Quality and quantity in primary care mixed-payment models: evidence from family health organizations in Ontario

Boris Kralj *Ontario Medical Association and Institute for Labor Studies (IZA)*
Jasmin Kantarevic *Ontario Medical Association*

Abstract. We study the impact of a mixed capitation model (the Family Health Organization, FHO) on quality and quantity outcomes among primary care physicians in Ontario. Using a panel of administrative data covering one year before and two years after the FHO model was introduced, we find that physicians in the FHO model provide about 6% to 7% fewer services and visits per day, but are between 7% and 11% more likely to achieve preventive care quality targets. These results suggest that the mixed capitation model with contractible quality indicators may be welfare improving relative to the FFS model. JEL classification: I10, I12, I18

Qualité et quantité dans les modèles de financement mixte des soins primaires: résultats pour les organismes de santé familiale en Ontario. On étudie l'impact d'un modèle mixte de capitation (les organismes de santé familiale de l'Ontario) sur les résultats en termes de qualité et quantité du travail des médecins de première ligne en Ontario. A l'aide d'un panel de données administratives couvrant la période allant d'un an avant la mise en place de l'arrangement à deux ans après, on montre que les médecins opérant dans ce cadre ont fourni de 6% à 7% de moins de services et de visites par jour, mais sont entre 7% et 11% plus susceptibles d'avoir réussi à atteindre les cibles de qualité de soins préventifs. Les résultats suggèrent que ce modèle avec des indicateurs contractuels pour assurer la qualité peut améliorer le niveau de bien-être par rapport aux résultats du modèle d'honoraires pour les services.

1. Introduction

The traditional fee-for-service (FFS) model, in which physicians receive a fee for each service they provide, has been and still remains the predominant payment

We thank two anonymous referees and participants at the 2011 Canadian Health Economics Study Group's annual meeting in London, Ontario, for useful comments. All remaining errors are ours. 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. Email: jasmin.kantarevic@oma.org

Canadian Journal of Economics / Revue canadienne d'Economique, Vol. 46, No. 1
February / février 2013. Printed in Canada / Imprimé au Canada

0008-4085 / 13 / 208–238 / © Canadian Economics Association

model in many developed countries, including Canada and the United States. This model has been criticized for long time because it may encourage over-provision of health care (see, e.g., Evans 1974) and because it typically lacks incentives to provide valuable but hard to observe quality of care (see, e.g., McGuire 2000). These two criticisms are particularly relevant in the current era of growing health expenditures and concerns about value for money.

As a promising alternative, the recent literature has advocated a mixed-payment model in which physicians receive a reduced fee for each service they provide and a fixed payment for each enrolled patient. Theoretically, this model may be designed to induce the socially optimal levels of both quantity and quality of health care (see, e.g., Léger 2008; McGuire 2008; Zweifel et al. 2009). Empirically, however, it is still not well understood how this model performs relative to the FFS model, especially among primary care physicians (for evidence on specialists, see, e.g., Dumont et al. 2008; Fortin et al. 2010).

In this paper, we provide new evidence on this question by studying a mixed-payment model known as the Family Health Organization (FHO) that was introduced in Ontario, Canada in 2007.[1] We compare the impact of this model on selected quantity and quality outcomes relative to an enhanced fee-for-service model known as the Family Health Group (FHG). The comparison between the two models is particularly revealing because the majority of physicians who joined the FHO model were previously in the FHG model. The comparison is also important from a policy perspective because the FHO and the FHG are currently the two most prevalent primary care models in Ontario, comprising about 60% of all family physicians.

Our analysis is based on a panel of rich administrative data that follows a cohort of FHG physicians from 2006 to 2009, including over one year before and over two years after the FHO model was introduced. Over this sample period, about one-third of FHG physicians switched to the new FHO model. Because this transition was voluntary, we use a two-step approach to address the concern that physicians who joined the FHO model were a selected, non-random sample of FHG physicians. In the first step, we use the propensity score method to match treatment and comparison physicians based on their observed characteristics prior to the introduction of the FHO model. In the second step, we use the propensity score weights to estimate a difference-in-difference model with fixed physician effects and physician-specific trends to control for unobserved, time-invariant physician heterogeneity and to account for differential trends in outcomes between treatment and comparison physicians.

We find that physicians in the FHO model provide about 6% to 7% fewer services and visits per day (about two fewer services or visits per day), but work the same number of days per year and enrol the same number of patients compared with physicians in the FHG model. On the other hand, physicians in the FHO

1 The FHO contract was finalized with the effective date of November 2006, but its implementation was delayed until the summer of 2007.

model are between 7% and 11% more likely to achieve preventive care bonuses for senior flu shots, toddler immunizations, colorectal screening, pap smears, and mammograms than physicians in the FHG model. These results are important because they suggest that the mixed capitation model with observable and contractible quality indicators may be welfare improving relative to the FFS model if, as the critics argue, physicians in the FFS model tend to over-provide quantity but under-provide quality of care.

We also study the performance of the FHO model with respect to two common concerns about capitation models: patient selection or 'cream-skimming' and excessive referrals (see, e.g., Léger 2008). We find that the FHO physicians have about 3% fewer referrals per rostered patient than the comparable FHG physicians. We also find that the FHO physicians enrol patients with expected primary care expenditures similar to patients enrolled with the FHG physicians.[2]

This analysis complements our earlier study of the transition of FFS physicians to the FHG model (see, e.g., Kantarevic, Kralj, and Weinkauf 2010). That transition, following the introduction of the FHG model in 2003, involved incremental changes to how physicians are paid, but the main method of payment remained the FFS system. In this paper, we study a more radical payment reform in which physicians transitioned from an enhanced FFS model (the FHG) to a mixed capitation model (the FHO). As figure 1 shows, these two transitions from the traditional FFS model – to the enhanced FFS model and to the mixed capitation model – represent two main stages in the Ontario primary care reform in the last decade.

Our study is also of direct policy relevance to other jurisdictions. The first relevance is to the concept of Patient-Centered Medical Homes (PCMH) that has recently become quite popular in the United States. The PMCHs are envisioned as multidisciplinary teams based on principles of coordinated and integrated care, quality, and safety, enhanced access, and payment system that reward value (for a brief review see, e.g., a report by Robert Graham Center 2007). In Ontario, the principles of the PCMH are closely embedded in the inter-disciplinary Family Health Teams in which most participating physicians are signatories to the FHO model (see, e.g., Rosser et al. (2010). The second policy relevance is to the Quality and Outcomes Framework that was introduced in the U.K. in 2004 (for a brief overview see, e.g., Smith et al. 2004).This Framework included 146 indicators typically of quality across seven areas of physician practice. About half of available quality points are for clinical indicators that are awarded if a minimum percentage of eligible patients receives the targeted type of care. This payment structure for clinical quality is quite similar to the preventive care bonuses in Ontario that we study in this paper.

The rest of the paper is organized as follows. The next section provides a detailed comparison between the FHO and FHG models. Section 3 then briefly

2 The expected expenditures are measured by the age-sex specific modifiers used to risk-adjust capitation payments. For more details, see section 2.

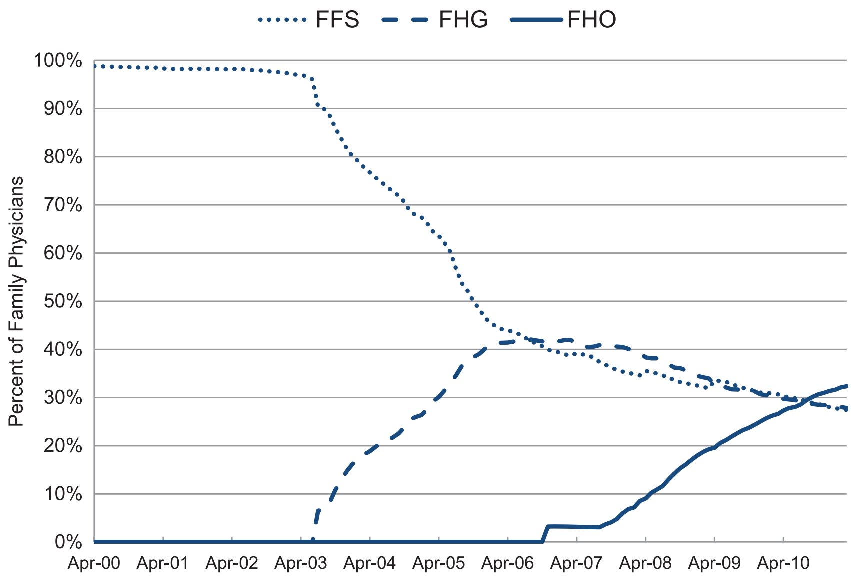

FIGURE 1 Distribution of family physicians in Ontario by various models, April 2000 to March 2011
NOTES: FFS = fee-for-service, FHG = Family Health Groups, and FHO = Family Health Organizations. The FFS group includes physicians not affiliated with any patient enrolment model introduced during the primary care reform. The figure does not show physicians in other, smaller patient enrolment models (e.g., Family Health Network, Comprehensive Care Model).
SOURCE: *Primary Health Care Status Report*, Ontario Ministry of Health and Long-Term Care

reviews the literature on the impact of mixed-payment model on the quantity and quality of health care. Section 4 describes the data and section 5 discusses our empirical strategy. The results are presented in sections 6 (quality and quantity outcomes) and 7 (patient selection and referrals). Section 8 concludes.

2. Institutional background

To appreciate changes involved in the transition from the FHG model to the FHO model, we compare the two models along three main dimensions of physician practice: organization, base compensation, and performance-based incentives (see table 1).

The organization of the two models is quite similar. Both models are group models with at least three physicians, with identical after-hours obligation,[3] and similar patient enrolment requirements. In the FHO model, the enrolment is

3 For example, the minimum requirement for the group of three physicians is to provide a three-hour block of care per week per physician during the after-hour period (evenings on weekdays or any time on weekends and holidays).

TABLE 1
Comparison of FHG and FHO models

Elements	Family Health Group (FHG)	Family Health Organization (FHO)
Organization		
Group size	≥3	≥3
Patient enrolment[a]	Yes	Yes
After-hours requirement	Yes	Yes
Base compensation		
Capitation rate	No	Yes
FFS billings[b]	100%	10%
CCC fee	Yes	Yes
Performance-based incentives		
Preventive care bonuses	Yes	Yes
Chronic disease management	Yes	Yes
Unattached patients	Yes	Yes
Special payments[c]	No	Yes

NOTE: CCC = Comprehensive Care Capitation.
a Mandatory in FHO, optional in FHG
b The FFS billings for FHG also include 10% comprehensive care premium and 20% after-hours premium. The 10% shadow billings for FHO physicians apply only to core services provided to enrolled patients. The FHO physicians receive 100% on non-core services and core services provided to non-enrolled patients, up to a ceiling. The FHO physicians also receive the 20% after-hours premium, but no comprehensive care premium.
c The FHG physicians are eligible for palliative care and serious mental illness special payments. As of October 2009, all primary care physicians are eligible for special payments.

stipulated by the contract; in the FHG model, the enrolment is optional but strongly encouraged because most financial incentives apply to enrolled patients only.

The main difference between the two models is in their base compensation. In the FHO model, physicians receive an age-sex adjusted capitation rate[4] for providing a set of core services to their enrolled patients.[5] In addition, the FHO physicians receive 10% of the FFS value of core services provided to their enrolled patients and 100% of the FFS value of core services provided to the non-enrolled patients (up to a hard cap[6]). For the non-core services, the physicians receive the full FFS value with no hard cap. Specifically, the base compensation in the FHO model can be represented as follows:

$$I_{FHO} = Rm + 0.1p_1q_1m + p_2q_2(m+n) + \min\{p_1q_1n, z\}, \tag{1}$$

4 The age-sex specific modifier includes 19 five-year age categories for each sex. The modifier ranges from 0.44 for males 10–14 years of age to 2.71 for females over 90 years of age, the provincial average being standardized to 1.
5 The FHO core basket includes over 100 comprehensive care services. The complete list of codes is available upon request.
6 In 2006, the value of the hard cap was C$47,500.

where R is the capitation rate,[7] m is the number of enrolled patients, p_1 and q_1 are the price and quantity of core services, p_2 and q_2 are the price and quantity of non-core services, n is the number of non-enrolled patients, and z is the hard cap on core services provided to non-enrolled patients.

In contrast, physicians in the FHG model receive no capitation payment, but they receive the full FFS value for all services provided to both enrolled and non-enrolled patients. In addition, the FHG physicians receive a 10% premium for a set of comprehensive care services provided to their enrolled patients.[8] Specifically, the base compensation in the FHG model can be represented as follows:

$$I_{FHG} = 1.1 p_1 q_1 m + p_1 q_1 n + p_2 q_2 (m + n), \tag{2}$$

where q_1 now represents services eligible for the 10% comprehensive care premium and q_2 represents other services.

Despite these differences, the base compensation contains two elements common to both models: a 20% premium for selected services provided to enrolled patients during after-hours[9] and the Comprehensive Care Capitation (CCC) fee for each enrolled patient.[10] The CCC fee is paid for commitment to provide comprehensive care services to enrolled patients but not for the actual provision of services. For this reason, the CCC fee is better interpreted as a transfer payment designed to meet the participation constraint of FFS physicians interested in joining a primary care model rather than the actual capitation payment.

Lastly, physicians in both models are eligible for a common set of performance-based incentives. These include preventive care bonuses, chronic disease management fees, and incentives to enrol patients with no regular family doctor. The preventive care bonuses are paid if the specific percentage of enrolled and eligible patients receives a defined type of care (pap smears, mammograms, flu shots, immunizations, and colorectal screening). The chronic disease management fees are paid annually for providing required elements of service and currently apply to patients with diabetes and congestive heart failure. The incentives to attach patients are paid as a one-time payment at the time of attachment, and differentiate between regular patients, patients discharged from hospital, complex or vulnerable patients, and unattached mothers with newborns.

7 In the FHO contract, physicians receive a net capitation rate plus an access bonus of up to 18.59% if their enrolled patients receive the core services exclusively from the physicians in the group. The average annual value of gross capitation rate (the net rate plus the full value of access bonus) is equal in value to about five office visits.

8 Services eligible for the 10% comprehensive care premium include assessments in office, emergency department and patient home; pap smear, immunization, flu shot, and annual health exam; primary mental health, HIV, and palliative care; and diabetic assessment. Most of these services are also included in the FHO core basket.

9 The main services eligible for the 20% after-hour premium include a subset of services eligible for the 10% comprehensive care premium.

10 The average annual value of this fee is slightly below the value of one office visit.

In addition to these common incentives, there also exists a set of special payments that until 2009 applied to the FHO physicians only. These bonuses were paid for providing services deemed to be in short supply (obstetrical deliveries, hospital services, prenatal care, home visits).[11] In 2009, the level of bonuses and the type of targeted services significantly changed and the eligibility was extended to all family physicians.

In summary, then, the FHG and FHO physicians share similar organizational structure and are eligible for a similar set of performance-based incentives. The main difference between the two models is in their base compensation. This difference resembles the stylized distinction between the FFS model, in which physicians are paid a full fee for each service they provide but receive no capitation payment, and the mixed capitation model, in which physicians receive a capitation payment and also a partial fee for each service they provide.

3. Expected impact of mixed capitation model

In this section, we briefly review theoretical and empirical literature that examines the relative merits of the FFS model and the mixed capitation model in achieving two important goals of primary care reforms: reducing costs and improving quality.

We start with the fee-for-service model. It is usually argued that physicians in this model tend to over-provide care because their income is directly related to the number of services provided (e.g., McGuire 2000; Scott 2000). This argument is further strengthened by the fact that physicians generally are better informed than patients and insurers about the appropriate level of care (e.g., Evans 1974). On the other hand, the argument about the over-provision of care in the FFS model is weakened if physicians are sufficiently altruistic or if the role of medical ethics is sufficiently strong (e.g., Evans 1974). The empirical evidence on the impact of the FFS model on the provision of care is in general mixed (see McGuire 2000 for a review), and where the evidence supports the over-provision argument, the estimated impact is relatively small (e.g., Gruber and Owings 1996; Yip 1998).[12]

Perhaps a more important concern about the FFS model is that it lacks incentives to provide the efficient level of quality. This concern arises because the quality is usually hard to observe and measure and therefore it is typically not rewarded in the FFS model. The concern remains even when the quality can be observed but imperfectly. In such a multitasking environment, in

11 There were also two special payments, for palliative care and serious mental illness, for which both the FHG and the FHO physicians were eligible during the sample period.

12 As one of the referees pointed out, this evidence can also be interpreted as supporting the claim that physicians may provide unnecessary services to recoup lost earnings. Furthermore, the lack of a large effect may reflect that inducing demand is difficult (because of monitoring and malpractice rewards).

which the compensation can be based on both the quantity and the quality of care, the optimal contract usually deviates from the high-powered FFS model (e.g., Holmstrom and Milgrom 1991; Chalkley and Malcomson 1998; Eggleston 2005).

Consider next the mixed-capitation model. By reducing the fee below the marginal cost of providing the service, it is possible to induce physicians to reduce the quantity of care, perhaps even to the efficient level (e.g., McGuire 2008; Léger 2008). The mixed-payment method can also help improve the quality of care. This improvement can arise if the patients choose their physicians based on the observed but unverifiable quality of care (McGuire 2000). However, the impact of physician competition for patients on the quality of care may depend on the costs of switching physicians and on the extent to which the patient can infer quality from observed health outcomes (Allard et al. 2006). The mixed-payment system can also provide the appropriate incentives for the quality of care when the patient demand for treatment depends on the observed but unverifiable quality (Ma and McGuire 1997).

The empirical evidence on the impact of the mixed-payment model is surprisingly scarce. In a recent study by Dumont et al. (2008), the authors study the introduction of a mixed-payment system (a per diem rate plus a prorated fee for service) for specialists in Quebec in 1999. Relative to the FFS model, physicians who participated in the mixed-payment model reduced billable services by about 6%, but increased their time per service by about 4% and also increased time spent on administration and teaching by about 8%. Fortin, Jacquemont, and Shearer (2010) study the same reform and find similar results, even though their study is limited to pediatricians. The main concern about these studies, as the authors acknowledge, is that time per service may not be a perfect measure of quality because it does not distinguish between time spent with patients and time spent in other activities, nor does it provide information on its effect on patient outcomes. In addition, both studies focus on specialists rather than primary care physicians.

This brief review of the literature suggests that physicians in the mixed-capitation model may provide lower quantity but higher quality of care than physicians in the FFS model. However, these results apply strictly to the case where quality cannot be observed or verified. Our study differs from this environment because the preventive care bonuses that we use as quality indicators are observable and contractible. We expect that these indicators are correlated, even if imperfectly, with the unobserved quality effort, such as time spent with patients. However, it is not a priori clear what impact the availability of these indicators has on the quantity and quality of care. In this paper, we present new evidence on this question by conducting an empirical study that examines whether the results that apply to the environment with the unobserved quality effort also extend to the case where a set of contractible quality indicators is available.

4. Data

4.1. Data sources and study sample

The data come from several administrative sources maintained by the Ontario Ministry of Health and Long-Term Care, described in detail in table A1. These sources can be linked using encrypted physician and patient numbers to construct a rich and comprehensive database that includes almost all family physicians and insured patients in Ontario.[13] From this database, we select the sample of all physicians affiliated with a FHG model as of 1 April 2006 (4,489 physicians, or about 40% of all family physicians in Ontario). This cohort of physicians is then followed for four fiscal years through 31 March 2010. This sample period includes over one year before and over two years after the FHO model was introduced in 2007.

4.2. Definition of treatment

Over the sample period, some physicians remained in the FHG model, while others switched to different models, predominantly the FHO. We define the treatment indicator FHO_{it} as a continuous variable between 0 and 1 that measures the proportion of year during which the physician was affiliated with the FHO model rather than the FHG model.[14] When we use this definition, the initial cohort consists of 4,133 full-year equivalent physicians in 2006, all of whom are in the FHG model (see figure 2).[15] By 2009, about two-thirds of this cohort remained in the FHG model (2,474 full-year equivalent physicians), while about one-third were now in the FHO model (1,358 full-year equivalent physicians).[16]

4.3. Choice of outcomes

We compare the treatment and comparison physicians with respect to selected measures of quantity and quality of care. The quantity measures include the number of services per day, the number of visits per day, the number of annual days of work, and the roster size (as of the last day of each fiscal year).[17] The quality

13 Physicians without any FFS billings or primary care payments, such as salaried physicians, are not included in our database. However, this group represents less than 1% of family physicians in Ontario.

14 We exclude observations in which the physician was not exclusively in either the FHO or the FHG group. However, our results are robust to two alternative definitions of treatment groups, as we discuss in section 6.

15 The number of full-year equivalent physicians in year t is calculated as $\Sigma_i FHO_{it}$ for the FHO model and $\Sigma_i(1 - FHO_{it})$ for the FHG model. This definition reflects the extent of treatment better than the simple head count of physicians at a given date because the switch to the FHO model may occur at any time during the year.

16 The total number of full-time equivalent physicians in 2009 is smaller than in 2006 because 301 physicians switched to a model other than the FHO or ceased to practise during the sample period.

17 The quantity measures include services and visits provided to both enrolled and non-enrolled patients.

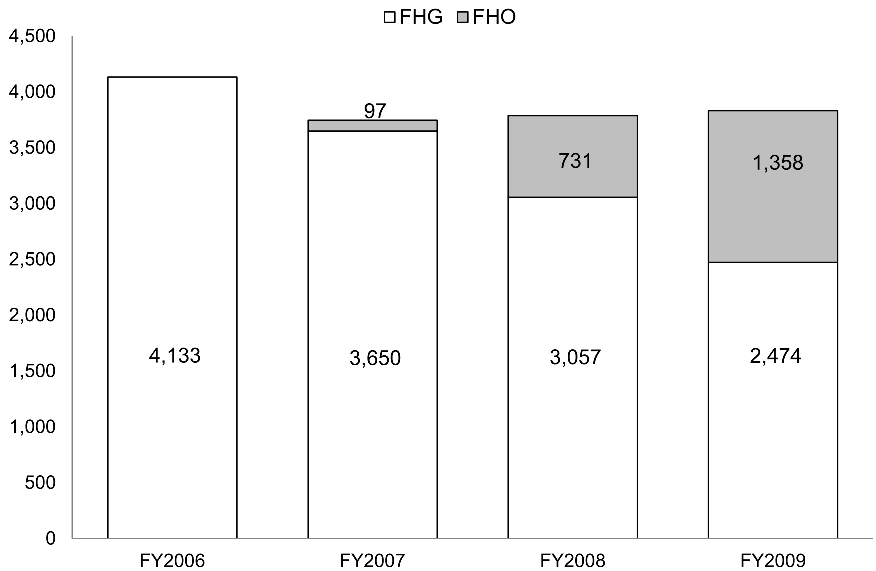

FIGURE 2 Distribution of sample physicians by year and model
NOTES: The number of (full-year equivalent) physicians in fiscal year t is calculated as $\Sigma_i FHO_{it}$ for the FHO model and $\Sigma_i(1 - FHO_{it})$ for the FHG model, where i indexes physicians and FHO represents the proportion of year during which the physician is affiliated with the FHO model rather than the FHG model.

measures include the five preventive care bonuses (PAP smears, mammograms, flu shots, immunizations, and colorectal cancer screening). In our empirical analysis, all quantity measures are used in the logarithmic form while the preventive care bonuses are used as 0–1 indicators for whether the physician received the bonus.

The preventive care bonuses were briefly discussed in the previous section. Further details are described in table A2. With the exception of colorectal screening, which was introduced in 2006, all bonuses have been available to both FHG and FHO physicians since 2007. For this reason, the sample period used for estimating the impact on preventive care bonuses is fiscal 2007 through 2009. Each bonus category specifies a target patient population (e.g., enrolled patients older than 65 years for the bonus for flu shots) as well as the minimum percentage of this population who received the targeted service that qualifies the physician for the bonus payment (e.g., 60%). The bonuses have several payment levels, ranging from C$220 to C$2,200, depending on the percentage of patients who received the targeted service.[18]

18 In 2008, two additional payment levels (C$3,300 and C$4,000) were introduced for the colorectal cancer screening.

TABLE 2
Summary statistics, by treatment status, fiscal year 2006–2007

	Treatment (FHO)	Comparison (FHG) Full sample	Comparison (FHG) Weighted matched sample
Number of physicians	1,530	2,947	2,626
Quantity measures			
Services per day	39.5	43.8*	40.1
Visits per day	27.9	29.7*	28.1
Annual workdays	248	245*	251
Roster size (1 April 2006)	1,167	972*	1,199
Eligibility for preventive care bonuses			
Colorectal screening	0.46	0.34*	0.45
Flu shots for seniors	0.46	0.34*	0.45
PAP smear	0.50	0.37*	0.47
Mammogram	0.52	0.36*	0.49
Toddler immunizations	0.51	0.36*	0.50
Covariates			
Average age	48.6	50.6*	48.7
Percent male	0.62	0.65	0.62
Percent in Toronto Central Region	0.11	0.13*	0.10
Expected income gain (C$)	58,178	−6,389*	60,501

NOTES: FHO = Family Health Organization, FHG = Family Health Group. * 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 unweighted 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.

We exclude performance incentives other than preventive care bonuses from our analysis. Specifically, the special payments are excluded because they were significantly restructured in 2009 and because most of them applied only to the FHO physicians until 2009. In addition, we exclude the chronic disease management and incentives to attach patients because many of them were introduced only towards the end of our sample period (the incentives to attach complex and vulnerable patients and unattached mothers with newborns were introduced in 2009, while the chronic disease management fee for congestive heart failure was introduced in 2008). The analysis of the impact of the mixed capitation model on these incentives remains an area for future research.

4.4. *Summary statistics*

The summary statistics for the sample physicians, as of 2006 when the FHO model was not yet introduced, are presented in the first two columns of table 2. In this table, the treatment physicians are defined as physicians who switched to the FHO model over the sample period, while the comparison physicians are defined as physicians who remained in the FHG model throughout the sample period.

These statistics show some significant differences between the treatment and comparison physicians. In particular, the treatment physicians provide a significantly smaller number of services and visits per day, but they work significantly more days per year and have a significantly larger roster size than the comparison physicians. In addition, the treatment physicians are significantly more likely to be eligible for all preventive care bonuses.

In terms of covariates, the treatment physicians are significantly younger than the comparison physicians, but there is no significant difference in gender composition or geographic location between the two groups. Perhaps most significantly, there is a large and significant difference in the expected change in income from joining the FHO model: a gain of about C$60,000 for the treatment physicians and a loss of about −C$6,000 for the comparison physicians.[19] These descriptive results clearly show that physicians who joined the FHO model were a selected, non-random group of FHG physicians. In the next section, we discuss our empirical strategy to address this selection problem.

5. Empirical strategy

5.1. Difference-in-differences with fixed effects and differential trends

The main challenge with identifying the causal impact of the FHO model is that physicians voluntarily choose which model to join. As a consequence, the observed difference in outcomes between the FHG and FHO physicians may reflect not only the impact of joining the FHO model (the treatment effect) but also any systematic differences between the two groups of physicians that would exist even if they practised in the same model (the selection effect).

To address this concern, we exploit the longitudinal nature of our data to control for unobserved differences between physicians that are time invariant and for physician-specific trends in outcomes. Specifically, our main model is as follows:

$$y_{it} = \gamma_i + \lambda_t + \theta_i t + w_{it}'\beta + \delta FHO_{it} + u_{it}, \tag{3}$$

where y_{it} represents the outcome of interest for physician i in year t; γ_i is the set of physician fixed effects; λ_t is the set of year fixed effects; θ_i is the trend for physician i; w_{it} is the set of time-varying physician characteristics; and FHO_{it} is the treatment indicator equal to the proportion of year during which the physician was affiliated with the FHO model rather than the FHG model. This model is sometimes called the correlated random-trend model and resembles the

19 The methodology for calculating the expected change in income from joining the FHO model is presented in table A3.

standard difference-in-difference model (see, e.g., Imbens and Wooldridge 2006), except that the differences are calculated for the same physicians over time rather than the same groups.

The main coefficient of interest is δ, which represents the mean difference in outcomes between the treatment and comparison physicians. This coefficient identifies the treatment effect of joining the FHO model, provided that the treatment status is strictly exogenous conditional on unobserved fixed effects and physician-specific trends in outcomes. This assumption implies that the treatment status in any time period is not correlated with the idiosyncratic error term in any time period ($E[FHO_{it}u_{is}] = 0$ for s, $t = 1,\dots,$T) (see, e.g., Wooldridge 2002, 252–4).

The strict exogeneity assumption may be violated in our study in two main ways. First, differences in past outcomes may predict future treatment status ($E[FHO_{it}u_{is}] \neq 0$, for $t > s$). In our study, this feedback effect may not be a serious concern because we allow physicians to self-select based on the time-invariant characteristics and linear trends in outcomes. In addition, physicians who joined the FHO model in our sample never reverted to the FHG model, which implies that we have to assume that the timing of switch is not determined by variation in past outcomes, conditional on fixed effects and trends in outcomes.

Second, the strict exogeneity assumption may fail if the FHO impact depends on how long the physician has been in the FHO model ($E[FHO_{it}u_{is}] \neq 0$, for $t < s$). In our data, almost all physicians who switched to the FHO model did so in either 2008 or 2009. Therefore, we can check the assumption that the FHO impact is independent of the timing of switch by testing the significance of δ_1 in the following model:

$$y_{it} = \gamma_i + \lambda_t + \theta_i t + w'_{it}\beta + \delta_0 FHO_{it} + \delta_1 FHO_{it} 2009_i + u_{it}, \tag{4a}$$

where 2009_i is the indicator for physicians who switched to the FHO model in 2009. We also provide further evidence on this assumption by testing whether the FHO impact occurs only in years after the physician switches to the FHO model, but not in the prior years. Specifically, we estimate

$$y_{it} = \gamma_i + \lambda_t + \theta_i t + w'_{it}\beta + \delta_0 FHO_{it-1} + \delta_1 FHO_{it} + \delta_2 FHO_{it+1} + u_{it}, \tag{4b}$$

and we expect that only δ_0 and δ_1 are significant if joining the FHO had a causal impact. While these two dynamic tests are informative, it must be recognized that this approach is limited by the shortness of our panel.

We also conduct formal tests of the significance of fixed effects and physician-specific trends. To conduct the test for the physician-specific trends, we first-difference equation (3) to obtain

$$\Delta y_{it} = \Delta\lambda_t + \theta_i + \Delta w'_{it}\beta + \delta\Delta FHO_{it} + \Delta u_{it}, \tag{5}$$

where Δy_{it} represents $y_{it} - y_{it-1}$, and so on. In this equation, the θ_is are the new fixed effects and the standard F-test can be applied to test whether the θ_is are jointly significant, which is equivalent to testing for the significance of physician-specific trends. To test for the significance of the fixed effects, we estimate the standard fixed-effects or within model, which does not control for trends in outcomes,

$$y_{it} = \gamma_i + \lambda_t + w'_{it}\beta + \delta FHO_{it} + u_{it}, \tag{6}$$

and then apply the standard F-test for the joint significance of γ_is. For completeness, we also estimate the pooled OLS, which controls for neither fixed effects nor trend in outcomes:

$$y_{it} = \lambda_t + w'_{it}\beta + \delta FHO_{it} + u_{it}. \tag{7}$$

This estimator will be consistent only if the treatment status is contemporaneously uncorrelated with the idiosyncratic error term ($E[FHO_{it}\mathrm{u_{it}}] = 0$ for $t = 1, .., T$), but it will be inconsistent whenever the *FHO* and u are correlated in any two time periods.

5.2. *Combining DID with propensity score weights*

As shown in table 2, the treatment and comparison physicians are significantly different in terms of outcomes and covariates prior to the introduction of the FHO model. To the extent that these differences are fixed over time, or at most follow a linear trend, our DID model is expected to work well. However, the availability of pre-treatment data allows us to conduct a further specification test. This test is based on the idea that physicians in the comparison group may not be equally comparable to physicians in the treatment group. Therefore, instead of weighing each comparison physician equally (as in our main model), we can construct alternative weights that reflect how comparable each comparison physician is to the group of treatment physicians.[20]

We construct these alternative weights using the propensity score matching. Specifically, we first estimate the probability of joining the FHO model (the propensity score) using all available outcomes and covariates as of 2006. These include age, sex, geographic location of practice, the expected gain from joining the FHO model, services per day, visits per day, roster size, annual days, and eligibility for each of the five preventive care bonuses. Clearly, all of these covariates and outcomes are independent of treatment because the FHO model was

20 There are several papers in the literature that use the propensity score matching in a difference-in-difference framework to reweight treatment and control groups so that the distribution of pre-treatment covariates is similar across groups. See, for example, Blundell et al. (2004) and discussion in Nichols (2007) and references therein. For theoretical reviews of the propensity score methods, see for example Rosenbaum and Rubin (1983, 1985) and Dehejia and Wahba (2002). For implementation in STATA, see Leuven and Sianesi (2003) and Becker and Ichino (2002).

not introduced until 2007. The more difficult issue is how to correctly specify the propensity score model. We rely on the algorithm by Dehejia and Wahba (2002), which starts with a linear specification and then adds higher-order terms, if required, until the treatment and comparison samples are balanced on each pre-treatment covariate and outcome.[21] This approach attempts to mimic the experimental designs in which randomization ensures that treatment and comparison individuals are balanced on all factors correlated with the treatment.

Given the propensity score, the matching estimators construct a counterfactual outcome for each treatment individual using a sample of comparison individuals. The frequency with which each control comparison individual is used as a match (i.e., weight) depends on the type of matching estimator. In our analysis we compare three most commonly used estimators. First, the nearest neighbour (NN) estimator assigns a weight of one to the closest comparison individual and zero for all others.[22] When a replacement option is used, which allows each comparison individual to be matched to more than one treatment individual, the weight each comparison physician receives is the number of treatment individuals to which each comparison individual is matched. Second, the kernel estimator defines a neighbourhood for each treatment individual and assigns a positive weight to all observations within the neighbourhood and a zero weight to the remaining observations. The weight each comparison individual receives depends on the type of weighting scheme used (the kernel function) as well as the size of neighbourhood (determined by the bandwidth). In our application, we employ the commonly used bi-weight kernel, but we also test the sensitivity of our results to using the normal, Epanechnikov, uniform, and tricube kernels. Similarly, we approach the problem of bandwidth selection empirically by using a fixed bandwidth of 0.2 and testing the sensitivity of our results to alternative specifications.[23] The last matching estimator we consider is the local linear regression (LLR) estimator. This estimator is quite similar to the kernel matching, except that the weighting scheme combines the kernel function and local least squares regression.

The kernel and LLR estimators are in general more efficient than the NN estimator. In addition, the standard errors generated using bootstrap re-sampling methods can be applied for the kernel and LLR estimators, but these methods are not valid for the NN estimator (Abadie and Imbens 2008). For these reasons, the kernel and LLR estimators are typically preferred to the NN estimator. Furthermore, the LLR estimator is in general preferred to the kernel estimator because the bias of the LLR estimator does not depend on the design density of the data

21 Alternative methods are discussed in the literature. For example, Heckman, Ichimura, and Todd (1997) choose the set of matching variables to maximize the equally weighted percentage of observations correctly classified under the probability model.

22 If we use n, rather than only one, nearest control individuals to construct the counterfactual for the treatment physician, each control individual receives the weight of $1/n$.

23 There is a large literature on choosing bandwidths in non-parametric estimation. See, for example, the survey article by Jones, Marron, and Sheather (1996).

and the LLR estimator also avoids the boundary bias problems associated with the kernel estimator (Fan 1992, 1993). Because of these considerations, we rely mainly on the LLR estimator in our analysis, but, as mentioned, we also examine the sensitivity of our results to using the alternative NN and kernel estimators.

5.3. Estimation

We estimate our main model described in equation (5) using a weighted fixed-effects estimator, where the weights come from the propensity score matching. In this model, the weight for each treatment physician is one, while the weight for each observation in the comparison group is the sum of the weights implied by the match to each individual member of the treatment group. To account for possible serial correlation of the outcomes over time, we use the robust Huber-White standard errors clustered at the physician level. This potentially mitigates the over-rejection problem for DID estimates when the inference of the regular t-statistic does not adjust standard error.[24] We also bootstrap the estimate of δ and its standard error using 500 replications to account for the estimation error in the propensity score and the variation that it induces in the matching process.

6. Impact on quantity and quality of care

6.1. Propensity score specification and balancing tests

Table 3 reports the results from the propensity score model that uses available covariates and outcomes as of 2006 to predict whether the physician ever joins the FHO model. The model has considerable predictive power, with pseudo-R^2 of about 0.17, and the likelihood ratio test clearly rejects the hypothesis that the included variables in the model all have zero coefficients.[25] In addition, the model correctly predicts 69.5% of observations (2,857 out of 4,109) classified under the probability model.[26]

Based on the estimated propensity scores, we then construct a weight for each comparison physician using the local linear regression estimator with a bi-weight kernel and a bandwidth of 0.2. Comparison physicians with a positive weight are then retained in the analysis, while the physicians with a zero weight are excluded from the analysis. As mentioned earlier, we select the specification of the propensity score model based on the balancing test that ensures that there is

24 See Bertrand, Duflo, and Mullainathan (2004). Specifically, these authors note that adjusting standard errors for intra-group correlation over time for each cluster works well when the number of clusters is large. The number of clusters (physicians) in our sample is over 4,000.

25 Note that the coefficients in the model do not have a structural interpretation because both outcomes and covariates are included in the model.

26 This is calculated as the sum of treated physicians with the propensity score of at least 0.3609 and comparison physicians with the propensity score below 0.3609 divided by the total number of sample physicians, where 0.3609 represents the percentage of sample physicians who eventually joined the FHO model.

TABLE 3
Propensity score estimates (dependent variable = 0–1 indicator forever in FHO)

Independent variables	Coefficient	Standard error
Age	0.0413	0.0333
Age squared	−0.0007**	0.0003
Male	0.1642*	0.0895
Income gain (×1,000)	0.0109***	0.0007
Quantity measures		
Services per day	−0.0108***	0.0044
Visits per day	0.0315***	0.0079
Annual workdays	0.0030***	0.0008
Roster size	−0.0006***	0.0001
Preventive care bonuses		
Colorectal screening	−0.0736	0.1136
Flu shots for seniors	0.0836	0.1231
Pap smear	−0.1969	0.1566
Mammogram	0.2849*	0.1543
Toddler immunizations	0.2446**	0.1251

NOTES: Logistic regression. The model also includes 14 geographic indicators for regional health areas (LHINs). Sample size is 4,345 physicians. Pseudo-R^2 = 0.1720. LR χ^2 (37) = 925.9, with p-value <0.0001. *** indicates statistical significance at 1% level, ** at 5% level, and * at 10% level. The test of joint insignificance of coefficients on the preventive care bonuses has χ^2 (37) = 21.84, with p-value = 0.0006.

no significant difference in means of all pre-treatment covariates and outcomes between the treatment and matched comparison physicians. The results of this analysis are shown in the third column of table 2. Our propensity score matching retains 2,626 physicians in the matched comparison group out of the full sample of 2,947 comparison physicians. In addition, the treatment and matched comparison physicians are balanced on all pre-treatment covariates and outcomes.

6.2. *Main results*

Our main results are presented in table 4. For comparison purposes, we present results from the least squares model, the fixed-effects model, and the correlated random-trend model. As discussed in section 5, these models use progressively less restrictive identification assumptions about the fixed effects and trends in outcomes. In addition, we present the results using both weighted and unweighted samples of comparison physicians to assess the robustness of our results with respect to the choice of comparison group.

The results from the correlated random-trend model with the weighted sample of comparison physicians are presented in the first column. These results suggest that joining the FHO model has a significant negative impact on the number of services and visits per day and a significant positive impact on the preventive care bonuses. On the other hand, the results indicate no significant impact on the annual days of work and the roster size.

TABLE 4
Impact of joining FHO model: main results

	Weighted results			Unweighted results		
Dependent variable	Correlated trend	Fixed effects	Least squares	Correlated trend	Fixed effects	Least squares
Quantity measures						
Log of services per day	**−0.0594*****	−0.0579***	−0.1775***	−0.0571***	−0.0500***	−0.1988***
	(0.0095)	(0.0075)	(0.0182)	(0.0093)	(0.0072)	(0.0156)
Log of visits per day	**−0.0704*****	−0.0955***	−0.1912***	−0.0685***	−0.0902***	−0.2045***
	(0.0084)	(0.0066)	(0.0161)	(0.0082)	(0.0065)	(0.0142)
Log of annual workdays	**−0.0088**	0.0028	−0.0238	−0.0065	0.0049	0.0067
	(0.0202)	(0.0100)	(0.0154)	(0.0197)	(0.0078)	(0.0130)
Log of roster size	**0.0123**	0.0975***	−0.0022	0.0061***	−0.0272	0.1759***
	(0.0221)	(0.0207)	(0.0335)	(0.0240)	(0.0222)	(0.0319)
Preventive care bonuses						
Colorectal screening	0.0537*	**0.0759*****	0.0384**	0.0512*	0.0629***	0.1300***
	(0.0308)	**(0.0229)**	(0.0162)	(0.0291)	(0.0201)	(0.0162)
Flu shots for seniors	0.1445***	**0.1143*****	0.0816***	0.0717**	0.0855***	0.1699***
	(0.0342)	**(0.0219)**	(0.0191)	(0.0330)	(0.0206)	(0.0172)
Pap smear	0.1196***	**0.1031*****	0.1132***	0.0541*	0.0813***	0.1960***
	(0.0319)	**(0.0225)**	(0.0180)	(0.0307)	(0.0201)	(0.0158)
Mammogram	0.0715**	**0.0743*****	0.1022***	−0.0040	0.0531***	0.1916***
	(0.0308)	**(0.0228)**	(0.0169)	(0.0296)	(0.0199)	(0.0153)
Toddler immunizations	0.0816**	**0.0792*****	0.0552***	−0.0071	0.0432**	0.1586***
	(0.0351)	**(0.0231)**	(0.0162)	(0.0037)	(0.0207)	(0.0162)
Observations	10,402	14,465	14,465	10,402	14,465	14,465
[Physicians]	[3,957]	[4,208]	[4,208]	[3,957]	[4,208]	[4,208]

NOTES: Each cell represents an estimate of δ from model (5) for the dependent variable in the leftmost column. The LS model also includes a quadratic in age, a male indicator, 4-year indicators, and 14 regional indicators. The FE and CRT models include 4-year indicators and 14 regional indicators. Bootstrap standard errors are in parentheses for the matched sample; robust standard errors, clustered at the physician level, are in parentheses for the full sample. *** Indicates statistical significance at 1% level, ** at 5% level, and * at 10% level.

The results from the fixed effects and least squares models are presented in the second and third columns, respectively. These results confirm a negative significant impact on services and visits and a positive significant impact on preventive care bonuses. However, the estimates from the least squares model for services and visits are more than twice as large as those from the other two models, suggesting the importance of controlling for unobserved physician heterogeneity. In addition, the significance of the FHO impact on the roster size is significant only in the fixed-effects model.

We further explore the robustness of these results by using the unweighted sample of comparison physicians.[27] These results, presented in the last three

27 These unweighted results exclude comparison physicians who are not matched to any treatment physician (i.e., those outside the common support region). Our results are qualitatively similar if these physicians are included in the unweighted models. Results are available upon request.

columns, indicate that the weighting of comparison group does not qualitatively change our conclusions regarding the impact on services, visits, and bonuses for flu shots, colorectal screening, and pap smears. However, using the propensity score weights affects the significance of FHO impact for the roster size and the bonuses for mammograms and immunizations.

Our tests of the significance of fixed effects and trends in outcomes, described in section 5, indicate that the fixed effects are significant for all outcomes, but that the physician-specific trends in outcomes are significant for only the quantity measures.[28] As a consequence, our preferred model in the following analysis is the correlated trend model for the quantity outcomes and the fixed-effects model for the preventive care bonuses.[29] Using this preferred model, our results indicate that the reduction in services and visits is about 6% and 7% per day, respectively, while the increase in the probability of qualifying for one of the bonuses ranges between 7% and 11%.[30]

6.3. Dynamics and general equilibrium effects

In this section, we explore two complementary approaches to further test the causal interpretation of our results. The first approach is to examine the dynamics of the FHO impact. As discussed in section 5, we study this issue by testing whether the FHO impact occurs only in years after the physician switches to the FHO model, but not in the prior years, and by testing whether the FHO impact depends on when the physician joined the FHO model.

The results from this analysis are shown in table 5. The reported coefficients in the first three columns represent the FHO impact in a given year relative to two years prior to joining the FHO model. Therefore, these coefficients should be insignificant in the year prior to the switch and significant in the year after the switch and possibly also in the year of switch. These expectations are largely confirmed for services, visits, and preventive care bonuses for flu shots, pap smears, mammograms, and immunizations. However, the impact on the roster size and the bonus for colorectal screening is significant even in the year prior to the switch, which suggests some anticipatory physician behaviour. Nevertheless, the impact after the switch is significantly larger than in any previous year, suggesting that the FHO model had an impact over and above the pre-existing trends. The last two columns in table 6 represent the overall FHO impact and the

28 Results are available upon request.

29 One consequence of this is that the sample size is roughly identical for the two types of models; that is, we lose one year for the correlated random model because of differencing for the quantity measures, and we lose 2006 for the preventive care bonuses estimated using the fixed effects because four out of five bonuses were not introduced until 2007.

30 We have also performed a battery of specification checks that test the robustness of our results to the alternative matching estimators, kernel types, bandwidths, and two alternative definitions of treatment indicator (extending the definition of treatment group to physicians in any capitation model and not only the FHO model, and using only physicians in the FHG and FHO models for the entire year). These variations do not have a significant impact on our results. All results are available upon request.

TABLE 5
Dynamics of FHO Impact

	Impact by year of switch (Model 4b)			Impact by cohort (Model 4a)	
Dependent variable	One year before switch (δ_2)	Year of switch (δ_1)	One year after switch (δ_0)	FHO impact (δ_0)	2009 cohort (δ_1)
Quantity measures					
Log of services per day	−0.0029	−0.0229***	−0.0603***	−0.0665***	0.0384
	(0.0049)	(0.0074)	(0.0098)	(0.0095)	(0.0290
Log of visits per day	−0.0066	−0.0380***	−0.1006***	−0.0712***	0.0041
	(0.0044)	(0.0065)	(0.0085)	(0.0086)	(0.0240)
Log of annual workdays	−0.0059	0.0115	−0.0029	−0.0329	0.1306
	(0.0063)	(0.0087)	(0.0114)	(0.0212)	(0.0962)
Log of roster size	0.0484**	0.0967***	0.1330***	0.0100	0.0127
	(0.0191)	(0.0256)	(0.0262)	(0.0241)	(0.0739)
Preventive care bonuses					
Colorectal screening	0.0414**	0.0468**	0.1182***	0.0821***	−0.0520
	(0.0174)	(0.0232)	(0.0250)	(0.0243)	(0.0459)
Flu shots for seniors	−0.0148	0.0389*	0.0852***	0.1122***	0.0179
	(0.0167)	(0.0220)	(0.0217)	(0.0233)	(0.0503)
Pap smear	−0.0025	0.0244	0.0962***	0.1122***	−0.0757
	(0.0160)	(0.0209)	(0.0210)	(0.0236)	(0.0500)
Mammogram	0.0194	0.0387*	0.0975***	0.0700***	0.0361
	(0.0160)	(0.0204)	(0.0198)	(0.0240)	(0.0487)
Toddler immunizations	0.0180	0.0447**	0.0772***	0.0696***	0.0808
	(0.0157)	(0.0198)	(0.0192)	(0.0242)	(0.0523)

NOTES: The estimates are obtained from the correlated random−trend model for quantity measures and the fixed-effects model for preventive care bonuses, using a matched sample of comparison physicians and controlling for 4-year indicators and 14 regional indicators. Bootstrap standard errors are in parentheses. *** Indicates statistical significance at 1% level, ** at 5% level, and * at 10% level.

incremental impact for physicians who joined the FHO model in 2009, respectively. These results support the conclusion that the FHO impact does not depend on the timing of the switch, since all coefficients for the incremental impact for the 2009 cohort are statistically insignificant. While not conclusive, these two sets of results support the causal interpretation of the FHO impact.

Our second approach tests for the presence of general equilibrium effects. These would occur, for example, if the reduction in the volume of services and visits by the FHO physicians led to an offsetting increase by the FHG physicians. If this were the case, it would compromise the validity of using the FHG physicians as a comparison group.

To examine this hypothesis, we use the sample of FHG physicians only and study whether their outcomes depend on the number of mixed capitation

TABLE 6
General equilibrium effects

	FHO physicians		Any capitation physicians	
Dependent variable	Sub-LHIN	LHIN	Sub-LHIN	LHIN
Quantity measures				
Log of services per day	0.0052	−0.0039	0.0036	0.0094
	(0.0033)	(0.0031)	(0.0035)	(0.0075)
Log of visits per day	0.0033	−0.0022	0.0017	0.0100
	(0.0028)	(0.0026)	(0.0034)	(0.0063)
Log of annual workdays	−0.0008	0.0029	0.0107**	0.0178**
	(0.0044)	(0.0035)	(0.0049)	(0.0072)
Log of roster size	0.0507	0.0295***	0.0327*	0.0310***
	(0.0124)	(0.0082)	(0.0115)	(0.0179)
Preventive care bonuses				
Colorectal screening	0.0094	−0.0283	0.0169	0.0282
	(0.0143)	(0.0280)	(0.0167)	(0.0516)
Flu shots for seniors	0.0113	−0.0218	0.0218	0.0122
	(0.0127)	(0.0217)	(0.0144)	(0.0422)
Pap smear	0.0057	−0.0230	0.0074	−0.0009
	(0.0145)	(0.0279)	(0.0167)	(0.0504)
Mammogram	0.0109	−0.0358	0.0176	0.0099
	(0.0145)	(0.0277)	(0.0172)	(0.0518)
Toddler immunizations	0.0029	−0.0030	0.0083	0.0160
	(0.0126)	(0.0249)	(0.0155)	(0.0441)
Observations	7,179	8,381	7,850	8,489

NOTES: Each cell represents an estimate of the coefficient on the number of FHO (any capitation physicians) for the dependent variable in the leftmost column. The estimates are obtained from the correlated random-trend model for quantity measures and the fixed-effects model for preventive care bonuses, using a sample of FHG physicians only. The models also include 4-year indicators and 14 regional indicators. Bootstrap standard errors are in parentheses. *** Indicates statistical significance at 1% level, ** at 5% level, and * at 10% level.

physicians practising in the same region. For robustness, we study this effect using both the FHO physicians only and using any mixed capitation physicians. In addition, we explore two measures of region of practice: the larger Local Health Integration Area (LHIN) as well as the more local sub-LHIN area. The results are presented in table 6. Overall, the results provide little evidence to support the presence of general equilibrium effects. The only exception is the positive and significant impact on the roster size and annual days in some specifications, but even in this case, the sign of the impact is inconsistent with general equilibrium effects.

6.4. Subgroup analysis

The results reported in table 4 represent the average impact of joining the FHO model. In this section, we examine how this impact varies for specific groups of treatment physicians.

TABLE 7
Impact by age, sex, and rurality

Dependent variable	Age <50	Age ≥50	Female	Male	Rural (RIO > 0)	Urban (RIO = 0)
Quantity measures						
Log of services per day	−0.0804***	−0.0385***	−0.0614***	−0.0586***	−0.0558***	−0.0609***
	(0.0135)	(0.0137)	(0.0161)	(0.0115)	(0.0123)	(0.0148)
Log of visits per day	−0.0921***	−0.0515***	−0.0755***	−0.0673***	−0.0705***	−0.0695***
	(0.0128)	(0.0114)	(0.0156)	(0.0095)	(0.0105)	(0.0148)
Log of annual workdays	−0.0440**	0.0181	−0.0450*	0.0134	0.0034	−0.0288
	(0.0191)	(0.0355)	(0.0235)	(0.0286)	(0.0295)	(0.0193)
Log of roster size	−0.0396	0.0692***	−0.0326	0.0333	0.0218	−0.0293
	(0.0384)	(0.0236)	(0.0382)	(0.0270)	(0.0261)	(0.0441)
Preventive care bonuses						
Colorectal sreening	0.0979***	0.0571*	0.1288***	0.0491*	0.0929***	0.0541*
	(0.0362)	(0.0311)	(0.0377)	(0.0288)	(0.0341)	(0.0319)
Flu shots for seniors	0.1311***	0.1021***	0.1242***	0.1104***	0.1562***	0.0664**
	(0.0333)	(0.0305)	(0.0377)	(0.0268)	(0.0313)	(0.0296)
Pap smear	0.1441***	0.0678**	0.1307***	0.0918***	0.0991***	0.1081***
	(0.0348)	(0.0310)	(0.0373)	(0.0277)	(0.0332)	(0.0317)
Mammogram	0.1213***	0.0379	0.1095***	0.0556*	0.0531	0.0935***
	(0.0352)	(0.0311)	(0.0366)	(0.0290)	(0.0343)	(0.0312)
Toddler immunizations	0.0979***	0.0727**	0.1420***	0.0431	0.0786**	0.0824**
	(0.0354)	(0.0319)	(0.0387)	(0.0286)	(0.0340)	(0.0320)
Observations	4,285	5,920	3,690	6,712	4,921	5,481

NOTE. Each cell represents an estimate of δ from model (5) for the dependent variable in the leftmost column using the sample defined in the uppermost row. The estimates are obtained from the correlated random−trend model for quantity measures and the fixed-effects model for preventive care bonuses, using a matched sample of comparison physicians and controlling for 4-year indicators and 14 regional indicators. Bootstrap standard errors are in parentheses. *** Indicates statistical significance at 1% level, ** at 5% level, and * at 10% level.

In table 7, we examine the FHO impact by age, sex, and rurality.[31] The results indicate that the negative impact on services and visits is stronger for younger physicians (under 50 years of age) compared with older physicians, but there are no significant differences between males and females and between rural and urban physicians. On the other hand, the positive impact on the preventive care bonuses seems to be stronger for female and younger physicians relative to male and older physicians, respectively.

In addition, we study the FHO impact by physician experience in a primary care model. Specifically, we divide the sample into two groups: those who were in the FHG model for at least 18 months as of April 2006 and those who were

31 The rurality is measured using the Rurality Index of Ontario (see Kralj 2000). This index is used by the Ontario Ministry of Health and Long-Term Care in many programs (e.g., Continuing Medical Education) that provide additional incentives to physicians living in rural or remote areas. The index ranges between 0 and 100, with a threshold of 45+ often used to identify rural and remote areas.

TABLE 8
Impact by time in FHG and income levels

Dependent variable	In FHG < 18 months	In FHG ≥ 18 months	Income > \$10K	Income > \$50K	Income > \$100K
Quantity measures					
Log of services per day	−0.0296**	−0.0771***	−0.0621***	−0.0640***	−0.0668***
	(0.0136)	(0.0128)	(0.0086)	(0.0083)	(0.0082)
Log of visits per day	−0.0478***	−0.0843***	−0.0726***	−0.0733***	−0.0765***
	(0.0131)	(0.0109)	(0.0079)	(0.0077)	(0.0076)
Log of annual workdays	−0.0279	0.0028	−0.0130	−0.0261***	−0.0288***
	(0.0212)	(0.0309)	(0.0158)	(0.0099)	(0.0094)
Log of roster size	−0.0116	0.0329**	0.0089	0.0088	0.0051
	(0.0516)	(0.0149)	(0.0220)	(0.0220)	(0.0184)
Preventive care bonuses					
	0.0834**	0.0686**	0.0742***	0.0745***	0.0711***
Colorectal screening	(0.0334)	(0.0315)	(0.0230)	(0.0232)	(0.0235)
Flu shots for seniors	0.1235***	0.1053***	0.1133***	0.1134***	0.1145***
	(0.0314)	(0.0305)	(0.0220)	(0.0221)	(0.0226)
Pap smear	0.1201***	0.0878***	0.1021***	0.1017***	0.0957***
	(0.0315)	(0.0319)	(0.0225)	(0.0227)	(0.0230)
Mammogram	0.0732**	0.0754**	0.0729***	0.0726***	0.0669***
	(0.0321)	(0.0324)	(0.0228)	(0.0230)	(0.0233)
Toddler immunizations	0.0912***	0.0686**	0.0779***	0.0794***	0.0739***
	(0.0349)	(0.0304)	(0.0231)	(0.0233)	(0.0237)
Observations	5,336	5,055	10,362	10,248	9,995

NOTES: Each cell represents an estimate of δ for the dependent variable in the leftmost column using the sample defined in the uppermost row. The estimates are obtained from the correlated random-trend model for quantity measures and the fixed-effects model for preventive care bonuses, using a matched sample of comparison physicians and controlling for 4-year indicators and 14 regional indicators. Bootstrap standard errors re in parentheses. *** Indicates statistical significance at 1% level, ** at 5% level, and * at 10% level.

in the FHG model less than 18 months.[32] The results are presented in the first two columns of table 8 and indicate that the negative impact on services and visits is relatively larger for physicians who were in the FHG model longer than 18 months and that the positive impact on preventive care bonuses is relatively stronger for physicians who were in the FHG model less than 18 months.

Lastly, we also examine the FHO impact by income levels. The results are shown in the last three columns of table 8. The results indicate that the FHO impact on services, visits, and preventive care bonuses is quite similar at all income levels. However, the negative impact on annual days is significant only for samples with income levels of over \$50,000 and \$100,000.

To summarize, these results suggest that the FHO impact varies across different physician groups. In particular, the negative impact on services and visits is more pronounced among younger physicians and physicians with a relatively

32 The period of 18 months was chosen to obtain two groups of similar sample size.

shorter experience with primary care models. On the other hand, the positive impact on the preventive care bonuses is stronger among female physicians, younger physicians, and physicians with a relatively longer experience with primary care models.

6.5. Related literature

Two most relevant studies to our results are Dumont et al. (2008) and Li et al. (2011). As discussed earlier, Dumont et al. (2008) study the impact of a mixed-payment system on specialists in Quebec. They find that relative to the FFS model, specialists who participated in the mixed-payment model reduced billable services by about 6%. Our study confirms that this result extends to primary care physicians and to another jurisdiction, as we document that physicians in a mixed-payment system in Ontario reduced their billable services and visits per day by about 6% to 7% relative to physicians in the enhanced FFS model.[33]

On the other hand, Li et al. (2011) study the same preventive care bonuses in Ontario that we focus on. However, their study differs from ours because they focus on identifying whether availability of these bonuses has any impact of physician behaviour.[34] Therefore, their treatment group includes all physicians eligible for the bonuses (e.g., FHG, FHO), while their comparison group includes pure FFS physicians not eligible for the bonuses. In contrast, our study focuses on whether the eligible physicians' response to these bonuses depends on how they are paid for their other services (i.e., capitation or FFS). In other words, we examine the differential impact of the preventive care bonuses on groups of physicians included in the treatment group by Li et al. As a consequence, the results in Li et al. (2011) are not directly comparable to our study. Instead, our study may be viewed as extending their results, because they show that being eligible for the preventive care bonuses has a positive impact on physician behaviour, while we show that even among eligible physicians this impact depends on how the physicians are paid for their other services.

7. Impact on patient selection and referrals

The capitation model, in which physicians receive a fixed payment for each enrolled patient but no additional payment for services they provide, has usually been criticized on two grounds. First, physicians in the capitation model may

33 Dumont et al. (2008) also find that the mixed-payment physicians increased their time per service by about 4%. This result is not directly comparable to our study because we lack data on time per service.

34 There are other differences between Li et al. (2011) and our study. First, they construct the physician provision of preventive care services based on the physician profile of services and patients, while we use the actual payment data to identify whether the physicians reached the targets or not. Second, Li et al. use data for fiscal 1998 to 2007, while our sample includes fiscal 2006 to 2009. Since the FHO model was introduced in late 2007, this implies that their sample had very few physicians in the FHO model. Lastly, Li et al. also study the special payment incentives that are not included in our analysis.

try to attract patients with lower than expected treatment costs (the patient selection or 'cream-skimming' problem). This incentive may arise if physicians observe patient characteristics that affect the expected treatment costs but are not captured in the risk-adjustment capitation formula (e.g., Ellis 1998). Second, the capitation physicians may excessively refer patients to other physicians because they receive no compensation for additional services they provide to their enrolled patients (e.g., Blomqvist and Léger 2005). In this section, we discuss whether these two concerns may also be relevant in the mixed-capitation model.

Theoretically, these concerns should be less important in a mixed-capitation model because physicians receive partial compensation for services they provide. Therefore, while the cost of treatment is fully shifted to the physician in a pure-capitation model, this cost is partially recovered by the physician in a mixed-capitation model. This difference in cost sharing is expected to reduce the incentive both to selectively enrol patients and excessively refer patients to specialists in a mixed-capitation model (see, e.g., Zweifel, Breyer, and Kiffman 2009).

Empirically, it is challenging to establish whether physicians selectively enrol patients because this selection is expected to occur along unobservable dimensions of patient's health. In this paper, we provide more limited evidence on this issue by testing whether physicians selectively enrol patients based on the expected cost of treatment for specific age and sex groups. Specifically, at the beginning of each fiscal year, we calculate for each physician the average roster 'complexity' as $\sum v_{as} m_{as} / . \sum v_{as}$, where v is the number of enrolled patients in age group a and of sex s, and m is the 38 age-sex specific multipliers used by the Ministry of Health and Long-Term Care to risk-adjust capitation payments in the FHO model.[35]

In this analysis, we restrict our sample to physicians who were in either the FHG model or the FHO model for the entire year, because physicians may switch models at any time and because the number and type of their enrolled patients may change over time. Our results, presented in the first column of table 9, indicate that there is no significant difference in the average roster 'complexity' between physicians in the two models. We also reach the same conclusion when restricting the sample to various physician groups. This evidence establishes that the patient selection does not occur along observable and contractible dimensions of patient's health (i.e., age and gender). Other approaches to gauging cream skimming seem a fruitful topic for future research.

Similarly, it is difficult to ascertain empirically whether physicians excessively refer patients. Again, we provide more limited evidence on this issue by testing

35 These age-sex modifiers have significant power in explaining the average primary care expenditures in Ontario. For example, Buckley et al. (2006) document that 'capitation payment models that adjust for age and sex or age, sex and prior diagnoses allocate payments to primary care practices in closer alignment with the needs of practice populations than fee-for-service payments or a capitation model based on age, sex and characteristics of the patients' area of residence.'

TABLE 9
Impact on patient selection and referrals

Specification	Log of age-sex modifier	Log of referrals per patient: Not controlling for age-sex modifier	Controlling for age-sex modifier
Base model	0.0287	−0.0389***	−0.0337***
	(0.1523)	(0.0106)	(0.0110)
Impact by subgroups			
Age < 50	−0.1667	−0.0619***	−0.0565***
	(0.2866)	(0.0180)	(0.0184)
Age ≥ 50	0.1351	−0.0190	−0.0170
	(0.1443)	(0.0124)	(0.0123)
Males	0.0101	−0.0320**	−0.0270**
	(0.2080)	(0.0124)	(0.134)
Females	0.0128	−0.0486**	−0.0431**
	(0.2135)	(0.0192)	(0.0190)
Rural	−0.0238	−0.1027	−0.0662
	(0.1820)	(0.0732)	(0.0660)
Urban	0.0343	−0.0370***	−0.0312***
	(0.2653)	(0.0107)	(0.0111)
In FHG < 18 months	0.1146	−0.0756***	−0.0658***
	(0.2733)	(0.0192)	(0.0187)
In FHG ≥ 18 months	−0.0474	−0.0137	−0.0113
	(0.1691)	(0.0126)	(0.0135)
Income > $10K	0.0027	−0.0404***	−0.0346***
	(0.1517)	(0.0104)	(0.0109)
Income > $50K	−0.0034	−0.0408***	−0.0351***
	(0.1521)	(0.0104)	(0.0109)
Income > $100K	−0.0181	−0.0409***	−0.0341***
	(0.1542)	(0.0103)	(0.0107)

NOTES: The estimates are obtained from the correlated random-trend model with a matched sample of comparison physicians that also includes 4-year indicators and 14 regional indicators. Bootstrap standard errors are in parentheses. *** Indicates statistical significance at 1% level, ** at 5% level, and * at 10% level. The sample for the log of age-sex modifier includes only physicians who were in the FHG or FHO model for the entire year. The estimates for the log of the age-sex modifier model are multiplied by 100.

whether the FHO physicians refer more patients than FHG physicians, without making any claims about the appropriateness of these referrals. Specifically, we obtain information on all clinical services for which the sample physicians were listed as referring physicians.[36] We then normalize the total number of referrals by the total number of enrolled patients and use the logarithm of this ratio as our outcome variable. The results are presented in the second column of table 9. These results indicate that physicians in the FHO model refer significantly fewer patients than comparable physicians in the FHG model. The average estimated

36 This includes referrals to both focused-practice family physicians and specialists. The results are qualitatively similar if we restrict the referrals to specialists only.

impact is about 4%, but there are significant differences across physician groups. The last column of table 9 shows the same analysis when we control for the average roster ‘complexity,’ as defined above. These results confirm that the FHO physicians have lower referral rates per patient than the FHG physicians, although the estimated impact is relatively smaller.

The results on referrals may seem counterintuitive, as we would expect the capitation physicians to refer more patients than the FFS physicians. However, as mentioned before, the FHO physicians are partially compensated for services they provide to their enrolled patients, which may limit their incentive to excessively refer patients. In addition, the result is not theoretically implausible if other factors are taken into account. For example, Allard, Jelovac, and Léger (2011) show that the difference in referral rates under capitation, fee-for-service, and fund holding depends in part on the extent of physician altruism and ability to correctly diagnose medical conditions. Again, this seems a promising topic for further research.

8. Conclusion

Understanding how physicians respond to incentives in various compensation models has been of policy interest for long time. We contribute to this debate by comparing selected quality and quantity outcomes of primary care physicians in a mixed-capitation model (the Family Health Organization) and an enhanced FFS model (the Family Health Group) in Ontario, Canada. Our results indicate that physicians in the mixed-capitation model provide about 6% to 7% fewer services and visits per day (about two services or visits per day), but are between 7% and 11% more likely to achieve preventive care bonuses than physicians in the enhanced FFS model. These results are important because they suggest that the mixed-capitation model with observable and contractible quality indicators may be welfare improving relative to the FFS model if, as the critics argue, physicians in the FFS model tend to over-provide quantity but under-provide quality of care. We also find that the FHO physicians have lower referral rates per patient and enrol patients of similar complexity compared with the FHG physicians.

There are two main limitations to our study. First, the estimated impact of a mixed-capitation model is necessarily short-term because the model was introduced in Ontario only about four years ago. Future research is needed to confirm whether this impact persists in the long run. Second, our results are perhaps best interpreted as the impact of a mixed-capitation model on physicians who are treated (i.e., those who found it beneficial to switch to this model). Therefore, our results may not generalize to the entire population of primary care physicians.

Future research can build on our analysis in several ways. For example, we have studied the impact of a mixed-capitation model on preventive care bonuses only. Future research can consider this impact on other performance-based quality incentives, such as chronic disease management and incentives to attach patients. In addition, our analysis focuses on the transition of physicians from the FHG model to the FHO model. It is also important to consider other new models for primary care physicians, focusing on determinants of transition between these models and the impact of this transition on physician behaviour.

Appendix

TABLE A1
Description of data sources

Data source	Variables extracted for the analysis
1. Ontario Health Insurance Plan	Physician clinical services, visits, and days
2. Corporate Provider Database	Physician age, sex, and location Physician type of primary care model
3. Architected Payment System	Preventive care bonus payments
4. Registered Persons Database	Patient age and sex
5. Client Agency Program Enrolment	Patient enrolment status Enrolling physician number

TABLE A2
Preventive care bonuses

Bonus type	Targeted patients	Payment levels
1 Flu shots (effective 1 April 2007)	Rostered patients, age 65 or more, who received the flu shot in the previous flu season	\$220 (60% of patients) \$440 (65% of patients) \$770 (70% of patients) \$1,100 (75% of patients) \$2,200 (80% of patients)
2 Pap smears (effective 1 April 2007)	Rostered female patients, age 35 to 69, who received a pap smear for cervical cancer during the last 30 months	\$220 (60% of patients) \$440 (65% of patients) \$660 (70% of patients) \$1,320 (75% of patients) \$2,200 (80% of patients)
3 Mammograms (effective 1 April 2007)	Rostered female patients, age 50 to 69, who received a mammogram for breast cancer during the last 30 months	\$220 (55% of patients) \$440 (60% of patients) \$770 (65% of patients) \$1,320 (70% of patients) \$2,200 (75% of patients)
4 Immunization (effective 1 April 2007)	Rostered children, age 30 to 42 months, who received 5 immunizations by the age of 30 months	\$440 (85% of patients) \$1,100 (90% of patients) \$2,200 (95% of patients)
5 Colorectal cancer screening* (effective 1 April 2006)	Rostered patients, age 50 to 74, who were administered a colorectal screening test by Fecal Occult Blood Testing during the last 30 months	\$220 (15% of patients) \$440 (20% of patients) \$1,100 (40% of patients) \$2,200 (50% of patients)

* In 2008, two additional payments levels were introduced for colorectal cancer screening: \$3,300 for 60% of patients and \$4,000 for 70% of patients.

TABLE A3
Expected income gain
The expected FHO income for the sample of FHG physicians in 2006 is based on the actual service profile and enrolled patients using the following assumptions.

Income source	Estimation methodology
Capitation rate	The gross capitation rate ($144.08) multiplied by the age-sex modifier for each enrolled patient as of 1 April 2006
Shadow claims	10% of FFS value for in-basket services to enrolled patients in 2006
FFS claims	100% of FFS value for out-of-basket services to any patients in 2006
Hard cap	Minimum of 100% FFS value for in-basket services to non-enrolled patients in 2006 or $47,500
Special payment	Calculated using the actual service profile in 2006 and the special payment eligibility rules
After-hour premium	Same as in FHG (actual AH Premium in 2006)
CCC fee	Same as in FHG (actual CCC fee in 2006)
Preventive care bonuses	Same as in FHG (actual bonuses in 2007)
Chronic disease management	Same as in FHG (actual Diabetes Management Fee in 2006)
Unattached patients	Same as in FHG (actual values for Q codes in 2006)

References

Abadie, Alberto, and Guido Imbens (2008) ‘On the failure of the bootstrap for matching estimators,’ *Econometrica* 76, 1537–57

Allard, Marie, Izabela Jelovac, and Pierre Thomas Léger (2011) ‘Treatment and referral decisions under different physician payment mechanism,’ *Journal of Health Economics* 30, 880–93

Allard, Marie, Pierre Thomas Léger, and Lise Rochaix (2009) ‘Provider competition in a dynamic setting,’ *Journal of Economics and Management Strategy* 18, 457–86

Becker, Sascha, and Andrea Ichino (2002) ‘Estimation of average treatment effects based on propensity scores,’ *Stata Journal* 2, 358–77

Bertrand, Marianne, Esther Duflo, and Sendhill Mullainathan (2004) ‘How much should we trust differences-in-differences estimates?’ *Quarterly Journal of Economics* 119, 249–75

Blomqvist, Åke, and Pierre Thomas Léger (2005) ‘Information asymmetry, insurance, and the decision to hospitalize,’ *Journal of Health Economics* 22, 617–35

Blundell, Richard, Monica Costa Dias, Costas Meghir, and John Van Reenen (2004) ‘Evaluating the employment impact of a mandatory job search assistance program,’ *Journal of the European Economics Association* 2, 596–606

Buckley, Neil, Brian Hutchison, Jeremiah Hurley, Paul Contoyannis, Andrea Ruskova, Gioia Buckley, Robert Reid, Murray Finkelstein, and Olga Rakita (2006) *Needs-Adjusted Primary Care Capitation Payment: Development and Comparison of Models*. Report to the Ontario Ministry of Health and Long-Term Care and the Needs-Adjusted Primary Care Capitation Funding Stakeholder Working Group

Chalkley, Martin, and James Malcomson (1998) ‘Contracting for health services when patient demand does not reflect quality,’ *Journal of Health Economics* 17, 1–19

Dehejia, Rajeev, and Sadek Wahba (2002) ‘Propensity-score matching methods for non-experimental causal studies,’ *Review of Economics and Statistics* 84, 151–61

Dumont, Etienne, Bernard Fortin, Nicholas Jacquement, and Bruce Shearer (2008) ‘Physicians’ multitasking and incentives: empirical evidence from a natural experiment,’ *Journal of Health Economics* 27, 1436–50

Eggleston, Karen (2005) 'Multitasking and mixed systems for provider payment,' *Journal of Health Economics* 24, 211–23
Ellis, Randall (1998) 'Creaming, skimping and dumping: provider competition on the intensive and extensive margins,' *Journal of Health Economics* 17, 537–55
Evans, Robert (1974) 'Supplier-induced demand: some empirical evidence and implications.' In *The Economics of Health and Medical Care*, ed. M. Perlman. London: Macmillan
Fan, Jianqing (1992) 'Design adaptive nonparametric regression,' *Journal of the American Statistical Association* 87, 998–1004
— (1993) 'Local linear regression smoothers and their minimax efficiencies,' *Annals of Statistics* 21, 196–216
Fortin, Bernard, Nicholas Jacquement, and Bruce Shearer (2010) 'Labour supply, work effort and contract choice: theory and evidence on physicians.' IZA Discussion Paper 5188
Gruber, Jonathan, and Maria Owings (1996) 'Physician financial incentives and cesarean section delivery,' *Rand Journal of Economics*, 27(1), 99–123
Heckman, James, Hahn Ichimura, and Petra Todd (1997) 'Matching as an econometric evaluation estimator: evidence from evaluating a job training program,' *Review of Economic Studies* 64, 605–54
Holmstrom, Bengt, and Paul Milgrom (1991) 'Multitask principal-agent analyses: incentive contracts, job design,' *Journal of Law, Economics and Organization* 7 (special issue), 24–52
Imbens, Guido, and Jeffrey Wooldridge (2007) 'What's new in econometrics?' NBER Summer Institute, chap. 10
Jones, M., J.S. Marron, and S.J. Sheather (1996) 'A brief survey of bandwidth selection for density estimation,' *Journal of the American Statistical Association* 91, 401–7
Kantarevic, Jasmin, Boris Kralj, and Darrel Weinkauf (2011) 'Enhanced fee-for-service model and physician productivity: evidence from family health groups in Ontario,' Journal of Health Economics 30, 99–111
Kralj, Boris (2000) 'Measuring "rurality" for purposes of health-care planning: an empirical measure for Ontario,' *Ontario Medical Review* 67, 33–49
Léger, Pierre Thomas (2008) 'Physician payment mechanisms.' In *Financing Health Care: New Ideas for a Changing Society*, ed. M. Lu and E. Jonsson. New York: Wiley
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. Version 3.1.5
Li, Jinhu, Jeremiah Hurley, Phil DeCicca, and Gioia Buckley (2011) 'Physician response to pay- for-performance – evidence from a natural experiment.' NBER Working Paper 16909
Ma, Ching-to Albert, and Thomas McGuire (1997) 'Optimal health insurance and provider payment,' *American Economic Review* 87, 685–704
McGuire, Thomas (2000) 'Physician agency.' In *Handbook of Health Economics*, ed. A.J. Culyer and J.P. Newhouse. Vol. 1A. Amsterdam: North-Holland
— (2008) 'Physician fees and behaviour: implications for structuring a fee schedule.' In *Incentives and Choice in Health Care*, ed. F.A. Sloan and H. Kasper. Cambridge, MA: MIT Press
Nichols, Austin (2007) 'Causal inference with observational data,' *Stata Journal* 7, 507–41
Robert Graham Center (2007) *The Patient Centered Medical Home: History, Seven Core Features, Evidence and Transformational Change*. Washington, DC. http://www.graham-center.org/online/etc/medialib/graham/documents/publications

Rosenbaum, Paul, and Donald Rubin (1983) 'The central role of the propensity score in observational studies for causal effects,' *Biometrika* 70, 41–55
— (1985) 'Constructing a control group using multivariate matched sampling methods that incorporate the propensity score,' *American Statistician* 39, 33–8
Rosser, Walter, Jack Colwill, Jan Kasperski, and Lynn Wilson (2010) 'Patient-centered medical homes in Ontario,' *New England Journal of Medicine* 362, e7(1)–e7(3)
Scott, Anthony (2000) 'Economics of general practice.' In *Handbook of Health Economics*, ed. A.J. Culyer and J.P. Newhouse. Vol. 1. Amsterdam: North-Holland
Smith, Peter, and Nick York (2004) 'Quality incentives: the case of U.K. general practitioners,' *Health Affairs* 23, 112–18
Wooldridge, Jeffrey (2002) *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press
Yip, Winnie (1998) 'Physician response to Medicare fee reductions: changes in the volume of coronary artery bypass graft surgeries in the Medicare and private sectors,' *Journal of Health Economics* 17, 675–99
Zweifel, Peter, Friedrich Breyer, and Mathias Kiffman (2009) *Health Economics*. 2nd ed. Chap. 10. Berlin and Heidelberg: Springer-Verlag