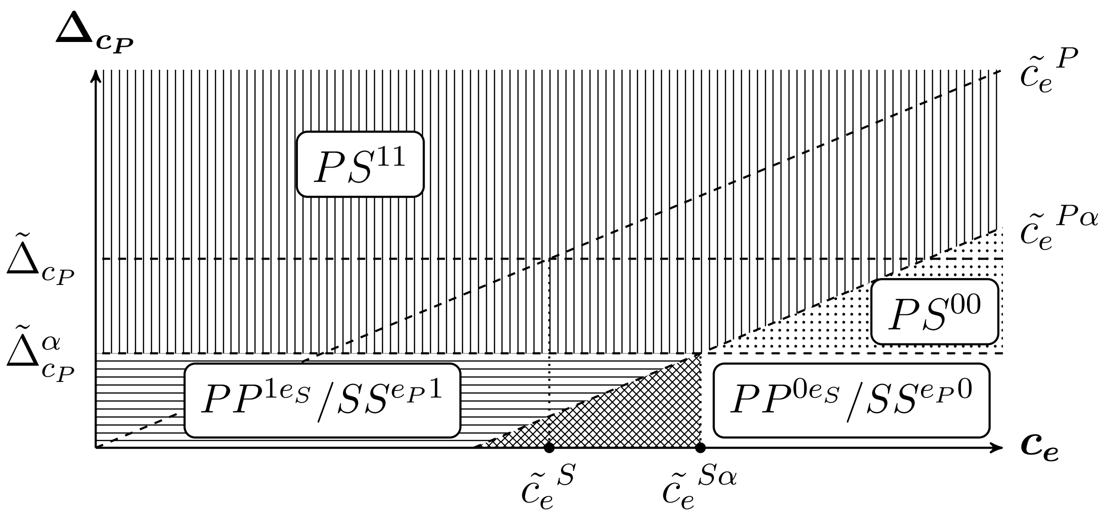


Figure 10: Second-best implementation in the altruistic team: crosshatch pattern = $PP^{0e_S}/SP^{11}/SS^{e_P1}$

If $c_e \leq \tilde{c_e}^{S\alpha}$ with

$$\tilde{c_e}^{k\alpha} := \frac{\Delta_p\beta(\alpha L + \Delta_{c_k})}{1+\beta(1-p_k^{ll}-p_k^{hh})}, k \in \{P, S\},$$

Condition (40) is stricter than (41). Otherwise, the opposite is true. PS can be implemented if and only if Conditions (40) and (42) can be fulfilled simultaneously:

$$\tilde{c_e}^{P\alpha} \geq \tilde{c_e}^{S\alpha} \iff$$
$$\frac{\Delta_{c_P}+\alpha L}{1+\beta(1-p_P^{ll}-p_P^{hh})} \geq \frac{\Delta_{c_S}+\alpha L}{1+\beta(1-p_S^{ll}-p_S^{hh})}.$$

As assumed, the specialist has a stronger treatment advantage for h-types: $p_P^{hh} - p_S^{hh} > p_S^{ll} - p_P^{ll}$. Thus follows step 1.:

$$\tilde{\Delta}_{c_P} > \tilde{\Delta}^{\alpha}_{c_P} := \frac{(\Delta_{c_S}+\alpha L)[1+\beta(1-p_P^{ll}-p_P^{hh})]}{1+\beta(1-p_S^{ll}-p_S^{hh})} - \alpha L$$

Then, if effort costs are low enough, i.e. $c_e \leq \tilde{c_e}^{P\alpha}$, PS^{11} can be implemented. Otherwise, PS^{00} can be implemented.

Figure 10 illustrates the result. Compared to the case without altruism, the area in which PS (especially PS^{11}) can be implemented grows. For $\Delta_{c_P} < \tilde{\Delta}^{\alpha}_{c_P}$, only blind treatment paths or SP^{11} can be implemented second-best optimally. In the special case $\alpha = 1$, the first-best can be always be implemented by a flat fee $\gamma_P = \gamma_S$.

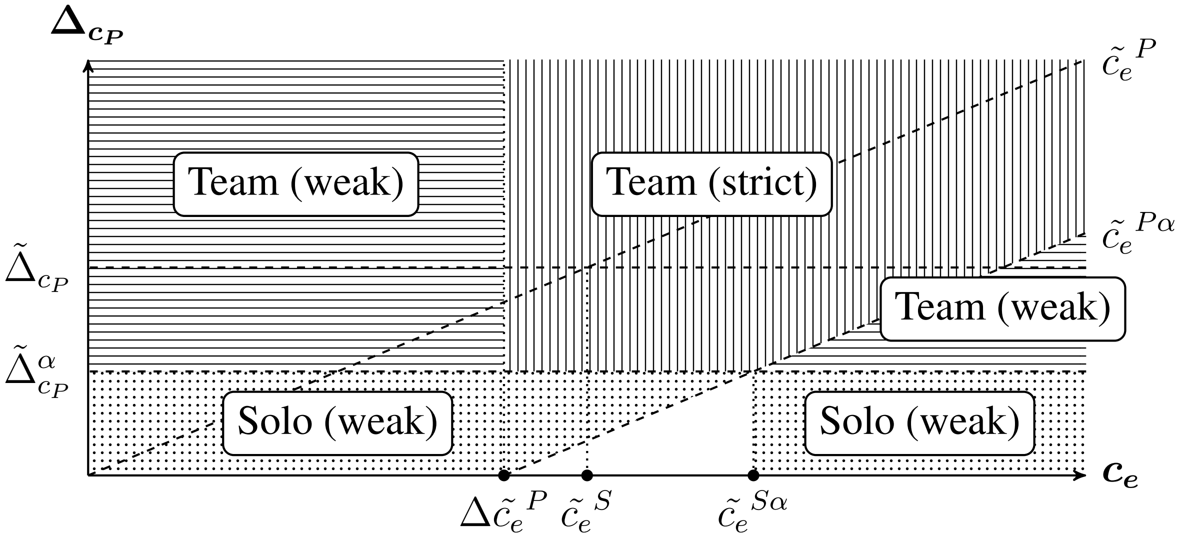


Figure 11: Second-best optimal organization of care for altruistic physicians (vertical lines: strict team dominance, horizontal lines: weak team dominance, dots: weak solo practice dominance, no pattern: no general dominance relation)

It remains to be proven that in the solo practices effort is provided only if $c_e \leq \tilde{c_e}^{P\alpha}$ for any treatment path. This is done in Appendix B.3. In particular, PS^{11} cannot be implemented for

$$c_e > \Delta\tilde{c_e}^P := \frac{\Delta_p \alpha L \beta}{1+\beta(1-p_P^{ll}-p_P^{hh})} = \tilde{c_e}^{P\alpha} - \tilde{c_e}^P.$$

It follows Proposition 10. Figure 11 visualizes the proposition.

Proposition 10. *If physicians are altruistic, the set of parameters for which the team weakly dominates the solo practices in terms of social welfare is increased. Furthermore, if*

$$c_e \in (\Delta\tilde{c_e}^P, \tilde{c_e}^{P\alpha}),$$

the team strictly dominates the solo practices given that Conditions (4) to (6) hold strictly.

This result can be interpreted in the following way. With increasing altruism, the team internalizes the patient's health loss. As there is no coordination problem between the physicians, the outcome approaches the first-best. In contrast, in the solo practices the coordination problem remains.

B.3. Proof for Proposition 10

Proof. I will show the proof for the specialist. The proof for the PCP is analogous. For each possible treatment path (excluding paths with hospital treatment) it needs to hold that effort provision is impossible to implement for $c_e > \tilde{c_e}^{P\alpha}$ given that $\Delta_{c_P} \geq \tilde{\Delta}^{\alpha}_{c_P}$.

$PS : s_P = (T,R)^{e^l_P e^h_P}$

In order to implement PS, the following conditions need to hold.

$e^l_P = 1:$

$$\begin{aligned}
U^{\alpha}_S[(R,T)^{e^l_S 1}] &\geq U^{\alpha}_S[(R,T)^{e^l_S 0}] \iff \\
u^{x\alpha}_S[(R,T)^{e^l_S 1}] &\geq u^{x\alpha}_S[(R,T)^{e^l_S 0}] \forall x \in \mathcal{X} \iff \\
\gamma_S &\leq c^h_S + \alpha L - \frac{c_e[1+\beta(1-p^{ll}_P - p^{hh}_S - \Delta_p)]}{\Delta_p \beta} \\
U^{\alpha}_S[(R,T)^{e^l_S 1}] &\geq U^{\alpha}_S[(R,R)^{e^l_S e^h_S}] \iff \\
u^{x\alpha}_S[(R,T)^{e^l_S 1}] &\geq u^{x\alpha}_S[(R,R)^{e^l_S e^h_S}] \forall x \in \mathcal{X} \iff \\
\gamma_S &\geq c^h_S - c_e - \frac{\alpha L \beta (p^{hh}_P - p^{hh}_S)}{1+\beta(1-p^{ll}_P - p^{hh}_P)}
\end{aligned}$$

$e^l_P = 0:$

$$\begin{aligned}
U^{\alpha}_S[(R,T)^{e^l_S 1}] &\geq U^{\alpha}_S[(R,T)^{e^l_S 0}] \iff \\
u^{x\alpha}_S[(R,T)^{e^l_S 1}] &\geq u^{x\alpha}_S[(R,T)^{e^l_S 0}] \forall x \in \mathcal{X} \iff \\
\gamma_S &\leq c^h_S + \alpha L - \frac{c_e[1+\beta(1-p^{ll}_P - p^{hh}_S)]}{\Delta_p \beta} \\
U^{\alpha}_S[(R,T)^{e^l_S 1}] &\geq U^{\alpha}_S[(R,R)^{e^l_S e^h_S}] \iff \\
u^{x\alpha}_S[(R,T)^{e^l_S 1}] &\geq u^{x\alpha}_S[(R,R)^{e^l_S e^h_S}] \forall x \in \mathcal{X} \iff \\
\gamma_S &\geq c^h_S - c_e - \frac{\alpha L \beta (p^{hh}_P - p^{hh}_S + \Delta_p)}{1+\beta(1-p^{ll}_P - p^{hh}_P)}
\end{aligned}$$

Either set of conditions can be fulfilled if and only if

$$c_e \leq \frac{\Delta_p \alpha L \beta}{1+\beta(1-p^{ll}_P - p^{hh}_P)} < \tilde{c_e}^{P\alpha}.$$

$SS/MS/SM$:

In all three paths, the specialist treats all received patients indefinitely. Effort is provided if and only if

$$\begin{aligned}
U_S^\alpha[(T,T)^{11}] &\geq U_S^\alpha[(T,T)^{01}], U_S^\alpha[(T,T)^{10}], U_S^\alpha[(T,T)^{00}] \iff \\
u_S^{ls\alpha}[(T,T)^{11}] &\geq u_S^{ls\alpha}[(T,T)^{01}], u_S^{ls\alpha}[(T,T)^{10}], u_S^{ls\alpha}[(T,T)^{00}] \iff \\
u_S^{hs\alpha}[(T,T)^{11}] &\geq u_S^{hs\alpha}[(T,T)^{01}], u_S^{hs\alpha}[(T,T)^{10}], u_S^{hs\alpha}[(T,T)^{00}] \iff \\
c_e &\leq \tilde{c_e}^{S\alpha}.
\end{aligned}$$

Furthermore,

$$\tilde{c_e}^{S\alpha} \leq \tilde{c_e}^{P\alpha} \iff \Delta_{c_P} \geq \tilde{\Delta}_{c_P}^\alpha.$$

$SP : s_P = (R,T)^{e_P^l e_P^h}$

$e_P^h = 1 :$

$$\begin{aligned}
U_S^\alpha[(T,R)^{1e_S^h}] &\geq U_S^\alpha[(T,R)^{0e_S^h}] \iff \\
u_S^{x\alpha}[(T,R)^{1e_S^h}] &\geq u_S^{x\alpha}[(T,R)^{0e_S^h}] \forall x \in \mathcal{X} \iff \\
\gamma_S &\geq c_S^l + c_e - \alpha L + \frac{c_e[1 + \beta(1 - p_S^{ll} - p_P^{hh} + \Delta_p)]}{\Delta_p \beta} \\
U_S^\alpha[(T,R)^{1e_S^h}] &\geq U_S^\alpha[(T,T)^{11}] \iff \\
u_S^{x\alpha}[(T,R)^{1e_S^h}] &\geq u_S^{x\alpha}[(T,T)^{11}] \forall x \in \mathcal{X} \iff \\
\gamma_S &\leq c_S^l + c_e - \alpha L + \frac{(\alpha L + c_S^h - c_S^l)1 + \beta(1 - p_S^{ll} - p_P^{hh})}{1 + \beta(1 - p_S^{ll} - p_S^{hh})}
\end{aligned}$$

$e_P^h = 0 :$

$$\begin{aligned}
U_S^\alpha[(T,R)^{1e_S^h}] &\geq U_S^\alpha[(T,R)^{0e_S^h}] \iff \\
u_S^{x\alpha}[(T,R)^{1e_S^h}] &\geq u_S^{x\alpha}[(T,R)^{0e_S^h}] \forall x \in \mathcal{X} \iff \\
\gamma_S &\geq c_S^l - \alpha L + \frac{c_e[1 + \beta(1 - p_S^{ll} - p_P^{hh} + \Delta_p)]}{\Delta_p \beta} \\
U_S^\alpha[(T,R)^{1e_S^h}] &\geq U_S^\alpha[(T,T)^{11}] \iff \\
u_S^{x\alpha}[(T,R)^{1e_S^h}] &\geq u_S^{x\alpha}[(T,T)^{11}] \forall x \in \mathcal{X} \iff \\
\gamma_S &\leq c_S^l + c_e - \alpha L + \frac{(\alpha L + c_S^h - c_S^l)1 + \beta(1 - p_S^{ll} - p_P^{hh} - \Delta_p)}{1 + \beta(1 - p_S^{ll} - p_S^{hh})}
\end{aligned}$$

Either set of conditions can be fulfilled if and only if

$$c_e \leq \tilde{c_e}^{S\alpha}.$$

$PM : s_P = (T, T)^{e_P^l e_P^h}$

$e_P^l = e_P^h = 1 :$

$$\begin{aligned}
U_S^\alpha[(R,T)^{e_S^l 1}] \geq U_S^\alpha[(R,T)^{e_S^l 0}] &\iff \\
u_S^{hs\alpha}[(R,T)^{e_S^l 1}] \geq u_S^{hs\alpha}[(R,T)^{e_S^l 1}] &\iff \\
\gamma_S \leq c_S^h + \frac{\alpha L[1+\beta(\Delta_p - p_P^{hh})]}{1+\beta(1-p_P^{ll}-p_P^{hh})} &- \frac{c_e(1-\beta p_S^{hh})}{\Delta_p \beta} \\
U_S^\alpha[(R,T)^{e_S^l 1}] \geq U_S^\alpha[(R,R)^{e_S^l e_S^h}] &\iff \\
u_S^{hs\alpha}[(R,T)^{e_S^l 1}] \geq u_S^{hs\alpha}[(R,R)^{e_S^l e_S^h}] &\iff \\
u_S^{ls\alpha}[(R,T)^{e_S^l 1}] \geq u_S^{ls\alpha}[(R,R)^{e_S^l e_S^h}] &\iff \\
\gamma_S \geq c_S^h + c_e - \frac{\alpha L\beta(p_P^{hh} - p_S^{hh})}{1+\beta(1-p_P^{ll}-p_P^{hh})}
\end{aligned}$$

$e_P^l = e_P^h = 0 :$

$$\begin{aligned}
U_S^\alpha[(R,T)^{e_S^l 1}] \geq U_S^\alpha[(R,T)^{e_S^l 0}] &\iff \\
u_S^{hs\alpha}[(R,T)^{e_S^l 1}] \geq u_S^{hs\alpha}[(R,T)^{e_S^l 1}] &\iff \\
\gamma_S \leq c_S^h + \frac{\alpha L[1+\beta(p_P^{hh})]}{1+\beta(1-p_P^{ll}-p_P^{hh})} &- \frac{c_e(1-\beta p_S^{hh})}{\Delta_p \beta} \\
U_S^\alpha[(R,T)^{e_S^l 1}] \geq U_S^\alpha[(R,R)^{e_S^l e_S^h}] &\iff \\
u_S^{hs\alpha}[(R,T)^{e_S^l 1}] \geq u_S^{hs\alpha}[(R,R)^{e_S^l e_S^h}] &\iff \\
u_S^{ls\alpha}[(R,T)^{e_S^l 1}] \geq u_S^{ls\alpha}[(R,R)^{e_S^l e_S^h}] &\iff \\
\gamma_S \geq c_S^h + c_e - \frac{\alpha L\beta(\Delta_p + p_P^{hh} - p_S^{hh})}{1+\beta(1-p_P^{ll}-p_P^{hh})}
\end{aligned}$$

Either set of conditions can be fulfilled if and only if

$$c_e \leq \frac{\Delta_p \alpha L \beta}{1+\beta(1-p_P^{ll}-p_P^{hh})} < \tilde{c_e}^{P\alpha}.$$

$MP : s_P = (T,T)^{e^l_P e^h_P}$

$e^l_P = e^h_P = 1:$

$$\begin{aligned}
U^\alpha_S[(T,R)^{1e^h_S}] \geq U^\alpha_S[(T,R)^{0e^h_S}] &\iff \\
u^{l_S\alpha}_S[(T,R)^{1e^h_S}] \geq u^{l_S\alpha}_S[(T,R)^{0e^h_S}] &\iff \\
\gamma_S \geq c^l_S + \frac{c_e(1-\beta p^{ll}_S)}{\Delta_p \beta} &- \frac{\alpha L[1-\beta(p^{ll}_P + \Delta_p)]}{1+\beta(1-p^{ll}_P - p^{hh}_P)} \\
U^\alpha_S[(T,R)^{1e^h_S}] \geq U^\alpha_S[(T,T)^{11}] &\iff \\
u^{l_S\alpha}_S[(T,R)^{1e^h_S}] \geq u^{l_S\alpha}_S[(T,T)^{11}] &\iff \\
u^{h_S\alpha}_S[(T,R)^{1e^h_S}] \geq u^{h_S\alpha}_S[(T,T)^{11}] &\iff \\
\gamma_S \leq \frac{(c^h_S + c_e)[1+\beta(1-p^{hh}_P - p^{ll}_P)] + \alpha L\beta(1+\Delta_p - p^{hh}_P)}{1+\beta(1-p^{ll}_P - p^{hh}_P)} & \\
- \frac{\beta(1+\Delta_p - p^{hh}_S)(c^h_S - c^l_S + \alpha L)}{1+\beta(1-p^{ll}_S - p^{hh}_S)} &
\end{aligned}$$

$e^l_P = e^h_P = 0:$

$$\begin{aligned}
U^\alpha_S[(T,R)^{1e^h_S}] \geq U^\alpha_S[(T,R)^{0e^h_S}] &\iff \\
u^{l_S\alpha}_S[(T,R)^{1e^h_S}] \geq u^{l_S\alpha}_S[(T,R)^{0e^h_S}] &\iff \\
\gamma_S \geq c^l_S + \frac{c_e(1-\beta p^{ll}_S)}{\Delta_p \beta} &- \frac{\alpha L[1-\beta(p^{ll}_P)]}{1+\beta(1-p^{ll}_P - p^{hh}_P)} \\
U^\alpha_S[(T,R)^{1e^h_S}] \geq U^\alpha_S[(T,T)^{11}] &\iff \\
u^{l_S\alpha}_S[(T,R)^{1e^h_S}] \geq u^{l_S\alpha}_S[(T,T)^{11}] &\iff \\
u^{h_S\alpha}_S[(T,R)^{1e^h_S}] \geq u^{h_S\alpha}_S[(T,T)^{11}] &\iff \\
\gamma_S \leq \frac{(c^h_S + c_e)[1+\beta(1-p^{hh}_P - p^{ll}_P)] + \alpha L\beta(1 - p^{hh}_P)}{1+\beta(1-p^{ll}_P - p^{hh}_P)} & \\
- \frac{\beta(1+\Delta_p - p^{hh}_S)(c^h_S - c^l_S + \alpha L)}{1+\beta(1-p^{ll}_S - p^{hh}_S)} &
\end{aligned}$$

Either set of conditions can be fulfilled if and only if

$$c_e \leq \tilde{c_e}^{S\alpha}.$$

□

References

Allard, Marie, Jelovac, Izabela, and Léger, Pierre Thomas (2011). "Treatment and referral decisions under different physician payment mechanisms". In: *Journal of Health Economics* 30.5, pp. 880–893. DOI: 10.1016/j.jhealeco.2011.05.016.

Blümel, Miriam and Busse, Reinhard (2015). *The German Health Care System*. URL: https://international.commonwealthfund.org/countries/germany/.

Bodenheimer, Thomas, Chen, Ellen, and Bennett, Heather D. (2009). "Confronting The Growing Burden Of Chronic Disease: Can The U.S. Health Care Workforce Do The Job?" In: *Health Affairs* 28.1, pp. 64–74. DOI: 10.1377/hlthaff.28.1.64.

Bruin, Simone R. de, Heijink, Richard, Lemmens, Lidwien C., Struijs, Jeroen N., and Baan, Caroline A. (2001). "Impact of disease management programs on healthcare expenditures for patients with diabetes, depression, heart failure or chronic obstructive pulmonary disease: A systematic review of the literature". In: *Health Policy* 101.2, pp. 105–121. DOI: 10.1016/j.healthpol.2011.03.006.

Buttorff, Christine (2017). *Multiple chronic conditions in the United States*. Santa Monica, CA: RAND.

Cheung, Winson Y., Neville, Bridget A., Cameron, Danielle B., Cook, E. Francis, and Earle, Craig C. (2009). "Comparisons of Patient and Physician Expectations for Cancer Survivorship Care". In: *Journal of Clinical Oncology* 27.15, pp. 2489–2495. DOI: 10.1200/jco.2008.20.3232.

Dulleck, Uwe and Kerschbamer, Rudolf (2006). "On doctors, mechanics, and computer specialists: The economics of credence goods". In: *Journal of Economic literature* 44.1, pp. 5–42. DOI: 10.1257/002205106776162717.

Dusheiko, Mark, Gravelle, Hugh, Martin, Stephen, Rice, Nigel, and Smith, Peter C. (2011). "Does better disease management in primary care reduce hospital costs? Evidence from English primary care". In: *Journal of Health Economics* 30.5, pp. 919–932. DOI: 10.1016/j.jhealeco.2011.08.001.

Eggleston, K. (2000). "Risk Selection and Optimal Health Insurance-Provider Payment Systems". In: *Journal of Risk and Insurance* 67, pp. 173–196.

Ellis, Randall P. and McGuire, Thomas G. (1986). "Provider behavior under prospective reimbursement: Cost sharing and supply". In: *Journal of Health Economics* 5.2, pp. 129–151. URL: http://EconPapers.repec.org/RePEc:eee:jhecon:v:5:y:1986:i:2:p:129-151.

Filar, Jerzy and Vrieze, Koos (1996). *Competitive Markov Decision Processes*. Springer New York. DOI: 10.1007/978-1-4612-4054-9.

Garcia-Mariñoso, Begoña and Jelovac, Izabela (2003). "GPs' payment contracts and their referral practice". In: *Journal of Health Economics* 22.4, pp. 617–635. DOI: 10.1016/S0167-6296(03)00008-0.

Gask, Linda (2005). “Role of specialists in common chronic diseases”. In: *BMJ* 330.7492, pp. 651–653. DOI: 10.1136/bmj.330.7492.651.

Grassi, Simona and Ma, Ching-to A. (2016). “Information acquisition, referral, and organization”. In: *The RAND Journal of Economics* 47.4, pp. 935–960. DOI: 10.1111/1756-2171.12160.

Griebenow, Malte and Kifmann, Mathias (2022). “Diagnostics and Treatment: On the Division of Labor between Primary Care Physicians and Specialists”. In: *Journal of Institutional and Theoretical Economics* 178.2, p. 191. DOI: 10.1628/jite-2022-0006.

Hafsteinsdottir, Elin J.G. and Siciliani, Luigi (2010). “DRG Prospective Payment Systems: Refine or not Refine?” In: *Health Economics* 19, pp. 1226–1239. DOI: 10.1002/hec.1547.

Hennessy, David A. (2008). “Prevention and cure efforts both substitute and complement”. In: *Health Economics* 17.4, pp. 503–511. DOI: 10.1002/hec.1270.

Hey, John D. and Patel, Mahesh S. (1983). “Prevention and cure? : Or: Is an ounce of prevention worth a pound of cure?” In: *Journal of Health Economics* 2.2, pp. 119–138. DOI: 10.1016/0167-6296(83)90002-4.

Ho, E.T.L., Tee, W., and Ng, H.S.H. (2013). “A right-siting model for chronic disease: clinical outcomes of patients right-sited under the SingHealth Delivering On Target (DOT) programme”. In: *International Journal of Integrated Care* 13.8. DOI: 10.5334/ijic.1489.

Iversen, Tor and Lurås, Hilde (2000). “The effect of capitation on GPs’ referral decisions”. In: *Health Economics* 9.3, pp. 199–210. DOI: 10.1002/(SICI)1099-1050(200004)9:3<199::AID-HEC514>3.0.CO;2-2.

Jelovac, Izabela and Macho-Stadler, Inés (2002). “Comparing organizational structures in health services”. In: *Journal of Economic Behavior & Organization* 49.4, pp. 501–522. DOI: 10.1016/s0167-2681(02)00008-2.

Kash, Bita and Tan, Debra (2016). “Physician Group Practice Trends: A Comprehensive Review”. In: *J Hosp Med Manage* 2.1.

Lemieux-Charles, Louise and McGuire, Wendy L. (2006). “What Do We Know about Health Care Team Effectiveness? A Review of the Literature”. In: *Medical Care Research and Review* 63.3, pp. 263–300. DOI: 10.1177/1077558706287003.

Li, R., Zhang, P., Barker, L. E., Chowdhury, F. M., and Zhang, X. (2010). “Cost-Effectiveness of Interventions to Prevent and Control Diabetes Mellitus: A Systematic Review”. In: *Diabetes Care* 33.8, pp. 1872–1894. DOI: 10.2337/dc10-0843.

Lim, Jeremy F. Y., Tan, Darren M. H., and Lee, Andrew L. (2008). “Consequences of right siting of endocrinology patients–a financial and caseload simulation.” In: *Annals of the Academy of Medicine, Singapore* 37 (2), pp. 109–113. ppublish.

Malcomson, James M. (2005). “Supplier discretion over provision: theory and an application to medical care.” In: *The RAND journal of economics* 36 (2), pp. 412–432.

Mehrotra, Ateev, Forrest, Christopher B., and Lin, CAROLINE Y. (2011). “Dropping the Baton: Specialty Referrals in the United States”. In: *Milbank Quarterly* 89.1, pp. 39–68. DOI: 10.1111/j.1468-0009.2011.00619.x.

Meiklejohn, Judith A., Mimery, Alexander, Martin, Jennifer H., Bailie, Ross, Garvey, Gail, Walpole, Euan T., Adams, Jon, Williamson, Daniel, and Valery, Patricia C. (2016). “The role of the GP in follow-up cancer care: a systematic literature review”. In: *Journal of Cancer Survivorship* 10.6, pp. 990–1011. DOI: 10.1007/s11764-016-0545-4.

Pascucci, Domenico, Sassano, Michele, Nurchis, Mario Cesare, Cicconi, Michela, Acampora, Anna, Park, Daejun, Morano, Carmen, and Damiani, Gianfranco (2020). “Impact of interprofessional collaboration on chronic disease management: findings from a systematic review of clinical trial and meta-analysis”. In: *Health Policy*. DOI: 10.1016/j.healthpol.2020.12.006.

Sarma, Sisira, Mehta, Nirav, Devlin, Rose Anne, Kpelitse, Koffi Ahoto, and Li, Lihua (2018). “Family physician remuneration schemes and specialist referrals: Quasi-experimental evidence from Ontario, Canada”. In: *Health Economics* 27.10, pp. 1533–1549. DOI: 10.1002/hec.3783.

Shumsky, Robert A. and Pinker, Edieal J. (2003). “Gatekeepers and Referrals in Services”. In: *Management Science* 49.7, pp. 839–856. DOI: 10.1287/mnsc.49.7.839.16387.

Song, Zirui, Sequist, Thomas D., and Barnett, Michael L. (2014). “Patient Referrals”. In: *JAMA* 312.6, p. 597. DOI: 10.1001/jama.2014.7878.

Wee, Shiou Liang, Tan, Caren G. P., Ng, Hilda S. H., Su, Scott, Tai, Virginia U. M., Flores, John V. P. G., and Khoo, Daphne H. C. (2008). “Diabetes outcomes in specialist and general practitioner settings in Singapore: challenges of right-siting.” In: *Annals of the Academy of Medicine, Singapore* 37 (11), pp. 929–935. ppublish.