CCHE CCES

Canadian Centre for Health Economics
Centre canadien en économie de la santé

Working Paper Series
Document de travail de la série

Physician payment contracts in the presence of moral hazard and adverse selection: theory and application to Ontario

Jasmin Kantarevic, Boris Kralj

Working Paper No: 150001

www.canadiancentreforhealtheconomics.ca

January 15, 2015

Canadian Centre for Health Economics
Centre canadien en économie de la santé
155 College Street
Toronto, Ontario

CCHE/CCES Working Paper No. 150001
January 15, 2015

Physician Payment Contracts In The Presence Of Moral Hazard And Adverse Selection: Theory And Application To Ontario[1]

Jasmin Kantarevic[a,b,c] and Boris Kralj[a]

[a] *Ontario Medical Association*
[b] *University of Toronto*
[c] *Institute for Labour Studies*

Abstract

We develop a simple principal-agent model with moral hazard and adverse selection to provide a unified framework for understanding some of the most salient features of the recent physician payment reform in Ontario and its impact on physician behaviour. These features include: (1) physicians can choose a payment contract from a menu that includes an enhanced fee-for-service contract and a blended capitation contract; (2) the capitation rate is higher and the cost-reimbursement rate is lower in the blended capitation contract; (3) physicians sort selectively into the contracts based on their initial productivity; (4) all else equal, physicians in the blended capitation model provide fewer services than physicians in the enhanced fee-for-service model.

JEL Classification: I10; I12; I18

Keywords: physician remuneration, moral hazard, adverse selection, Ontario

[1]The views expressed in this paper are strictly those of the authors. No official endorsement by the Ontario Medical Association is intended or should be inferred.

1 Introduction

Over the last decade, Ontario has launched a major primary health care reform. One of the unique features of this reform is that it introduced a menu of payment contracts, rather than a single contract, in which physicians can choose to participate. In contrast to the traditional fee-for-service (FFS) contract, under which physicians receive a fee for each service they provide, the new contracts blend prospective (per-patient) and retrospective (per-service) payments to varying degrees. Understanding the impact of this reform has been a focus of much recent empirical research (see e.g. [11,14,15,20,28]). However, this literature still lacks a unified framework that can explain seemingly unrelated findings and provide a rationale for the main features of the reform, such as the existence of a menu of contracts and the specific blend of prospective and retrospective elements in each contract.

In this paper, we take the first few steps needed to fill this gap. In the first part, we document some of the most salient stylized facts about the new payment contracts. These include the menu and shape of contracts, which are defined by the institutional framework, and the selection and incentive effects of contracts, which we estimate using the difference-in-difference methodology with the propensity score matching. In the second part, we then develop a simple principal-agent model with moral hazard and adverse selection in which these stylized facts arise as a solution to the problem of designing optimal payment contracts when neither physician type nor physician action is contractible.

The specific contribution of this paper is therefore to document a set of stylized facts about the physician payment reform in Ontario and to develop a relatively standard economic model to interpret this reform. More generally, the paper contributes to our understanding of how physicians respond to payment incentives. Traditionally, interest has centered on comparing the two main methods of payment - FFS and capitation - in terms of their impact on the access, quality, and cost of health care (see e.g. [18,24,29,31]). The conclusions from this literature are mostly based on a principal-agent model with moral hazard in which physician actions cannot be observed or verified (see e.g. [8,9,21,22]). More recent studies have also considered the adverse selection framework that focuses on unobserved heterogeneity among physicians such as ability, altruism, and efficiency (see e.g. [5,23]). Lastly, there is a small but growing literature, to which our study contributes, that combines both moral hazard and adverse selection to study the design of optimal contracts and their impact on physician behavior (see e.g. [1,2,13]).

Our study most closely parallels the analysis of the UK fundholding scheme by Jack (2005) and Dusheiko et al. (2006). Jack develops a theoretical model with moral hazard and adverse selection and concludes that the optimal menu of contracts includes a set of cost-sharing rules resembling the fundholding contract, while Dusheiko et al. use the difference-in-difference methodology to study the incentive and selection effects of the fundholding scheme on provider behaviour. We use similar analytical tools to study the reform in Ontario, which testifies to the value of these tools in understanding the design of payment reforms and their impact on physician behaviour in diverse health care environments[2].

The rest of the paper proceeds as follows. In the next section, we describe the payment contracts available to primary care physicians in Ontario and document selected stylized facts that we study in this paper. In Section 3, we set up the model and use it to explain the stylized facts. We conclude in Section 4.

[2]Allard et al. (2011, 2014) study how heterogeneous physicians choose between payment mechanisms and the impact of this choice on their treatment and referral behaviour. We also study the selection and incentive effects of the similar type of contracts, although on different outcomes, but we additionally examine the design of the optimal menu of contracts and their shape, which is taken as given in these two studies.

2 Stylized Facts about Physician Payment Contracts in Ontario

The physician payment reform in Ontario over the last decade aimed to address two main problems. The first was the chronic shortage of family physicians in the province, leaving a significant number of patients with no access to a regular family doctor. This problem was caused in part by the declining real incomes of physicians during the 1990s, and the reform aimed to reverse that trend, relative to both family physicians in other jurisdictions and to specialists in the province. The second problem was the practice style of family physicians. Specifically, almost all physicians practiced in the traditional FFS model, which was often criticized because of its excessive focus on the volume-based acute care and the lack of incentives to form teams with other physicians and health care providers (see e.g. [18,24]).

The reform addressed the two problems simultaneously by linking significant financial incentives to physician participation in the new models of care. These models centered on patient enrolment, comprehensive and preventive care, chronic disease management, after-hours access, group practices with interdisciplinary teams, and increased reliance on the electronic medical records.

As mentioned earlier, understanding the impact of this reform has been a focus of much recent empirical research. We contribute to this literature by developing a unified framework to study selected stylized facts of the reform, which we describe in the following subsections.

2.1 Menu and Shape of Payment Contracts

The first stylised fact we study is that the primary care reform introduced two main types of contracts rather than a single contract (see Table 1). The harmonised models, such as the Family Health Network and the Family Health Organization, are blended capitation models (capitation models hereafter). The non-harmonised models, such as the Family Health Groups and the Comprehensive Care Model, are enhanced *fee-for-service* models (FFS models hereafter). Currently, about two thirds of primary care physicians in Ontario choose to participate in these two payment models.

The second stylized fact relates to the shape of payment contracts in FFS and capitation models. In each model, the contract consists of two main elements: a payment per enrolled patient (the capitation rate) and a payment per service (the cost-reimbursement rate). Specifically, in the FFS models, the capitation rate consists of a comprehensive care management (CCM) fee, equal in the annual value to about one regular office visit, while in the capitation models, the capitation rate includes both the CCM fee and an age-sex adjusted capitation payment, equal in the annual value to about five regular office visits (see Table 2). On the other hand, in the capitation models, physicians receive 15 percent of the FFS value of services (for core services provided in the capitation basket), while physicians in the FFS models receive 100 percent of the FFS value services provided. Therefore, the capitation rate is higher in the capitation models, while the cost reimbursement rate is higher in the FFS models. Alternatively speaking, the prospective payment is higher and the retrospective payment is lower in the capitation models compared to the FFS models.

These two stylized facts are similar to the fundholding scheme that was available to general practitioners in the UK from 1991 to 1999. In this scheme, physician practices were also able to choose from a menu of two contracts. In the fundholding contract, physician practices were given a budget from which to finance selected non-emergency hospital-delivered secondary care and unused budget could be used to purchase new equipment or other services (Dusheiko et al., 2006). The other contract was the status quo, in which physician practices neither bore the costs of secondary care directly nor appropriated any savings. Further, the shape of the contracts was also similar to Ontario in the sense that the prospective payment (i.e., the budget) was larger and the cost-reimbursement rate lower for the fundholding contract than for the status quo contract. Therefore, in both the Ontario and UK fundholding scheme, there is a negative relationship between the prospective and retrospective payment elements across different contracts.

2.2 Selection and Incentive Effects of Payment Contracts

The last two stylized facts relate to the impact of payment contracts on physician behavior. Specifically, differences between the contracts may impact both the physician's choice of which contract to join (the selection effect) and his/her choice on how to practice (the incentive effect).

Discerning these two effects is an empirical issue that has been discussed to some extent in the literature (see e.g. [15,20]). In this paper, we present new evidence on this issue using a unique data set on a cohort of all 3,641 physicians who participated in the Family Health Groups (the enhanced FFS model) in fiscal 2006/7. Of this cohort, 1,563 physicians (43 percent) remained in the same model by fiscal 2013/14, while 2,078 physicians (57 percent) switched to the Family Health Organizations (the blended capitation model). The aspect of physician practice we study is volume-based: the number of services and visits per day. Admittedly, this is not the only or even the most important aspect of physician practice. Nevertheless, it is still an important aspect in a health care system such as Ontario in the late 1990s, which was characterized by a shortage of physicians, long wait times, and a significant number of patients with no regular family doctor. In addition, the quantitative aspect of physician practice tends to be accurately measured through the administrative claims databases, such as the Ontario Health Insurance Plan (OHIP) claims database that we use in this study.

To measure the selection effect, we compared the outcomes of interest - the number of services and visits per day - between physicians who remained in the FFS model throughout the sample period (the stayers) and physicians who switched to the capitation model at some point during the sample period (the switchers). This comparison is conducted as of fiscal 2006/7, when both stayers and switchers practiced in the FFS model and before the capitation model was introduced in 2007. The results, presented in Table 3, clearly indicate that the stayers provided significantly more services and visits than the switchers. This evidence of selective sorting across models is further confirmed by the fact that the expected income gain for the switchers was about C$45,000, while it was about -C$33,000 for the stayers.[3] The table also indicates that the switchers were on average younger than the stayers, but there were no significant differences between the two groups in terms of their gender composition or geographical distribution. This analysis generates the third stylized fact about the contracts: physicians with lower volumes tend to sort into the capitation models, while physicians with higher volumes tend to sort into the FFS models. This result - that higher-productivity individuals sort into jobs in which the marginal return to their productivity is higher - is well documented in the labor economics literature (see e.g. [17,30]).

The last stylized fact relates to the incentive effect of contracts, arising primarily because of the differential cost-reimbursement rate between the contracts. Specifically, all else equal, we expect that physicians in the FFS models tend to provide more services and visits per day than physicians in the capitation models, because of the higher cost-reimbursement rate.

The challenge of evaluating this proposition empirically is controlling for the selection effect that we documented earlier. Given that we observe this cohort of physicians both before (fiscal 2006/7) and after (fiscal 2013/14) the capitation model was introduced, one plausible empirical strategy is the difference-in-differences model, which allows us to control for time-invariant differences between

[3]The expected income gain is calculated by using the current profile of services and patients in the Family Health Groups and applying the payment rules from the Family Health Organization model.

the groups and the common trend in outcomes. We strengthen this strategy by combining it with the propensity score matching, which is expected to further improve comparability between the two groups (see e.g. [3,6,19,25,26,27]). Conceptually, this approach involves two steps. In the first step, we estimate the propensity score for each physician to switch to the capitation model using outcomes and covariates listed in Table 3. The distribution of estimated propensity scores for the stayers and switchers is shown in Figure 1. This figure shows that the empirical support of the two distributions is very similar, although, as expected, the switchers have a higher average probability of joining the capitation model than the stayers.

In the second step, we then compare the difference in outcomes between each switcher and a matched group of stayers based on the propensity scores. The results are presented in the first row of Table 4 and indicate that physicians in the capitation model provide significantly fewer services and visits per day than if they practiced in the FFS model. These results, performed in Stata 12.0 using the psmatch2 command, come from the local linear regression model with the bi-weight kernel, the bandwidth of 0.1, and the common support restriction. The remainder of rows in Table 4 shows that the results are quite robust to using the alternative estimators (nearest neighbor, kernel), the alternative kernel functions (normal, uniform, Epanechnikov, Tricube), alternative bandwidth values (0.05 and 0.20), and the alternative trimming levels (0 and 10 percent). Therefore, this analysis generates the fourth stylized fact about the contracts: all else equal, physicians provide more visits and services in the FFS model than in the capitation model. Again, this finding has been well documented in the literature (see e.g. [12]). Of particular interest here are the results for the UK fundholding scheme by Dusheiko et al. (2006), given its similarity to Ontario. Dusheiko et al. discuss and find empirical support for both the selection and incentive effects of the fundholding contract. They assume that physicians are heterogeneous with respect to a taste parameter, and argue that this parameter determines both the admission rates for secondary care (because of the perceived gross benefit of admission) and the propensity to choose fund-holder status, because of a fixed cost association with such status (e.g. direct transaction cost). Empirically, they find that fundholding incentives reduced elective admission rates by 3% and accounted for 57% of the difference between fund-holder and non-fund-holder elective admissions, with 43% being a selection effect. This evidence is consistent with our study of Ontario in at least two ways. First, the authors convincingly demonstrate the importance of controlling for selection when estimating the behavioral consequences of payment contracts. Second, the results for the UK indicate that the volume of services (i.e., elective admission rates) is negatively related to the cost-reimbursement rate, which is consistent with our finding for Ontario.

3 Model

3.1 Set up

To explain the stylized facts documented in the previous section, we study the classical problem of a payer (e.g. government) that wishes to design a payment contract for providers (e.g. physicians) to deliver health care services (e.g., patient visits).

The number of patients treated by each provider is given exogenously and normalized to one. The number of services per patient, denoted by q, depends stochastically on the provider effort according to $q = e + \epsilon$, where $e > 0$ denotes the provider effort (e.g. clinical hours) and ϵ is a mean-zero random variable (e.g. the stochastic component of patient demand for care). All health care services are identical and the price of each service (i.e. its value to the payer) is normalized to one.

The payment contract w consists of two parts: a fixed payment a and a variable payment bq, with $0 \leq b \leq 1$. We refer to a as the capitation rate and to b as the cost-reimbursement rate. While this contract is linear[4], it is general enough to encompass the common types of payment contracts observed in practice, such as the pure FFS model (a = 0, b = 1), the pure capitation model (a > 0, b = 0), and the blended model (a > 0, 0 < b < 1).

Provider utility is given by $U = w - c(e, \theta)$, where $\theta > 0$ is the provider type. For convenience, we assume that $c(e, \theta) = 0.5\theta e^2$. This cost function trivially satisfies the Spence-Mirrlees single-crossing property, since $c_{e\theta}(e, \theta) = e > 0$. Further, we assume that $\theta \in \{\theta_L, \theta_H\}$, with $\theta_L < \theta_H$. We refer to providers with $\theta = \theta_L$ as the low-cost type and to providers with $\theta = \theta_H$ as the high-cost type.[5] The proportion of the high-cost type is equal to α and of the low-cost type to $1 - \alpha$.

The payer's utility when contracting with a provider of type i = {L,H} is equal to $V_i = q_i - w_i$. This specification assumes that the payer focuses primarily on access to physician services, rather than other goals such as the quality and cost of care. This assumption is motivated by the chronic access problem that the government of Ontario had to address in the late 1990s, as discussed in Section 2. Further, this specification implies that the patient benefit is proportional to the number of medical services, which may be violated if physicians have the ability to induce demand (e.g., [10]). However, in an environment with a chronic shortage of physicians, the concern with the physician-induced demand is unlikely to be of first-order importance.

The last three assumptions we make is that both the payer and providers are risk neutral, that all providers have an identical outside option equal to u, and that the payer's outside option is 0.

The timing of the contracting game is as follows. First, the nature determines the provider type

[4] Our focus on the linear contracts is unlikely to be too restrictive, given that with more general non-linear payment mechanisms, it is often the case that the optimal contract can be approximated by a set of linear contracts (e.g., [13])

[5] Extending the model to a continuum of types is straightforward, but is not necessary given the revelation principle and our focus on only two types of contracts.

$\theta \in \{\theta_L, \theta_H\}$, which is observed by the provider but not by the payer. Second, the payer offers a menu of two contracts (a_i, b_i) for i = {L,H}. Third, the provider either accepts one of the contracts or rejects both contracts. If the provider rejects both contracts, the game ends and the provider receives its outside option u. If the provider accepts one of the contracts, it provides effort e that cannot be observed or verified by the payer. Lastly, the nature determines ϵ, which then determines the number of services q_i and payoffs U_i and V_i.

The problem for the payer is to design a menu of contracts to maximize the benefit of the health care services provided to each patient, net of the payment to providers. Such a menu must satisfy three constraints: the contracts must be acceptable to providers; each provider must choose the contract that is designed for its type; and each contract must be compatible with the provider's optimal choice of effort. Specifically, the problem of designing an optimal payment contract can be stated as

$$MaxE[V] = \alpha[e_H - a_H - b_H e_H] + (1 - \alpha)[e_L - a_L - b_L e_L] \tag{1}$$

subject to

$$(PC_i) \qquad a_i + b_i e_i - 0.5\theta_i e_i^2 \leq u \tag{2}$$

$$(AS_i) \qquad a_i + b_i e_i - 0.5\theta_i e_i^2 \geq a_k + b_k e_k - 0.5\theta_k e_k^2 \tag{3}$$

$$(IC_i) \qquad e_i = argmax \;\; a_i + b_i \tilde{e}_i - 0.5\theta_i \tilde{e}_i^2 \tag{4}$$

for i = {L,H} and i $\neq$ k, where equations (2)-(4) denote, respectively, the participation constraints, the adverse selection (or screening) constraints, and the incentive compatibility constraints.

3.2 First-Best

When the payer can observe and verify both the provider's effort and its type, the payment contracts must satisfy only the participation constraints. Furthermore, it is easy to verify that these constraints will bind at the optimum for each provider type. Using these constraints to substitute for the provider payment in the payer's expected utility yields

$$E[V] = \alpha[e_H - 0.5\theta_H e_H^2] + (1-\alpha)[e_L - 0.5\theta_L e_L^2] - u \tag{5}$$

Since this is a concave problem, the first-order condition for the effort level is both necessary and sufficient. Therefore, the first-best level of effort is given by

$$e_i^* = 1/\theta_i \tag{6}$$

for i = {L,H}. In this environment, it is not necessary to tie the provider's pay to the number of services because the provider's effort is verifiable and both parties are risk neutral. Therefore, substituting the first-best level in the participation constraint yields the optimal capitation rate

$$a_i^* = u + 1/2\theta_i \tag{7}$$

Therefore, in the full information environment, the low-cost type provides more effort and receives higher payment than the high-cost type ($e_L^* > e_H^*, a_L^* > a_H^*$).

3.3 Second Best - Moral Hazard Only

When the payer can observe and verify the provider's type but not its effort, the relevant constraints are the participation and incentive compatibility constraints. The provider's incentive compatibility constraint for each type i is given by the first-order condition from eq. (4), which can be written as

$$e_i = b_i/\theta_i \tag{8}$$

To induce the optimal level of effort from each provider, the payer should therefore set b_i equal to one; from the participation constraint, it then follows that $a_i = u - 1/2\theta_i$. Note that this high-powered contract can induce the efficient level of effort from both types of providers. In effect, the payer 'sells the job' to the provider in exchange for a type-specific fee a_i, and the provider then fully internalizes the benefit of its effort. This efficiency result is not surprising given that the providers are risk neutral and there are no limited liability constraints.

At the optimum, $a_i \leq 0$ given that it is efficient to contract with provider type i if and only if $u \leq 1/2\theta_i$. In other words, the optimal contract entails a non-positive capitation fee as a payment from the provider to the payer. This finding seems to be inconsistent with the FFS contracts typically observed in practice. However, the non-positive capitation fee arises in this model because the payer has all the bargaining power and there are no liability constraints. To see this, note first that if the provider had all the bargaining power, the contract would maximize $E[U_i]$ subject to the payer's participation constraint $E[V_i] \geq 0$. It is easy to verify that in this case the provider would choose the efficient level of effort and the payer's participation constraint would be binding, which implies that the capitation rate would be set to zero. In this case, a single pure FFS contract for both provider types ($b_i = 1$, $a_i = 0$) would induce an efficient outcome. On the other hand, when there is a binding liability constraint of the form $a_i \geq 0$, the optimal cost-reimbursement rate for provider type i will be $1/(1+\lambda_i)$, where $\lambda_i > 0$ is the Lagrange multiplier associated with the liability constraint. Again, the contract is a pure FFS contract with no capitation rate and with the type-specific cost-reimbursement rate. However, this contract will be inefficient because the optimal cost-reimbursement rate is below unity. In both cases, however, when either the provider has all bargaining power or there is a limited liability constraint, the optimal payment contract with moral hazard is a pure FFS contract.

3.4 Second Best - Adverse Selection Only

Consider now the case where the payer can observe and verify the provider's effort but not its type. In this case, the contract must satisfy the participation and screening constraints. Furthermore, the payment contract consists of a capitation rate and a contractually specified level of effort (a_i, e_i) for each provider type. As in the full information case, it is not necessary to tie the provider's pay to the number of services because the provider's effort is verifiable and both parties are risk-neutral. Therefore, the optimal payment contract with adverse selection is a pure capitation contract.

The capitation rate and level of contractual effort for each provider type must be such that each type chooses the contract designed for its type. This qualification is important because the optimal contract in the full information environment (e_i^*,a_i^*) will, in general, fail to induce the appropriate sorting. Specifically, when offered a menu of contracts (e_i^*,a_i^*), both provider types will choose the contract designed for the high-cost type (e_H^*,a_H^*). To see this, note that for both provider types, choosing the contract designed for their type yields u. On the other hand, the high-cost type gains $u - 0.5\Delta\theta/\theta_L^2 < u$ if it chooses the contract designed for the low-cost type, and the low-cost type gains $u + 0.5\Delta\theta/\theta_H^2 > u$ if it chooses the contract designed for the high-cost type, where

$\Delta\theta = \theta_H - \theta_L > 0$. Therefore, the low-cost type has an incentive to mimic the high-cost type and the payer must design a menu of contracts different from (e_i^*,a_i^*).

By following a standard approach for solving adverse selection models (see e.g. [4,16]), we assume that only the participation constraint for the high-cost type and the screening constraint for the low-cost type are binding and then verify ex-post that the other two constraints are not binding at the optimum. By using the two binding constraints to substitute in the payer's expected utility, we get

$$E[V] = \alpha[e_H - 0.5\theta_H e_H^2] + (1-\alpha)[e_L - 0.5\theta_L e_L^2] - u - (1-\alpha)0.5\Delta\theta e_H^2 \tag{9}$$

The first three terms represent the payer's expected utility in the full information environment, while the last term represents the expected information rent for the low-cost type. Therefore, the payer's problem entails a trade-off between productive efficiency and rent extraction. The first-order necessary and sufficient conditions for the effort levels of each provider type yield

$$\tilde{e}_L = \frac{1}{\theta_L} = e_L^* \tag{10}$$

$$\tilde{e}_H = \frac{\alpha}{\alpha\theta_H + \Delta\theta(1-\alpha)} < \frac{1}{\theta_H} = e_H^* \tag{11}$$

Therefore, the optimal contract achieves 'efficiency at the top', as the low-cost provider is induced to provide the efficient level of effort, while the high-cost provider provides less than the efficient level of effort. The inefficiency result for the high-cost type reflects the optimal resolution of the conflict between efficiency and rent extraction. Furthermore, substituting the effort levels into the two binding constraints yields the type-specific capitation rates that induce each provider to choose the contract designed for their type:

$$\tilde{a}_H = u + 0.5\theta_H \tilde{e}_H^2 > 0 \tag{12}$$

$$\tilde{a}_L = \tilde{a}_H + 0.5\theta_L(\tilde{e}_L^2 - \tilde{e}_H^2) > \tilde{a}_H \tag{13}$$

where the inequality in eq. (13) follows because $\tilde{e}_L^2 - \tilde{e}_H^2 > e_L^{2*} - e_H^{2*} > 0$ since $\tilde{e}_L^2 = e_L^{2*}$ and $\tilde{e}_H^2 < e_H^{2*}$. Therefore, the capitation rate for the low-cost type is higher than that of the high-cost type. Furthermore, it is easy to show that $\tilde{a}_H < \alpha_H^*$ and $\tilde{a}_L > \alpha_L^*$.

It remains to verify that the other two constraints are not binding at the optimum. First, the participation constraint for the low-cost type is non-binding since $U_L = \tilde{a}_L - 0.5\theta_L \tilde{e}_L^2 = u + 0.5\Delta\theta e_H^2 > u$. Therefore, the low-cost type earns a strictly positive information rent. Second, the adverse selection constraint for the high-cost type implies $(\tilde{a}_H - 0.5\theta_H \tilde{e}_H^2) - (\tilde{a}_L - 0.5\theta_H \tilde{e}_L^2) = 0.5\Delta\theta(\tilde{e}_L^2 - \tilde{e}_H^2) > 0$. Therefore, the high-cost type is strictly better off when choosing the contract designed for it than by mimicking the low-cost effort type.

3.5 Third Best

Finally, consider the more realistic case where the payer cannot observe or verify neither the provider's effort nor its type. In this environment, the contracts must satisfy all three types of constraints described in (2)-(4). The optimal contract is summarized in Proposition 1, with the proof delegated to the appendix.

Proposition 1: The optimal payment contract with moral hazard and adverse selection is as follows:

$$
\begin{aligned}
b_L =& 1 && (14)\\
b_H =& \frac{\alpha\theta_L}{\alpha\theta_L + \Delta\theta(1-\alpha)} \in (0,1) && (15)\\
a_H =& u - 0.5b_H^2/\theta_H && (16)\\
a_L =& a_H - \frac{0.5(1-b_H^2)}{\theta_L} < \alpha_H && (17)
\end{aligned}
$$

Therefore, the optimal contract with both moral hazard and adverse selection consists of a menu of two contracts, in which the contract designed for the high-cost type has a higher capitation rate and a lower cost-reimbursement rate than the contract designed for the low-cost type.

These results are closely related to those derived by Jack (2005), despite some main differences in the model setup. Specifically, Jack focuses on the quality and cost of health care as primary health care goals (rather than access), the providers in his model differ in altruism (rather than the disutility of effort), and there is a continuum of provider types (rather than only two types). Jack shows that the optimal contract can be approximated by a menu of the linear contracts of the form $\alpha(\theta) - \tau(\theta)c$, where $\alpha(.)$ is a fixed salary component, $1 - \tau(.)$ is a cost-reimbursement rate, θ is the degree of provider altruism, and c is the financial cost of providing treatment (e.g., the cost of labor services of other staff). In this model, $\alpha(\theta)$ and $\tau(\theta)$ are both increasing in θ, so that more altruistic physicians choose contracts with higher fixed payment and lower cost-reimbursement rate. Our model is similar to Jack's in two important ways. First, the optimal contract is a menu of contracts rather than a single contract. Second, the relationship between the fixed payment and the cost-reimbursement rate is negative across contracts. However, because Jack focuses on the quality and cost of health care and we focus on the access to health care, our contracts are linear in the number of services, while Jack's contracts are (at least approximately) linear in the financial cost of services.

3.6 Summary

To summarize, there are four main implications of our model with moral hazard and adverse selection. First, the model shows that the optimal contract is a menu of two contracts $[(a_L, b_L), (a_H, b_H)]$ that optimally trades off efficiency and rent extraction. Second, the capitation rate and the cost-reimbursement rate are negatively related across contracts $(a_H > a_L, b_H < b_L)$. Third, for any contract (a, b), the θ_L-type physicians provide more services than the θ_H-type physicians, since $E[q_i] = e_i = b/\theta_i$ and $\theta_L < \theta_H$. In other words, there is a selective sorting of physicians across contracts based on their productivity. Lastly, for any physician type θ, the contract (a_L, b_L) induces more services than the contract (a_H, b_H) since $e_i = b_i/\theta$ and $b_L > b_H$. Therefore, even after controlling for the physician type, the two contracts have differential incentive effects. It is easy to verify that these results are consistent with the stylized facts we discussed in Section 2, once we define physicians who switched to the blended capitation model as the high-cost type, and physicians who remained in the enhanced FFS model as the low-cost type.

4 Conclusion

In this study, we analyzed the design of payment contracts aiming to maximize patient access to physician services when the payer has limited information about physician actions and his/her type. The main policy implications of this analysis are as follows.

First, the design of optimal physician payment contracts may depend critically on the information available to the payer. Specifically, in the class of linear contracts we analyzed, the optimal contract is an FFS contract when the payer can verify the physician type but not her action; in the opposite case, when the payer can verify the physician action but not her type, the optimal contract is a capitation contract. When the payer can verify neither the physician type nor her action, the optimal contract is a menu of contracts that blend FFS and capitation to varying degrees. Second, the analysis draws attention to the potential benefit of screening, whenever there is unobserved heterogeneity among providers, and therefore of offering a menu of payment contracts rather than mandating a single contract. While this point is well understood in the principal-agent literature, it has yet to receive wider recognition in health care. Lastly, and related specifically to the evaluation of payment reforms, our study provides another warning about drawing causal inference based on a simple comparison of providers practicing in different payment contracts. Again, it is well understood that any differences in such comparisons include both incentive and selection effects of contracts, but it is still not unusual to read policy evaluations making no attempt to separate the two effects.

In conclusion, while it is significant that a relatively standard economic model developed in this study can provide a unified explanation for some of the most important features of the primary care reform in Ontario, the model nevertheless comes short in fully explaining all complexities of the reform. For example, other features of the reform such as the introduction and impact of pay for performance bonuses, interdisciplinary teams, and preventive care incentives are not examined. In addition, the model abstracts from questions often studied in the literature such as the cost and quality of healthcare and risk selection in capitation contracts. Extending the model to address this richer set of questions and other features of the reform in Ontario and other jurisdictions remains an important area for future research.

References

[1] Allard, M, Jelovac, I, Lèger, P.T. 2011. Treatment and referral decisions under different physician payment mechanisms. Journal of Health Economics; 30(5): 880-893.

[2] Allard, M, Jelovac, I, Lèger, P.T. 2014. Payment mechanism and GP self-selection: capitation versus fee for service. International Journal of Health Care Finance and Economics; 14(2): 143-160.

[3] Blundell, R, Dias, M.C., Meghir, C, Van Reenen, J. 2004. Evaluating the employment impact of a mandatory job search assistance program. Journal of the European Economics Association; 2(4): 596-606.

[4] Bolton, P, Dewatripont, M. 2004. Contract theory. MIT Press.

[5] Chone, P, Ma, C.A. 2007. Optimal health care contracts under physician agency. Annals of Economics and Statistics; 101/102: 229-256.

[6] Dehejia, R, Wahba, S. 2002. Propensity-score matching methods for nonexperimental causal studies. The Review of Economics and Statistics; 84(1): 151-161.

[7] Dusheiko, M, Gravelle, H, Jacobs, R, Smith, P. 2006. The effect of financial incentives on gatekeeping doctors: Evidence from a natural experiment. Journal of Health Economics; 25: 449-478.

[8] Ellis R.P, McGuire T.G. 1986. Provider behavior under prospective reimbursement. Journal of Health Economics; 5: 129-151.

[9] Ellis, R.P, McGuire, T.G., 1990. Optimal payment systems for health services. Journal of Health Economics; 9: 375-396.

[10] Evans, R.G, 1974. Supplier-induced demand: Empirical evidence and implications. In: Perlman, M. (Ed.). The Economics of Health and Medical Care, Amsterdam: North Holland; 162-173.

[11] Glazier, R.H, Klein-Geltnik, J, Kopp, A, Sibley, L.M, 2009. Capitation and enhanced fee-

for-service models for primary care reform: A population-based evaluation. Canadian Medical Association Journal; 180(11): E72-E81.

[12] Iversen, T, Lurås, H. 2012. Capitation and incentives in primary care. In: The Elgar Companion to Health Economics, Second Edition. Ed. A.M. Jones. (North Hampton: Edward Elgar Publishing Limited).

[13] Jack W. 2005. Purchasing health care services from providers with unknown altruism. Journal of Health Economics; 24: 73-93.

[14] Kantarevic, J, Kralj, B, Weinkauf, D. 2011 Enhanced Fee-for-Service Model and Physician Productivity: Evidence from Family Health Groups in Ontario. Journal of Health Economics; 30(1): 99-111.

[15] Kralj, B, Kantarevic, J. 2013. Quality and quantity in primary care mixed payment models: Evidence from family health organizations in Ontario. Canadian Journal of Economics; 46(1): 208-238.

[16] Laffont, J, Martimort, D. 2002. The theory of incentives: The principal-agent model. Princeton University Press.

[17] Lazear, E. 2000. Performance Pay and Productivity. American Economic Review; 90(5): 1346-1361.

[18] Lèger, P.T., 2008. Physician payment mechanisms. In: Financing Health Care: New Ideas for a Changing Society, Lu, M. and Jonsson, E. (Eds.), Wiley; 149-176.

[19] Leuven, Edwin, and Barbara Sianesi (2003) 'PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing' http://ideas.repec.org/c/boc/bocode/s432001.html.

[20] Li, J, Hurley, J, DeCicca, P, Buckley, G. 2014. Physician response to pay-for-performance -

evidence from a natural experiment. Health Economics; 23(8): 962-78.

[21] Ma, C.A. 1994. Health care payment systems: cost and quality incentives. Journal of Economics and Management Strategy; 3(1): 93-112.

[22] Ma, C.A, McGuire, T. 1997. Optimal health insurance and provider payment. American Economic Review; 87: 685-704.

[23] Makris, M, Siciliani, L. 2013. Optimal incentive schemes for altruistic providers. Journal of Public Economic Theory; 15(5): 675-699.

[24] McGuire, T. 2000. Physician agency. In: Culyer, A.J., Newhouse, J.P. (Eds.), Handbook of Health Economics, vol. 1A. North-Holland, Amsterdam; 461-536.

[25] Nichols, A. 2007. Causal inference with observational data. The Stata Journal; 7(4): 507-541.

[26] Rosenbaum, P, Rubin, D. 1983. The central role of the propensity score in observational studies for causal effects. Biometrika; 70(1): 41-55.

[27] Rosenbaum, P, Rubin, D. 1985. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. The American Statistician; 39(1): 33-38.

[28] Rudoler, D, Deber, R, Barnsley, J, Glazier, R.H, Laporte, A. Paying for primary care: the factors associated with physician self-selection into payment models. Canadian Centre for Health Economics; Working Paper No: 2014-06.

[29] Scott A. 2000. Economics of General Practice. Handbook of Health Economics, Vol. 1, ed. A.L. Culyer and J.P. Newhouse (Amsterdam: North-Holland).

[30] Shearer, B. 2004. Piece rates, fixed wages and incentives: evidence from a field experiment. Review of Economic Studies; 71: 513-534.

[31] Zweifel P, Breyer F, Kiffman M. Health economics. Springer-Verlag:Berlin; 2009.

Appendix
Proof of Proposition 1

The solution to this problem proceeds by assuming that the only relevant constraints are the participation constraint for the high-cost type and the adverse selection constraint for the low-cost type, and then by verifying ex-post that the other two constraints are not binding at the optimum. Therefore, the relevant constraints are

$$a_H + b_H e_H - 0.5\theta_H e_H^2 \geq u \tag{18}$$

$$a_L + b_L e_L - 0.5\theta_L e_L^2 \geq a_H + b_H e_L(b_H) - 0.5\theta_L e_L^2(b_H) \tag{19}$$

At the optimum, both these constraints will be binding. In addition, by using the incentive compatibility constraint ($e_H = b_H/\theta_H$), the expected payment for the high-cost type can be expressed as

$$a_H + b_H e_H = u + 0.5 b_H^2/\theta_H \tag{20}$$

Furthermore, by using the expression for a$_H$ from eq. (19) and the incentive compatibility constraint ($e_L = b_L/\theta_L$), eq. (18) can be rewritten as

$$a_L + b_L e_L = u + 0.5 b_L^2/\theta_H + 0.5 b_H^2 \Delta\theta/\theta_L\theta_H \tag{21}$$

By using eqs. (19) and (20), the payer's expected utility can be expressed as

$$E[V] = \alpha \left[\frac{b_H}{\theta_H} - \frac{0.5b_H^2}{\theta_H} \right] + (1-\alpha) \left[\frac{b_L}{\theta_L} - \frac{0.5b_L^2}{\theta_L} \right] - (1-\alpha) \left[\frac{0.5b_H^2 \Delta\theta}{\theta_H \theta_L} \right] - u \tag{22}$$

Given that this is a concave program, the following first-order conditions for b$_H$ and b$_L$ are both necessary and sufficient:

$$\frac{\partial E[V]}{\partial b_H} = \alpha \left[\frac{1}{\theta_H} - \frac{b_H}{\theta_H} \right] - (1-\alpha) \left[\frac{b_H \Delta\theta}{\theta_H \theta_L} \right] = 0 \tag{23}$$

$$\frac{\partial E[V]}{\partial b_L} = (1-\alpha) \left[\frac{1}{\theta_L} - \frac{b_H}{\theta_L} \right] = 0 \tag{24}$$

Solving these two conditions for b$_H$ and b$_L$ then yields eqs. (13) and (14) in the text. Lastly, the participation constraint (17) can be rewritten as eq. (15) by using the incentive compatibility constraint $e_H = b_H/\theta_H$ and simplifying.

Table 1: New Primary Care Payment Models in Ontario, April 2014

Payment Model	Year of Introduction	Physicians	% of Family Physicians
Harmonized (Blended Capitation)			
Family Health Network	2002	269	2
Family Health Organization	2007	4,591	36
Non-Harmonized (Enhanced FFS)			
Family Health Group	2003	2,749	21
Comprehensive Care Model	2005	333	3

Table 2: Comparison of Elements in Patient Enrolment Models

	Harmonized Models	Non-Harmonized Models
Compensation Elements		
FFS Billings (cost-reimbursement rate)	15%	100%
Capitation		
Comprehensive care management[1]	C$30	C$30
Age-sex adjusted capitation rate[2]	C$170	C$0
Incentives and Bonuses[3]	Yes	Yes
Organizational Elements		
Group Size	≥ 3	≥ 3
Patient Enrolment	Yes	Yes
After-Hours Requirement	Yes	Yes

[1]Approximate rate per patient per year as of April 1, 2014, which is then age-sex adjusted.
[2]Approximate gross rate per patient per year, as of April 1, 2014 which is then age-sex adjusted.
[3]Incentives and bonuses include preventive care bonuses (pap smears, mammograms, childhood immunisations, flu shots, colorectal screening), special payments (obstetrical deliveries, hospital services, palliative care, prenatal care, home visits), chronic disease management fees (diabetes, congestive heart failure), and incentives to enrol unattached patients.

Table 3: Summary Statistics, Fiscal Year 2006/07

	Full Sample	Switchers[1]	Stayers[2,3] (All)	Stayers[4] (Matched)
Number of Physicians	3,641	2,078	1,563	1,350
Services per day	43.1	39.8	47.6*	39.6
Visits per day	29.6	27.9	31.9*	28
Average age	49.2	48.4	50.2*	48.9
Percent male	63.3	62.1	64.8	59.8
Percent in Toronto Central Region	12.3	12.8	11.5	12.6
Expected Income gain (C$)	11,095	44,456	-33,257*	35,616

[1]Includes physicians who were in the Family Health Group (FHG) model both in fiscal 2006/07 and in fiscal 2013/14.
[2]Includes physicians who were in the FHG as of fiscal 2006/07 but switched to the capitation model (FHO) by fiscal 2013/14.
[3]* indicates that the difference from the FHO group is significant at 0.05 level using the two-tail t-test. The t-tests are based on a regression of each variable on the treatment indicator. Before matching, this is an un-weighted regression on the whole sample; after matching, the regression is weighted using the propensity score weights obtained from the local linear regression model with the bi-weight kernel and a bandwidth of 0.2.
[4]Includes physicians who were in the Family Health Group (FHG) model both in fiscal 2006/07 and in fiscal 2013/14 who were matched based on the propensity score to the group of switchers.

Table 4: Difference-in-Difference Matching Estimates of Incentive Effects

	Services per day[2]	Visits per day[3]
Baseline Model[1]	-5.67***	-4.10***
	(0.45)	(0.29)
Alternative Estimators		
Nearest Neighbour (1 neighbour)	-5.06***	-3.87***
	(0.58)	(0.36)
Nearest Neighbour (10 neighbours)	-5.48***	-4.00***
	(0.50)	(0.32)
Kernel	-5.62***	-4.08***
	(0.42)	(0.28)
Alternative Bandwidth Values		
0.05	-5.49***	-4.02***
	(0.47)	(0.30)
0.2	-5.77***	-4.14***
	(0.45)	(0.29)
Alternative Kernel Functions		
Normal	-5.76***	-4.13***
	(0.43)	(0.29)
Uniform	-5.77***	-4.15***
	(0.46)	(0.30)
Epanechnikov	-5.72***	-4.12***
	(0.45)	(0.29)
Tricube	-5.64***	-4.08***
	(0.45)	(0.29)
Alternative Trimming Levels		
No trimming	-6.01***	-4.24***
	(0.50)	(0.32)
10 percent	-5.64***	-4.06***
	(0.43)	(0.28)

NOTES

[1]The baseline model is the local linear regression model, using the bi-weight kernel, the bandwidth of 0.1, and imposing a common support by dropping treatment observation whose propensity score is higher than the maximum or less than the minimum propensity score of the comparison physicians and by dropping 5 percent of the treatment observations at which the propensity score density of the comparison observations is the lowest. The sample size for both dependent variables is 3,428 physicians.

[2]Bootstrap standard errors in parentheses, using 200 bootstrap repetitions. *** indicates significance at 1% level, ** significance at 5% level, * significance at 10% level.

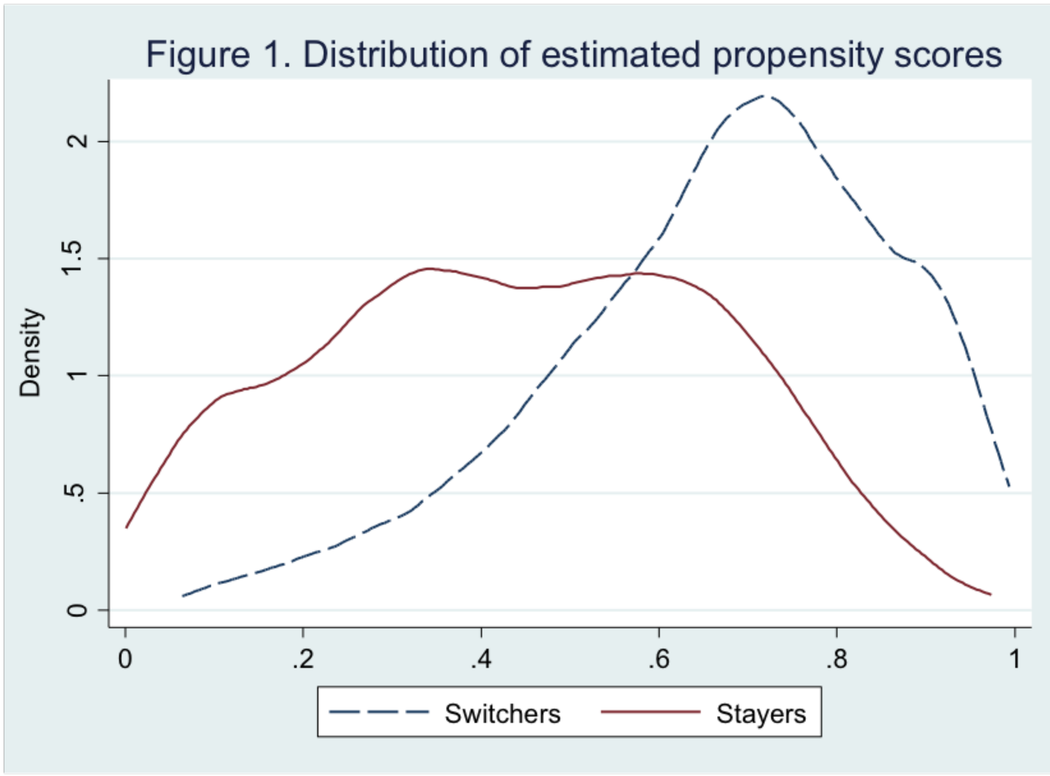
Figure 1. Distribution of estimated propensity scores
Density
0
.5
1
1.5
2
0
.2
.4
.6
.8
1
Switchers
Stayers