Journal of Health Economics 30 (2011) 99–111

Contents lists available at ScienceDirect

Journal of Health Economics

journal homepage: www.elsevier.com/locate/econbase

Enhanced fee-for-service model and physician productivity: Evidence from Family Health Groups in Ontario

Jasmin Kantarevic[a,b,*], Boris Kralj[a], Darrel Weinkauf[a]

[a] *Ontario Medical Association, Toronto, Canada*
[b] *Institute for Labour Study (IZA), Bonn, Germany*

ARTICLE INFO

Article history:
Received 2 March 2010
Received in revised form 14 October 2010
Accepted 18 October 2010
Available online 29 October 2010

JEL classification:
I10
I12
I18

Keywords:
Physician productivity
Remuneration
Primary care
Family Health Groups
Canada

ABSTRACT

We study an enhanced fee-for-service model for primary care physicians in the Family Health Groups (FHG) in Ontario, Canada. In contrast to the traditional fee-for-service (FFS) model, the FHG model includes targeted fee increases, extended hours, performance-based initiatives, and patient enrolment. Using a long panel of claims data, we find that the FHG model significantly increases physician productivity relative to the FFS model, as measured by the number of services, patient visits, and distinct patients seen. We also find that the FHG physicians have lower referral rates and treat slightly more complex patients than the comparable FFS physicians. These results suggest that the FHG model offers a promising alternative to the FFS model for increasing physician productivity.

1. Introduction

Understanding how primary care physicians respond to payment incentives has been an important policy question for decades. The early literature has focused in large part on how three main methods of payment – salary, fee-for-service, and capitation – influence physician behaviour[1]. These traditional methods of payment have been recently reformed in many countries to include incentives for desired healthcare outcomes, such as reaching preventive care targets, improving chronic disease management, and attaching patients with no family doctor[2]. Despite the emerging empirical literature, however, it is still largely unknown how physicians respond to payment incentives in these new payment models[3].

In this paper, we provide new empirical evidence on this question by studying a primary care model known as the Family Health Group (FHG) that was introduced in Ontario, Canada in 2003. The FHG model is an enhanced fee-for-service (FFS) model that includes payment incentives for improving patient access and quality of care, such as premiums for extended hours, bonuses for chronic disease management, and incentives for patient enrolment. We study the impact of joining the FHG relative to the traditional FFS model on three measures of physician productivity: the number of clinical services, visits, and distinct patients seen. Our analysis is based on claims data for almost all primary care physicians in Ontario for eleven years before and five years after the FHG model was introduced.

Our study contributes to the emerging literature on how physicians respond to new payment incentives in several ways. First, we study a primary care model that is based on the targeted healthcare outcomes that are at the front and centre of recent primary care reforms in many countries, such as chronic disease management, enhanced access, and comprehensive care. In addition, we develop a stylized economic model of physician behaviour in

* Corresponding author. Tel.: +1 416 340 2945; fax: +1 416 340 2248.
E-mail address: jasmin.kantarevic@oma.org (J. Kantarevic).

[1] For a review and excellent discussion, see Scott (2000).

[2] For an overview, see the Robert Graham Center report (2007) for the Patient-Centered Medical Homes in the U.S., Smith and York (2004) for the Quality Outcomes Framework in the U.K., and Russell et al. (2009) and Rosser et al. (2010) for the Patient Enrolment Models in Ontario, Canada.

[3] For the U.K. evidence, see for example Campbell et al. (2007) and Sutton et al. (2010). For Ontario, see Glazier et al. (2009) and Russell et al. (2009). See also and Devlin and Sarma (2008) for a more general study of the impact of remuneration schemes on Canadian family physicians. For the U.S., see for example Rosenthal (2010).

0167-6296/$ – see front matter
doi:10.1016/j.jhealeco.2010.10.005

the FHG model. This model is useful as a framework for understanding how physicians respond to the FHG incentives and as a guide for our empirical analysis. The model can also serve as a starting point to study incentive structures in other jurisdictions. Lastly, we use an empirical methodology that can be fruitfully exploited to evaluate how physicians respond to payment incentives when only observational data is available. Specifically, we use the propensity score matching to select control groups of FFS physicians and we use the difference-in-difference model with fixed physician effects and linear trends to evaluate the FHG impact. We also explore multiple 'experiments' and dynamics of the FHG impact to further validate the interpretation of changes in physician behaviour.

We find that joining the FHG model has a meaningful impact on physician productivity relative to the traditional FFS model, as measured by the number of services, visits, and patients. The estimated productivity gain is about six to ten percent, equivalent to about two to three additional weeks of work per year. Furthermore, the impact occurs within the first year of joining the FHG model and persists over time. The impact is also stable across physician groups defined by age, sex, and location of practice. We also find that FHG physicians have significantly lower referral rates to specialists and treat slightly more complex patients than the comparable FFS physicians. These results suggest that the payment incentives in the FHG model significantly improve physician productivity relative to the traditional FFS model.

The rest of the paper is organized as follows. The next section provides a brief introduction to the primary care models in Ontario, including a detailed comparison between the traditional FFS model and the FHG model. Section 3 presents a theoretical analysis of the decision of FFS physicians to join the FHG model and the impact of this decision on their practice profile. Section 4 describes our data and empirical strategy. Section 5 discusses the results and Section 6 concludes.

2. Institutional background

Primary care physicians in Ontario participate in a wide spectrum of patient enrolment models (PEM). These models were introduced in a recent primary care reform that aimed to provide alternatives to the traditional FFS model. The reform dramatically changed how primary care is provided in Ontario. Between 2002 and 2008, the percent of primary care physicians participating in the PEMs increased from less than 5 percent (400 physicians) to over 70 percent (8000 physicians)[4].

The PEMs are of two main types. The harmonized models, such as the Family Health Network (FHN) and the Family Health Organization (FHO), are blended capitation models. The non-harmonized models, such as the Family Health Groups and the Comprehensive Care Model (CCM), are enhanced fee-for-service models. Physicians can choose which PEM to join, but they can also remain in the traditional FFS model. As shown in Table 1, about two thirds of physicians currently participate in the PEMs, while the rest of physicians practice in the FFS model.

The PEMs share four main elements. First, all PEMs are group models, with the exception of the solo CCM model. In most models, the minimum size of the group is three physicians. Second, the PEMs are based on a formal enrolment of patients. The patient enrolment is a required contractual obligation in the harmonized models, but it is strongly encouraged through the enrolment-based fees and payments in all models[5]. Third, most physicians in the PEMs are eligible for performance-based initiatives such as preventive care bonuses and chronic disease incentives. Lastly, the PEM contract stipulates that the physician group provides scheduled extended hours. 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
Primary care compensation models in Ontario, January 2010.

Compensation model	Physicians	% of Family physicians
Patient Enrolment Models		
Harmonized (Blended Capitation)		
Family Health Network	430	3.8%
Family Health Organization	2839	25.0%
Other Harmonized	311	2.7%
Non-Harmonized (Enhanced FFS)		
Family Health Group	3414	30.1%
Comprehensive Care Model	318	2.8%
Other Patient Enrolment Models	235	2.1%
Fee-for-Service (FFS) Model	3796	33.5%

While the PEMs share common elements, there are also subtle differences between various models. These differences are often important from the policy perspective. For example, it is important to know how cost-effective each PEM is relative to the traditional FFS model. In addition, studying a specific PEM can help in understanding how physicians respond to well-defined organizational and compensation elements. Therefore, the policy makers are often interested not in the performance of the entire group of PEMs relative to the FFS model, but in the performance of each PEM individually.

In this paper, we focus on the performance of physicians in the FHG model relative to the traditional FFS model. We focus on the FHG model for several reasons. First, the FHG model is the most popular single compensation model for the primary care physicians in Ontario, especially among full-time physicians. Second, the FHG model is often the first transit point for physicians who migrate from the FFS environment to the PEMs. For example, about 1500 physicians who joined the capitation models between 2007 and 2008 were previously in the FHG model. Third, the quality of data for the FHG model is considered superior compared to the data available for the capitation models. Specifically, physicians in the FHG model receive the full value of their claims, while physicians in the capitation models receive only a fraction of the full value. This difference in the return to accurate reporting may seriously undermine the quality of data for physicians in the capitation models. Lastly, as we explain below, the FHG model is conceptually the closest PEM comparator to the traditional FFS model.

To clarify the comparison between the FFS and FHG models, we present their main organizational and compensation elements in Table 2. As mentioned earlier, the FHG model is a group model based on the patient enrolment and extended hours. While the FFS physicians can also practice in groups and offer extended hours, these elements are not contractual requirements in the FFS model. On the other hand, the FHG physicians receive the full fee-for-service value for services they provide, just like the FFS physicians. However, the FHG model also provides financial incentives that are entirely absent from the FFS model. These incentives include targeted fee increases for comprehensive care services provided during regular hours (10 percent premium) and during after-hours

[4] Estimates are based on the Primary Health Care Status reports by the Ontario Ministry of Health and Long-Term Care.

[5] Even though enrolment is optional in the non-harmonized models, almost all financial incentives incremental to the traditional FFS model are based on patient enrolment.

Table 2
Comparison of elements in FFS and FHG models.

	FFS Model	FHG Model
Organizational Elements		
Group Size	1	≥3
Patient Enrolment	No	Yes
After-Hours Requirement	No	Yes
Compensation Elements		
FFS Billings	100%	100%
FFS Premiums	No CC Premium	10% CC Premium
	No AH Premium	20% AH Premium
CCC Fee	No	Yes
Incentives and Bonuses	No	Yes

Note. FFS = Fee-for-Service; FHG = Family Health Group; CCC = Comprehensive Care Capitation; CC = Comprehensive Care; AH = After-hours. Incentives and bonuses include preventive care bonuses (pap smears, mammograms, childhood immunizations, 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 enroll unattached patients.

(20 percent premium)[6]. To encourage patient enrolment, these premiums are paid only for services provided to enrolled patients. In addition, the FHG physicians are also paid a small Comprehensive Care Capitation (CCC) fee for each enrolled patient[7]. This fee, which is adjusted for the patient age and sex, is paid for commitment to provide comprehensive care services to enrolled patients, and not for the actual provision of services[8]. For this reason, the CCC fee is perhaps best interpreted as a transfer designed to meet the participation constraint of FFS physicians interested in joining the FHG model. Lastly, the FHG physicians are also eligible for performance-based initiatives. These include preventive care bonuses (pap smears, mammograms, childhood immunizations, 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 enroll patients who have no regular family doctor[9]. Most of these initiatives reward physicians if they reach specific service targets or provide specific type of services. In this respect, these initiatives resemble standard fee premiums.

3. Economic model of Family Health Groups

To analyze how differences between the FFS and FHG models affect physician behaviour, we utilize a stylized labour supply model that distinguishes between regular hours of work and after hours. We then use this model to analyze the decision of FFS physicians to join the FHG model and the impact of this decision on their practice profile.

3.1. Fee-for-Service Environment

The physician problem in the FFS environment can be stated as follows:

$$p_1x_1 + p_2x_2 + m = c \tag{1}$$

$$T = l + t_1x_1 + t_2x_2 \tag{2}$$

$$U = u(c, l, x_2) \tag{3}$$

$$x_1 > 0, \qquad x_2 \geq x_{\min} = 0 \tag{4}$$

Eq. (1) specifies the budget constraint. We assume that the physician receives income from three main sources: services provided during regular hours (p_1x_1), services provided during after hours (p_2x_2), and non-labour income (m), where p_i and x_i can be interpreted as the average fee and count of services provided during period i, with $i = 1$ for regular hours and $i = 2$ for after hours. We also assume that the physician spends her entire income on the composite consumption good c. Eq. (2) describes the time constraint. The physician has T units of time which she can allocate to either leisure l or to provision of medical services during regular hours (t_1x_1) or during after hours (t_2x_2), where t_i represents the units of time required to provide one unit of service x_i. Eq. (3) describes physician preferences. We assume that marginal utilities of consumption and leisure are positive (u_C, $u_l > 0$) but non-increasing (u_{CC}, $u_{ll} \leq 0$)[10]. To allow for preferences for the timing of work, we include x_2 as a separate argument in the utility function and assume that $u_2 < 0$ and $u_{22} \leq 0$[11]. For convenience, we also assume that the utility function is separable. Lastly, the two constraints in (4) require that the physician provides some services during regular hours but may choose whether to provide any after-hour services. In the second inequality, $x_{\min}$ represents the minimum after-hours requirement, which is equal to zero in the FFS environment.

This stylized model therefore captures three potential differences between work during regular hours and after hours: prices (p_i), technology (t_i), and preferences.

The first-order conditions for this problem are:

$$u_cp_1 - u_lt_1 = 0 \tag{5}$$

$$u_cp_2 - u_1t_2 + u_2 + \lambda = 0 \tag{6}$$

$$\lambda(x_2 - x_{\min}) = 0 \tag{7}$$

where $\lambda \geq 0$ is the Lagrangean multiplier associated with the after-hours constraint and Eq. (7) holds with complementary slackness. These conditions can be simplified into:

$$\frac{p_1}{t_1} = \frac{u_l}{u_c} \geq \frac{p_2}{t_2} + \frac{u_2}{t_2u_c} \tag{8}$$

The first term (p_1/t_1) represents the marginal return per unit of time to work during regular hours. Similarly, p_1/t_1 and

[6] Nineteen services are eligible for the 10 percent premium during regular hours, which 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. Ten services are eligible for the 20 percent premium during after-hours, which 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. Eight out of ten services eligible for the premium during after-hours are contained in the list of nineteen services eligible for the premium during regular hours.

[7] The average annual value of this fee is C$25.8. For comparison, the fee value for a single intermediate assessment that constitutes the bulk of physician billings is C$32.35.

[8] The CCC fee is adjusted by the age-sex specific modifier which 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, with the provincial average standardized to 1.

[9] Preventive care bonuses have targets based on the ratio of patients receiving the service to the number of patients eligible for the service, similar to how the QOF indicators in the UK are defined. Special payments are based on achieving specific minimum service targets (e.g. the bonus is C$5000 if selected hospital services are at least C$2000 per annum). Chronic disease management fees are paid annually for providing required elements of service. The incentive to attach patients with no family doctor are paid as a one-time payment at the time of attachment, and differentiate between regular patients, patients discharged from hospital, complex/vulnerable patients, and mothers with newborns.

[10] In the literature, the case with $u_{CC} \to 0$ is known as the profit maximization hypothesis, while the case with $u_{CC} \to -\infty$ as the income target hypothesis. See for example McGuire and Pauly (1991) and McGuire (2000).

[11] With $u_2 = 0$, the physician will completely specialize in one type of service, depending on the relation between p_1/t_1 and p_2/t_2.

$p_2/t_2 + u_2/t_2u_c$ represent the marginal returns per unit of time to leisure and work during after-hours, respectively. Therefore, the physician allocates her time between the three activities by comparing their marginal monetary returns per unit of time. At the interior solution ($\lambda = 0$), these returns are equalized and the physician is indifferent between allocating an additional unit of time to work during regular hours, leisure, and work during after-hours. At the corner solution ($\lambda > 0$), the physician is indifferent between work during regular hours and leisure, but she provides no after-hour services.

3.2. Family Health Group Environment

The physician problem in the FHG environment differs in three main ways. First, the physician is now required to provide a minimum number of after-hour services, which modifies the after-hour constraint by $\Delta x_{\min} > 0$. Second, the physician receives a comprehensive care premium of k_1 for selected services provided to enrolled patients during regular hours and an after-hours premium of k_2 for selected services provided to enrolled patients provided during after hours. If we let a represent the fraction of services eligible for the premiums and b the fraction of patients who are enrolled, the fee increases can be written as $\Delta p_1 = abk_1p_1 > 0$ and $\Delta p_2 = abk_2p_2 > 0$. Lastly, the physician also receives a Comprehensive Care Capitation fee for each enrolled patient. We interpret this fee as an increase in the non-labour income by $\Delta m > 0$, where Δm is the product of the per-patient capitation fee and the number of patients to whom the physician is the primary care provider. For simplicity, we do not separately model the physician decision on how many patients to enroll. We assume that the physician decides on how many services to provide and this choice then determines the number of patients that the physician can enroll, given some minimum number of services that each patient can be expected to receive. Once the number of patients is determined, the CCC fee can be interpreted as a type of non-labour income because this fee is paid in lieu of commitment to provide comprehensive care services, not for the actual provision of services, and this commitment does not require additional physician time.

We also abstract from differences in the group size between the FFS and FHG models. The FHG physician groups are loosely defined and the FHG contract does not require that physicians share the same physical office. In addition, all payments are made to individual physicians, with the exception of Telephone Health Advisory, which represents a very minor source of income for the FHG physicians. In addition, we abstract from performance-based initiatives for analytical convenience and because these incentives reward physicians if they reach specific service targets, or provide selected services. These initiatives can then be considered as fee premiums, which would be incorporated in our model with an appropriate re-interpretation of p_1 and p_2.

3.3. Decision to join Family Health Group

Let $(x_1^0, \quad x_2^0)$ denote the optimal solution to the physician problem in the FFS environment and let v be the associated value function. Using the envelope theorem, the change in v from joining the FHG model can be approximated by:

$$\Delta v = u_c x_1^0 \, \Delta p_1 + u_c x_2^0 \, \Delta p_2 + u_c \, \Delta m - \lambda \, \Delta x_{\min} \tag{9}$$

Rearranging the terms, the decision to join the FHG model can be written as:

$$x_1^0 \, \Delta p_1 + x_2^0 \, \Delta p_2 + \Delta m \geq \left(\frac{\lambda}{u_c} \right) \, \Delta x_{\min} \tag{10}$$

The left side of Eq. (10) represents the expected gain in income evaluated at the service profile prior to joining the FHG model. This expected gain is similar to what the Ministry of Health and Long-Term Care provides in its free revenue reports to FFS physicians interested in joining the FHG model. In our empirical analysis, we construct a similar measure of the expected income gain, using the actual service and patient profiles prior to joining the FHG model.

The right side of Eq. (10) can be interpreted as the monetary value of disutility from increasing the after-hours requirement in the FHG model relative to the FFS environment. When the after-hour constraint is not binding ($\lambda = 0$), the decision rule in (10) implies that the FFS physician would always join the FHG model. When the after-hours constraint is binding ($\lambda > 0$), the first order conditions (5) and (6) can be used to write the right side of (10) as:

$$\frac{\lambda}{u_c} \, \Delta x_{\min} = \left(\frac{t_2}{t_1} p_1 - p_2 - \frac{u_2(0)}{u_c(p_1 x_1^0 + m)} \right) \, \Delta x_{\min} \tag{11}$$

This expression implies that the physician is less likely to join the FHG model the more productive she is during regular hours relative to after-hours, the higher is the average fee for services during regular hours, the lower is the average fee for services during after hours, the higher is the physician non-labour income, and the higher is the physician disutility associated with working during after hours. The physician is also less likely to join the FHG model the higher is the after-hours constraint relative to their current practice. In our empirical analysis, we proxy for these factors by using age, gender, location, the expected income gain, and the number of working weekends and holidays to predict physician decision to join the FHG model.

3.4. Impact of joining Family Health Group on practice profile

The FHG impact on the physician practice profile depends on the relative strength of income and substitution effects arising from changes in fees and non-labour income. When income effects are negligible and the physician is at least as productive during after-hours as during regular hours ($t_2 \leq t_1$), the number of total services ($x_1 + x_2$) will unambiguously increase[12]. These two conditions are sufficient, but not necessary, for both cases when the after-hour constraint is binding and when it is not. However, in the general case the impact of joining the FHG model is ambigious and remains an empirical question.

4. Data and empirical framework

4.1. Data

The data comes from the Ontario Health Insurance Plan (OHIP) fee-for-service claims for the fiscal years 1992–2008. This period includes eleven years before and five years after the FHG model was introduced in 2003.

The OHIP data has several advantages for our analysis. It includes virtually all family physicians in Ontario who are potentially affected by the introduction of the FHG model[13]. In addition, the seventeen-year panel improves our chances to distinguish significant deviations in physician behaviour from long-term secular trends. The data is also reported for each physician–patient encounter and for each type of service provided during this encounter. Such detailed data enables us to examine a rich set

[12] See Appendix A.

[13] Physicians with no fee-for-service claims are not included in the OHIP data. This group includes mainly salaried physicians who represent less than 1 percent of all family physicians.

of outcomes, including the number and type of clinical services, the number of total patient visits and distinct patients seen, and the number of referrals to specialists. Lastly, the claims data alleviates problems such as measurement error and recall bias that are sometimes present in self-reported surveys. On the other hand, the main disadvantage of the OHIP data is that it contains limited demographic information for patients and physicians (age, sex, and location only).

Our full sample includes 10,111 family physicians with some fee-for-service claims in 2002, the immediate year before the FHG model was introduced[14]. Therefore, our sample excludes physicians who ceased to practice prior to 2002 and those who started to practice after 2002. The full sample is divided between 5260 physicians who joined the FHG model in any year between 2003 and 2008 (the treatment group) and 4851 physicians who never joined the FHG model (the comparison group).

The summary statistics for this sample, as of 2002, are presented in the first two columns in Table 3. These statistics show some striking differences in pre-treatment covariates between the treatment and comparison physicians. Specifically, the treatment sample is on average 4 years younger and has about 5 percent fewer male physicians and about 6 percent fewer physicians residing in the Toronto Central region relative to the comparison sample. More dramatically, the expected gain from joining the FHG model is twice as high for the treatment physicians compared to the comparison physicians. The treatment physicians also work about 50 percent more weekends and holidays than the comparison physicians. All of these differences are statistically significant. Moreover, the differences in the pre-treatment outcomes are also statistically significant, indicating that the treatment physicians provide substantially more annual services and visits and see more distinct patients than the comparison physicians.

These results indicate that physicians who joined the FHG model were positively selected from the population of all family physicians. To partially address this selection problem, we use the propensity score matching to select a sub-sample of comparison physicians with observed covariates most similar to the treatment physicians[15]. Specifically, we first estimate the probability of joining the FHG model (the propensity score) based on physician age, gender, location, expected gain from joining the FHG model, and the number of working weekends and holidays using the full sample of family physicians in 2002[16]. Our specification of the probability model is based 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 covariate. Based on this algorithm, our probability model includes quartic functions of age, expected gain, and working weekends and holidays; an indicator for male physicians; an interaction term between the male indicator and age; and 14 regional indicators[17].

In the second step, we use the nearest neighbor matching to select which comparison physicians to include in the final sample. In the nearest neighbor matching, each treatment physician is matched on the propensity score to the nearest comparison

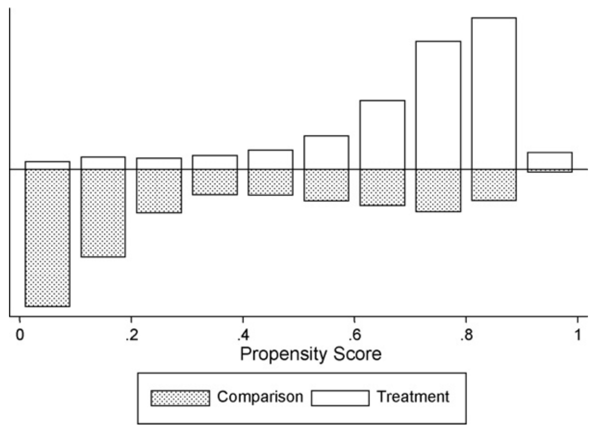

Fig. 1. Distribution of physicians on estimated propensity score.

physician. We also use the replacement option with the nearest neighbor matching which allows a comparison physician to be matched to more than one treatment physician. This option is preferred to matching without replacement if the distribution of propensity scores is very different between the comparison and treatment groups[18]. In our sample, the replacement option seems more appropriate given the empirical distribution of propensity scores shown in Fig. 1.

The summary statistics for the matched comparison group are shown in the third column of Table 3**.** This sample consists of 1734 physicians compared to 4851 physicians in the full comparison sample. The matched comparison physicians are quite similar to the treatment physicians with respect to the pre-treatment covariates and outcomes. Moreover, none of the differences observed in the full sample are statistically significant. Therefore, the propensity score matching seems to significantly reduce the pre-treatment imbalances between treatment and comparison physicians in our sample.

4.2. Empirical framework

Our empirical strategy relies on contrasting changes in outcomes for the treatment and comparison physicians before and after the FHG model was introduced. This strategy critically depends on comparability of physicians in the two groups. While the propensity score matching ensures that the treatment and matched comparison physicians are similar at one point in time (just prior to the introduction of the FHG model), the two groups also need to have comparable trends in outcomes over time.

To examine the common trend assumption, we calculated the weighted average for three main outcomes (log of annual services, visits, and distinct patients) for each group over the sample period. The results are presented in Figs. 2–4. The figures show that the outcome trends for the two groups were quite similar until 2002, but then significantly diverged. In addition, as shown by the small bars in the figures, the most significant changes occurred between 2003 and 2005, when over thousand physicians joined the FHG model in each year. These figures suggest that the treatment physicians permanently shifted their productivity profiles upward after 2002. Moreover, the figures imply that this shift occurred soon after the physicians joined the FHG model.

While these changes coincide with the introduction of the FHG model, this relationship of course may not be causal. The main con-

[14] The sample excludes 379 physicians who were in harmonized models in 2002.

[15] For theoretical reviews, see for example Rosenbaum and Rubin (1983, 1985) and Dehejia and Wahba (2002). For implementation in STATA, see Leuven and Sianesi (2003).

[16] We also used the patient case mix based on age and sex as an additional matching control, with no significant impact on our main conclusions. The results are available upon request.

[17] The role of propensity score is to solve the dimensionality problem associated with the exact matching on multiple covariates and as such has no behavioural assumptions attached to it.

[18] See for example Dehejia and Wahba (2002).

Table 3
Summary statistics, 2002.

	Treatment (FHG) Group	Comparison (Never FHG) Group	
		Full Sample	Matched Sample
Number of Physicians	5260	4851	1734
Covariates			
Average Age	46.5	50.3*	46.4
Percent Male	0.65	0.70*	0.66
Percent in Toronto Central Region	0.12	0.18*	0.13
Expected Income Gain (C$)	42844	18222*	42629
Working Weekends and Holidays	31.6	21.0*	32.1
Outcomes			
Log of Annual Services	8.92	7.83*	8.93
Log of Annual Visits	8.65	7.33*	8.65
Log of Annual Distinct Patients	7.49	6.37*	7.49

Note. FHG = Family Health Group. The *t*-tests are based on a regression of each covariate on the treatment indicator. Before matching, this is an unweighted regression on the whole sample; after matching, the regression is weighted using the number of times each comparison physician is matched to a physician in the treatment group.
* Difference from the FHG group is significant at 0.05 level using the two-tail *t*-test.

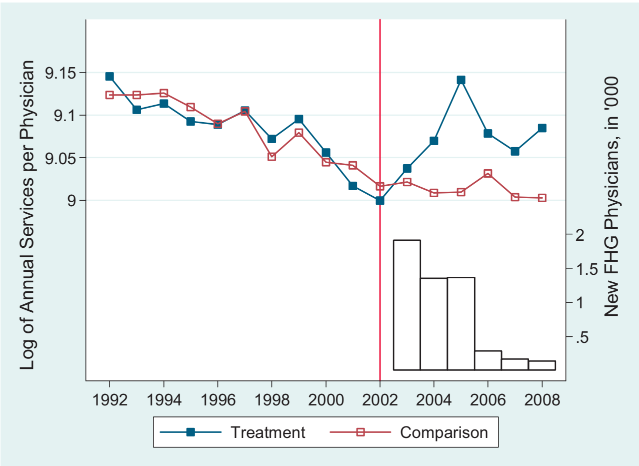

Fig. 2. Log of annual services per physician, 1992–2008.

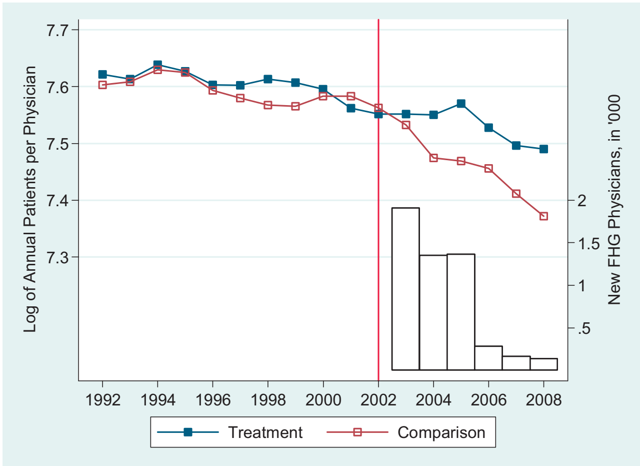

Fig. 4. Log of annual distinct patients per physician, 1992–2008.

cern is that physicians choose whether to join the FHG model, and factors that determine this choice may also be correlated with their productivity. To address this concern, we use the correlated random trend model (Wooldridge, 2005). This model resembles the standard difference-in-differences model, except that we calculate the differences for the same physicians over time rather than the same groups. Importantly, the model also controls for physician-specific linear trends in outcomes[19].

Specifically, our baseline model is:

$$y_{it} = \gamma_i + \lambda_t + \theta_i t + w'_{it}\beta + \delta \mathrm{FHG}_{it} + u_{it} \quad (12)$$

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 FHG_{it} is the treatment indicator equal to 1 if the physician is in the FHG model at time t and 0 if the physician is in the FFS model.

In this model, γ_i controls for mean differences in outcomes across physicians, λ_t controls for trends in outcomes common to all physicians, and θ_i captures the physician-specific linear trend in outcomes. Therefore, the coefficient δ represents the difference in outcomes for treatment and comparison physicians, controlling for fixed physician and year effects and physician-specific linear trends. This difference may be interpreted as a causal impact of joining the FHG model provided that idiosyncratic deviations from the linear trend in outcomes do not vary systematically between treatment and comparison physicians except for the FHG impact.

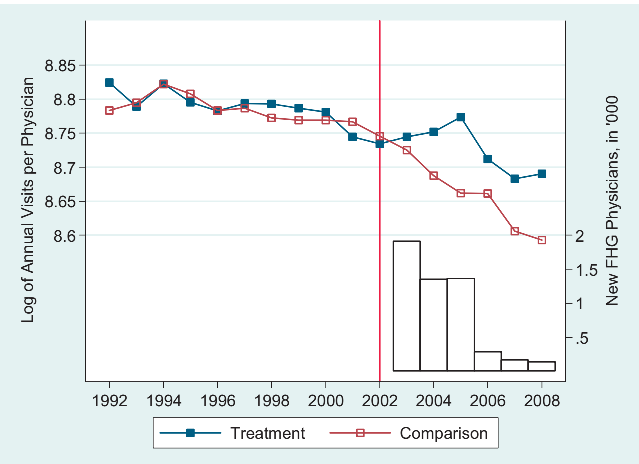

Fig. 3. Log of annual visits per physician, 1992–2008.

[19] We have also estimated models with the quadratic physician-specific trends, with no significant impact on our main conclusions. These results are available upon request.

Table 4
Impact of joining FHG on physician productivity (sample selection using nearest neighbor matching).

Specification	Sample Size [Physicians]	Dependent Variable		
		Log of Services	Log of Visits	Log of Patients
OLS	95890 [6938]	0.1107 (0.0256)	0.1012 (0.0251)	0.0911 (0.0344)
Fixed Effects	95890 [6938]	0.1374 (0.0219)	0.1209 (0.0212)	0.1167 (0.0237)
Correlated Random Trend	89741 [6929]	0.0936 (0.0090)	0.0682 (0.0090)	0.0610 (0.0088)

Note. FHG = Family Health Group. Bootstrap standard errors in parentheses. The OLS model includes age, age squared, male, age-male interaction term, expected income gain in 2002, number of working holidays and weekends in 2002, 14 regional indicators, and 17 year effects. The fixed-effect model and the correlated random trend model include 14 regional indicators and 17 year effects. The sample excludes observations with annual fee-for-service billings below C$30K.

Table 5
Impact of joining FHG on physician productivity (sample selection using caliper matching).

Specification	Sample Size [Physicians]	Dependent Variable		
		Log of Services	Log of Visits	Log of Patients
Distance = 0.0005	86771 [6695]	0.0975 (0.0095)	0.0723 (0.0095)	0.0649 (0.0093)
Distance = 0.0001	52737 [4103]	0.0987 (0.0139)	0.0749 (0.0141)	0.0657 (0.0138)
Distance = 0.00005	34401 [2704]	0.0998 (0.0184)	0.0763 (0.0187)	0.0667 (0.0184)

Note. FHG = Family Health Group. Bootstrap standard errors in parentheses. Estimates from correlated random trend model which also includes 14 regional indicators and 17 year effects. The sample excludes observations with annual fee-for-service billings below C$30K.

The model is estimated by first differencing Eq. (12) to remove γ_i and then applying a fixed effects estimator. In this estimation, we use weights from the matching step to account for the fact that some comparison physicians were matched to more than one treatment physician. We also use robust Huber-White standard errors clustered at the physician level to account for clustering and serial correlation. Because of the estimation error in the propensity score and the variation that it induces in the matching process, we bootstrap the estimate of δ and its standard error using the following procedure. We first draw a bootstrap sample with replacement from the full sample of 10,111 family physicians in 2002. We then estimate the propensity scores for this sample, using the same specification of the probability model as in the previous section, and use the nearest neighbor matching with replacement to select which comparison physicians to include in the final sample. This final sample is then matched to the full 1992–2008 period and the model is estimated as described above. The coefficients and standard errors for δ reported in the following sections are averages over 500 such replications.

5. Results

5.1. Initial estimates

Our initial estimates of the FHG impact are presented in Table 4. These estimates are based on the sample that excludes years with annual billings below C$30,000, a common income threshold used to identify physicians with the minimum attachment to the labour force. For comparison, we present results from the OLS model, the fixed effects model, and the correlated random trend model. These models progressively add more physician-specific effects: the OLS model includes two observed fixed effects: the expected gain in 2002 and the number of working weekends and holidays in 2002; the fixed effects model includes a full set of physician fixed effects; and the correlated random trend includes both fixed effects and physician-specific linear trends. All three models indicate a positive and statistically significant difference in each outcome between the treatment and comparison physicians[20]. The estimates from our preferred correlated random trend model suggest that joining the FHG model had a meaningful impact on physician behaviour, increasing the number of annual services, visits, and distinct patients by about 9.8, 7.1, and 6.3 percent, respectively[21]. Based on the summary statistics in Table 3 and an average of 235 annual days of work, these changes are equivalent to about two to three additional weeks of work per year[22].

The estimates in Table 4 are based on samples selected using the nearest neighbor matching, where each treatment physician is matched to the single nearest comparison physician. In Table 5, we report results for the caliper matching that uses all comparison physicians that have a propensity score within a specified distance from the matched treatment physician. Theoretically, the choice between the two matching methods depends on the trade-off between bias and efficiency, as the caliper matching can improve standard errors relative to nearest neighbor matching, although at the cost of greater bias. In our application, this choice seems to be of little consequence. The point estimates from the correlated random trend model in Table 5 are within the 95 percent confidence interval for our initial estimates in Table 4. Based on these results, we continue to use the nearest neighbor matching in the remainder of our analysis.

5.2. Multiple 'Experiments' and dynamics of FHG impact

As mentioned previously, the estimates from the correlated random trend model may have a causal interpretation provided that deviations from the linear trend in outcomes do not vary system-

[20] The full set of results for each of the three models is available upon request.

[21] The reported percent estimates are calculated by taking the exponential value of the coefficient estimates from Table 4 and subtracting one.

[22] We also estimated models with shorter time windows around fiscal year 2003 to examine the assumption that unobservable factors may change significantly over the seventeen year period, but the main results remain essentially unchanged. All results are available upon request.

Table 6
Impact by Cohort.

Sample	Sample Size [Physicians]	Dependent Variable		
		Log of Services	Log of Visits	Log of Patients
2003 Cohort	44194 [3633]	0.0875 (0.0192)	0.0793 (0.0188)	0.0635 (0.0177)
2004 Cohort	39002 [3073]	0.0857 (0.0213)	0.0770 (0.0204)	0.0777 (0.0196)
2005 Cohort	39721 [3089]	0.1799 (0.0251)	0.1173 (0.0249)	0.1036 (0.0245)

Note. Each cohort regression uses physicians who switched to the Family Health Group (FHG) in a given year as the treatment group and physicians who never switched to the FHG as the comparison group. Bootstrap standard errors in parentheses. Estimates from correlated random trend model which also includes 14 regional indicators and 17 year effects. The sample excludes observations with annual fee-for-service billings below C$30K.

atically between treatment and comparison physicians except for the FHG impact. While this assumption cannot be tested directly, we explore two complementary approaches to further probe the causal interpretation of joining the FHG model.

The first approach is based on the fact that there were three main cohorts of physicians joining the FHG model, in 2003, 2004, and 2005. These cohorts present us with multiple 'experiments' to study the impact of joining the FHG model. While these experiments may not be independent of each other, the consistent results across cohorts support the causal interpretation of the FHG impact.

To explore this issue, we estimated the correlated random trend model for each treatment cohort separately. To facilitate the comparison of estimates across cohorts, we used the same group of comparison physicians in each model. The results are presented in Table 6. The estimated impact for each outcome is quite similar for the 2003 and 2004 cohorts and also to our initial results in Table 4, but the estimated impact for the 2005 cohort is significantly larger. These results are consistent with the interpretation that physicians who had to change their practices the least to meet the FHG contractual requirements joined the FHG model earlier than other physicians. As a result, the estimated impact for the earlier cohorts appears smaller than for the later cohorts. However, despite the variation in the magnitude across different cohorts, the FHG impact remains relatively large and statistically significant for each treatment cohort.

The second approach to infer causality is to examine the dynamics of the estimated FHG impact. Specifically, we estimate the correlated random trend model for each year before and after the physician switches to the FHG model. If joining the FHG model has a causal impact, we expect to observe this impact only in years after the physician switches to the model but not in prior years. The results are shown in Table 7 and Figs. 5–7[23].

The results for the log of annual services and distinct patients closely confirm our prior expectations: the difference between the treatment and comparison physicians is insignificant in all years prior to joining the FHG model, but positive and significant in all years after joining the FHG model. The results for the log of annual visits are quite similar, with insignificant differences in almost all years prior to the switch and significant differences in all years after

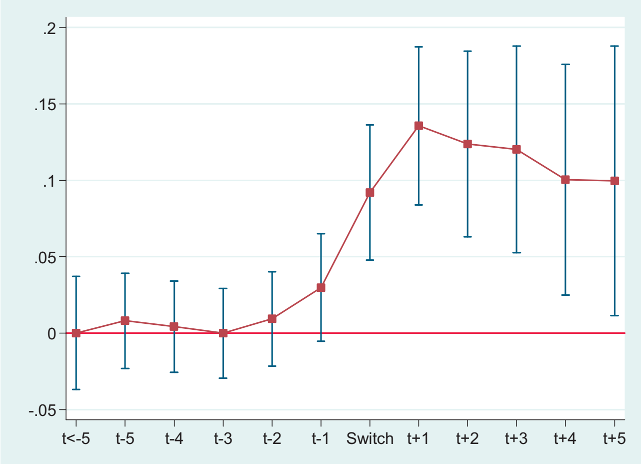

Fig. 5. Log of annual services per physician – 95 percent confidence interval of FHG impact by lead and lag years.

Table 7
Impact by year.

Year	Dependent Variable		
	Log of Services	Log of Visits	Log of Patients
Year $\leq t-6$	0.0001 (0.0189)	0.0060 (0.0173)	0.0181 (0.0226)
Year $t-5$	0.0080 (0.0159)	0.0103 (0.0148)	0.0302 (0.0232)
Year $t-4$	0.0042 (0.0152)	0.0084 (0.0141)	0.0234 (0.0233)
Year $t-3$	−0.0001 (0.0150)	0.0076 (0.0138)	0.0126 (0.0239)
Year $t-2$	0.0093 (0.0158)	0.0219 (0.0148)	0.0229 (0.0254)
Year $t-1$	0.0298 (0.0180)	0.0500 (0.0171)	0.0431 (0.0276)
Year of Switch	0.0920 (0.0226)	0.0917 (0.0217)	0.0836 (0.0319)
Year $t+1$	0.1357 (0.0264)	0.1180 (0.0256)	0.1030 (0.0361)
Year $t+2$	0.1238 (0.0310)	0.1185 (0.0303)	0.1011 (0.0411)
Year $t+3$	0.1203 (0.0345)	0.1139 (0.0343)	0.0957 (0.0451)
Year $t+4$	0.1004 (0.0385)	0.0915 (0.0381)	0.0932 (0.0488)
Year $t+5$	0.0997 (0.0450)	0.0904 (0.0448)	0.1308 (0.0553)

Note. Bootstrap standard errors in parentheses. Estimates from correlated random trend model which also includes 14 regional indicators and 17 year effects. The sample size is 89,741 observations (6929 distinct physicians). The sample excludes observations with annual fee-for-service billings below C$30K.

[23] For presentation purposes, we use a single indicator for observations that are six or more years prior to joining the FHG model. The results are very similar if we use a separate indicator for each of these years.

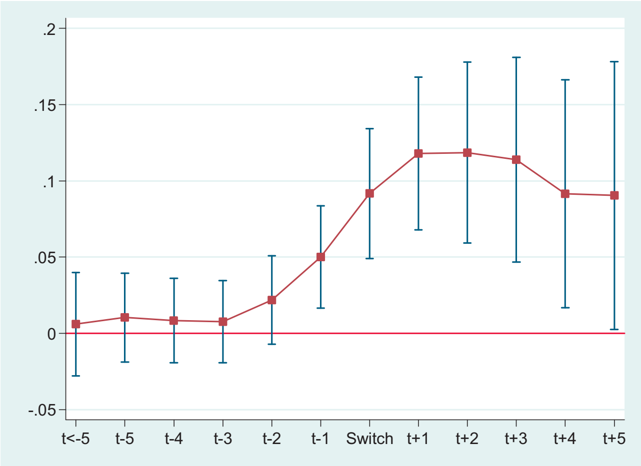

Fig. 6. Log of annual visits per physician – 95 percent confidence interval of FHG impact by lead and lag years.

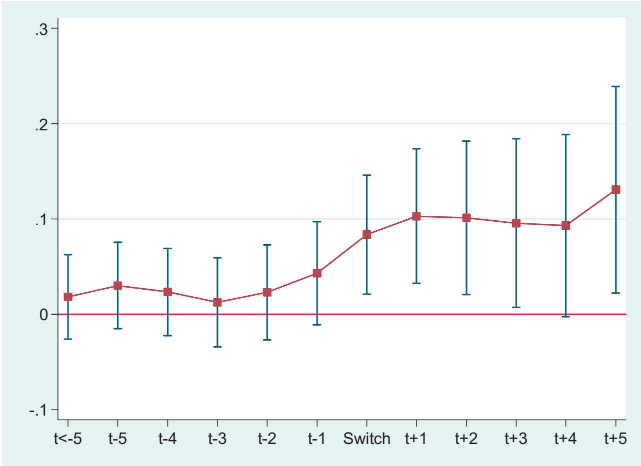

Fig. 7. Log of annual distinct patients per physician – 95 percent confidence interval of FHG impact by lead and lag years.

the switch. However, the difference is significant in the immediate year prior to the switch, suggesting that physicians may have started changing their practice in anticipation of joining the FHG model.

The results also indicate that the adjustment in practice profile was relatively quick, with a complete adjustment taking one to two years, and that the change in practice profile seems permanent, extending to the end of our sample period.

5.3. Alternative samples

Our initial estimates in Table 4 are based on a sample that includes years when physicians were in either FFS or FHG model and had at least C$30,000 in annual FFS billings. In this section, we examine how this sample selection rule affects the interpretation of our results.

The first concern is that many physicians also receive clinical income from alternative payment plans (APPs), such as the Emergency Department Alternative Funding Arrangement. In general, services provided in these plans are not captured in the fee-for-service claims. This raises the concern that changes in physician behaviour in the FFS environment may reflect the impact of changes in the APPs rather than the FHG impact. For example, suppose that the incentives to provide services in the APPs substantially increased over the sample period and that only comparison physicians participate in the APPs. In this case, we may observe a reduction in the FFS claims for the comparison physicians, even though their total clinical services have not changed because they increased their services in the APPs. Using the FFS claims alone, we would incorrectly interpret the resulting differential in outcomes between the treatment and comparison physicians as the FHG impact.

To address this concern, we obtained data on the shadow billing claims that physicians use when providing services in the APPs[24]. We then estimated the correlated random trend model using the sample that excludes years when physicians reported these claims. The results are presented in the first panel of Table 8. The point estimates are very similar to our initial estimates in Table 4. This result can perhaps be explained by the fact that both groups of physicians participate in the APPs and changes in the APPs affect both groups in a similar way. Based on these results, it appears that observed differences in outcomes between the treatment and comparison physicians cannot be entirely explained by physician participation in the APPs.

Another concern with the causal interpretation of the FHG impact is that our results may be driven by an idiosyncratic group of physicians who joined the FHG model but subsequently switched to a harmonized model. This concern is particularly important given that over 1500 FHG physicians switched to the harmonized models between 2007 and 2008. In our previous results, we used samples that exclude years when physicians were not in the FFS or the FHG model. While this restriction facilitates the comparison of the FHG impact relative to the FFS model, the concern is that the estimated impact may reflect a higher productivity of physicians who were in the FHG model only temporarily.

To address this concern, we estimated the correlated random trend model using samples that exclude physicians who switched to a harmonized model at any point during the sample period. The results are presented in the second panel of Table 8. The point estimates are slightly larger than our initial estimates in Table 8, suggesting that the impact was smaller for physicians who subsequently switched to a harmonized model. Importantly, the estimated impact for the included sample of physicians remains relatively large and statistically significant for each outcome.

The last panel of Table 8 presents the results using alternative income cut-offs (C$0, C$10,000K, C$50,000, and C$100,000) for deciding which observations to include in the estimation sample. These results show that the estimated FHG impact decreases monotonically as we use higher income cut-offs. For example, the estimated impact on annual services ranges between 15 percent when we use no income cut-off to 7.2 percent when we use the highest cut-off of C$100,000. Similarly, the estimated impact for annual visits and distinct patients is more than twice as high when using no income cut-off compared to using the cut-off of C$100,000. These results indicate that the magnitude of the estimated FHG impact depends on the choice of specific income cut-off and they are consistent with the interpretation that part-time physicians had to adjust their practices relatively more than full-time physicians to join the FHG model, probably because of the after-hour requirement. However, the overall impact remains positive and statistically significant regardless of which income cut-off is used.

[24] The shadow billing claims, which pay only a fraction of the full value of clinical services, are expected to be of lower quality than the fee-for-service claims. This issue is particularly important when estimating the volume of services based on the shadow billing claims. However, the quality of shadow claims is of less concern in our study because we use these claims only to create an indicator for physicians who had any shadow claims and not to infer the actual volume of services.

Table 8
Alternative samples.

Sample	Sample Size [Physicians]	Dependent Variable		
		Log of Services	Log of Visits	Log of Patients
Excluding Years with Shadow Claims	86362 [6865]	0.0892 (0.0092)	0.0613 (0.0091)	0.0568 (0.0086)
Excluding Switchers to Harmonized Models	63910 [4659]	0.1099 (0.0115)	0.0789 (0.0114)	0.0704 (0.0110)
Income Restrictions				
No Restriction	92748 [6981]	0.1400 (0.0127)	0.1158 (0.0129)	0.1014 (0.0120)
>C$10,000	91512 [6964]	0.1124 (0.0101)	0.0873 (0.0101)	0.0753 (0.0095)
>C$50,000	87818 [6879]	0.0823 (0.0084)	0.0566 (0.0083)	0.0541 (0.0083)
>C$100,000	80140 [6652]	0.0699 (0.0078)	0.0441 (0.0076)	0.0434 (0.0076)

Note. Shadow claims = claims that are not paid at full fee-for-service (FFS) value that physicians use in alternative payment plans. Harmonized models = primary care blended capitation models. Bootstrap standard errors in parentheses. Estimates from correlated random trend model which also includes 14 regional indicators and 17 year effects. The regressions using samples that exclude years with shadow claims and switchers to harmonized models exclude observations with annual fee-for-service billings below C$30K.

5.4. Estimates by age, gender, and location

The results reported in Table 4 represent the average impact of joining the FHG model. In this section, we examine how this impact varies for specific groups of treatment physicians defined by age, gender, and location of practice. In addition to two gender groups, we also sliced the sample into three age groups (less than 41 years, between 41 and 51 years, and over 51 years) and four regional groups (South East, Central, South West, and North).

The results, presented in Table 9, can be summarized by two main points. First, the estimates suggest that the impact is smaller for male physicians relative to female physicians and for physicians over 41 years relative to younger physicians. There is also some regional variation, but with no consistent pattern. However, the estimated impact overall seems reasonably stable across different physician groups. Second, the estimated impact is positive and statistically significant for each physician group and for each outcome, with the exception of the small Northern Ontario sample. These results suggest that the FHG impact was not limited to a particular age or sex group or to a particular region.

5.5. Anatomy of the FHG impact

The results presented so far indicate that the FHG physicians provide more services and visits and see more patients than the comparable FFS physicians. However, it is not clear how the FHG physicians accomplished this improvement. In this section, we examine this question by decomposing the annual number of services into annual days worked and services per day and then analyze the impact of joining the FHG model on each of these components separately.

Table 9
Impact by age, gender, and location.

Sample	Sample Size [Physicians]	Dependent Variable		
		Log of Services	Log of Visits	Log of Patients
Males	61985 [4608]	0.0872 (0.0107)	0.0621 (0.0107)	0.0575 (0.0105)
Females	27756 [2321]	0.1068 (0.0163)	0.0802 (0.0164)	0.0679 (0.0155)
Age in 2002: <41	24513 [2324]	0.0990 (0.0191)	0.0776 (0.0190)	0.0684 (0.0181)
Age in 2002: 41–51	28952 [2093]	0.0842 (0.0134)	0.0632 (0.0134)	0.0558 (0.0133)
Age in 2002: >51	36276 [2512]	0.0869 (0.0123)	0.0539 (0.0121)	0.0502 (0.0122)
South East Ontario	19772 [1577]	0.0630 (0.0160)	0.0542 (0.0159)	0.0504 (0.0173)
Central Ontario	37502 [2792]	0.1124 (0.0128)	0.0703 (0.0125)	0.0580 (0.0115)
South West Ontario	24668 [1898]	0.1007 (0.0205)	0.0799 (0.0204)	0.0729 (0.0204)
Northern Ontario	7799 [682]	0.0672 (0.0342)	0.0614 (0.0349)	0.0729 (0.0329)

Note. Bootstrap standard errors in parentheses. Estimates from correlated random trend model which also includes 14 regional indicators and 17 year effects. The sample excludes observations with annual fee-for-service billings below C$30K.

Table 10
Decomposition of FHG impact on annual services.

Dependent Variable	Coefficient on FHG Indicator	Bootstrap Standard Error
Log of Annual Services	0.0936	(0.0090)
Log of Annual Days	0.0448	(0.0067)
Log of Services per Day	0.0498	(0.0046)
Log of Services Per Day		
All Days	0.0498	(0.0046)
Regular Days only	−0.1388	(0.0163)
After-Hour Days only	0.3419	(0.0159)

Note. Estimates from correlated random trend model which also includes 14 regional indicators and 17 year effects. The sample size is 89,741 observations (6929 distinct physicians). The sample excludes observations with annual fee-for-service billings below C$30K. Regular days refer to days in which no Q012 after-hour indicator was claimed, while after-hour days refer to days in which Q012 indicator was claimed.

Table 11
Impact on targeted and non-targeted services.

Specification	Log of Targeted Services	Log of Non-Targeted Services	Log of Share of Targeted Services
OLS	0.2169	0.3805	0.1286
	(0.0326)	(0.0387)	(0.0306)
Fixed Effects	0.1931	0.3173	0.0712
	(0.0296)	(0.0283)	(0.0274)
Correlated Random Trend	0.0813	0.2092	−0.0045
	(0.0079)	(0.0091)	(0.0054)

Note. Targeted services are comprehensive care services and selected after-hour services that receive a fee premium in the FHG model. Bootstrap standard errors in parentheses. The OLS model includes age, age squared, male, age-male interaction term, expected income gain in 2002, number of working holidays and weekends in 2002, 14 regional indicators, and 17 year effects. The fixed-effect model and the correlated random trend model include 14 regional indicators and 17 year effects. The sample excludes observations with annual fee-for-service billings below C$30K.

The results are presented in the upper panel of Table 10. The results indicate that the 9.8 percent increase in services can be decomposed into 5.1 percent increase in services per day and 4.6 percent increase in annual days. Therefore, it seems that about half of the FHG impact comes simply from the increased days of work.

To further examine the impact on the services per day, we analyzed days when the FHG physicians provided after-hours services ('after-hour' days) and days when they did not ('regular' days)[25]. The results are presented in the lower panel of Table 10 and indicate that the FHG physicians provide about 40 percent more services on their after-hour days and 13 percent fewer services on their regular days compared to the FFS physicians. These results provide some evidence against the hypothesis that the increase in the number of services per day comes predominantly from the increase in the intensity of services, as it is not clear why the FHG physicians would provide more intensive services on the after-hour days only. A more plausible hypothesis is that the FHG physicians work on average more hours per day because of their contractual after-hours obligations. This hypothesis, together with the finding that the FHG physicians work more days per year, suggest that the FHG impact can in large part be interpreted as an increase in the physician labor supply.

Further evidence for this interpretation comes from the analysis of the FHG impact on targeted services (i.e. services receiving comprehensive care and after-hour premiums) and non-targeted services. The results are presented in Table 11 and indicate that the FHG physicians increased their provision of targeted services, as expected. However, the FHG physicians also increased their provision of non-targeted services. Moreover, the results from our preferred correlated random trend model indicate that the share of targeted services for the FHG physicians does not seem to be significantly different than for the FFS physicians. These results suggest that the FHG physicians uniformly increased their provision of services, without significant changes in their service profiles. Again, a most plausible interpretation of these results is that the FHG impact can be interpreted as an increase in physician labor supply, as it is hard to see why the physicians would increase the intensity of providing non-targeted services.

[25] The distinction between two types of days is based on a special indicator, Q012A, that the FHG physicians submit with their claims to trigger the after-hour premiums.

Table 12
Impact on referrals and patient complexity.

Dependent Variable	Coefficient on FHG Indicator	Bootstrap Standard Error
Log of Referrals per Service	−0.0421	(0.0086)
Log of Referrals per Visit	−0.0169	(0.0081)
Log of Referrals per Patient	−0.0098	(0.0092)
Log of Complexity Modifier	0.0278	(0.0019)

Note. Estimates from correlated random trend model which also includes 14 regional indicators and 17 year effects. The sample size is 89,741 observations (6929 distinct physicians). The sample excludes observations with annual fee-for-service billings below C$30K.

5.6. *Impact on referrals and patient selection*

Although our results indicate that the FHG physicians provide more services and visits and see more distinct patients, other considerations are also important. One such consideration is the impact of remuneration scheme on the referrals to specialists. From the perspective of public purse, the increased throughput of FHG physicians would obviously have less value if it came at the cost of increased referrals to specialists. To address this issue, we estimated the correlated random trend model using the log of referral rates as our dependent variable. The results are presented in the first panel of Table 12. The results indicate that the referral rates per service, visit, or distinct patient are all significantly lower for the treatment physicians relative to the comparison physicians, by about one to four percent. These results suggest that joining the FHG model does not have an unintended consequence of increasing the referral rates to specialists. The actual decline in the referral rates can perhaps be explained by the wider range of comprehensive care services that the FHG physicians are required to provide compared to the FFS physicians.

The other important consideration is the impact of remuneration scheme on the type of patients that family physicians see. Of particular concern is whether the increase in physician throughput comes at the cost of limiting access to more complex patients. To explore this issue, we calculated the average patient complexity for each physician using the following measure:

$$\frac{\sum v_{as} m_{as}}{\sum v_{as}} \tag{13}$$

where v is the number of visits by patients in age group a and of sex s, and m is the age-sex specific complexity modifier used by the Ministry of Health and Long-Term Care to adjust the Comprehensive Care Capitation fee.

We then estimated the correlated random trend model using the log of this measure of patient complexity as our dependent variable. The results are presented in the second panel of Table 12. The point estimate indicates that the average patient complexity is about three percent higher for the treatment physicians than for the comparison physicians. Therefore, to the extent that age and sex of patients appropriately capture patient complexity, it appears that the treatment physicians actually treat slightly more complex patients than the comparison physicians. This finding can perhaps be explained by the built-in incentives in the Comprehensive Care Capitation fee to enroll more complex types of patients. Alternatively, the more complex patients could gravitate to the FHG physicians because of the wider range of services these physicians provide or because of the enhanced access through the after-hours requirement. In both cases, the enrolment of patients is voluntary and occurs by the mutual agreement of both the patient and the physician.

6. Conclusion

In this paper, we compare productivity of primary care physicians in a new payment model introduced in Ontario known as the Family Health Group to the traditional fee-for-service model. Our results indicate that the FHG physicians provide more services and visits and treat more patients than the comparable FFS physicians.

We believe that these results are important for at least two reasons. First, the results show that how physicians are paid can affect their total productivity. This insight is particularly important in jurisdictions experiencing physician shortages, which are sometimes exacerbated by an aging physician population, an increasing proportion of female physicians, and changing preferences and personal constraints of younger physicians[26]. Second, the results show that physician productivity can improve despite significant pay increases in the new payment models[27]. Specifically, most new payment models provide financial incentives for quality outcomes, but these incentives have opposing income and substitution effects on physician labour supply. As a result, incentives targeted to improve the quality of care may actually reduce the quantity of care depending on the relative strength of income and substitution effects.

However, there are two main limitations to our study. First, the FHG model is an enhanced fee-for-service model that includes a combination of payment incentives for improving patient access and quality of care. As in many other program evaluation studies, we cannot separately identify which of the payment incentives contributed to the change in physician behaviour and to what extent. However, the combination of payment incentives in the FHG model is sufficiently common in many recent primary care reforms to make us believe that our study can be relevant for many other jurisdictions.

The second limitation of our study is that we focus on outcomes that measure the quantity aspect of health care only. Therefore, improved physician productivity may not necessarily be welfare improving if additional care is inappropriate and unnecessary. While this quantity-quality trade-off represents an important consideration in any study of the impact of payment incentives on physician behaviour, good data on quality of care are hard to come by. However, we believe that the quality concerns about the FHG model can be attenuated by at least three considerations. First, our results suggest that the FHG impact can be interpreted in large part as an increase in the physician labour supply. Second, we show that the FHG model does not have adverse effects on the patient case-mix or on the referrals to specialists. Lastly, the complementary evidence from Li et al. (2010) documents that the 'quality' outcomes in the primary care models in Ontario, as measured by preventive care and special service targets, is no worse and sometimes better than in the traditional FFS model.

Future research can build on our analysis in at least two ways. First, we document that the FHG model improves physician productivity, but it is also important to understand how the model affects the cost of delivering primary health care. Second, our analysis focuses on the transition of physicians from the FFS model to the FHG model. Future research could consider the entire spectrum of payment models for primary care physicians, focusing on determinants of transition between the models and the impact of this transition on physician behaviour.

Acknowledgements

We thank the editor, two anonymous referees, and seminar participants at York University in Toronto, Canada, McMaster University in Hamilton, Canada, and the 2010 Annual Meeting of the Canadian Health Economics Study Group in Montreal, Canada for useful comments. As usual, all errors are ours.

Appendix A.

Comparative statics when the after-hours is not binding

The first-order necessary conditions for the interior solution to the problem described by (1)-(4) are:

$$u_C p_1 - t_1 u_l = 0$$

$$u_C p_2 - t_2 u_l + u_2 = 0$$

Totally differentiating these conditions and rearranging, we have

$$\begin{bmatrix} u_{cc}p_1^2 + u_{ll}t_1^2 & u_{cc}p_1p_2 + u_{ll}t_1t_2 \\ u_{cc}p_1p_2 + u_{ll}t_1t_2 & u_{cc}p_2^2 + u_{ll}t_2^2 + u_{22} \end{bmatrix} \begin{bmatrix} \mathrm{d}x_1 \\ \mathrm{d}x_2 \end{bmatrix}$$

$$= -\begin{bmatrix} u_{cc}p_1x_1 + u_c & u_{cc}p_1x_2 & u_{cc}p_1 \\ u_{cc}p_2x_1 & u_{cc}p_1x_1 + u_c & u_{cc}p_2 \end{bmatrix} \begin{bmatrix} \mathrm{d}p_1 \\ \mathrm{d}p_2 \\ \mathrm{d}m \end{bmatrix}$$

Let D denote the determinant of the matrix of coefficients associated with changes in the endogenous variables. By the second-order condition, $D > 0$.

With negligible income effects ($u_{cc} \to 0$), by Cramer's rule we have:

$$\frac{\partial x_1}{\partial p_1} = -u_c(u_{ll}t_2^2 + u_{22})/D > 0$$

$$\frac{\partial x_2}{\partial p_2} = -u_c u_{ll} t_1^2/D > 0$$

$$\frac{\partial x_i}{\partial p_j} = u_c u_{ll} t_1 t_2/D < 0$$

The total change in x_1 and x_2 is then given by:

$$\Delta(x_1 + x_2) \propto u_{ll}(t_2 - t_1)[t_1\Delta p_2 - t_2\Delta p_1] - u_{22}\Delta p_1$$

where the factor of proportionality is u_cD.

We want to show that a sufficient condition for $\Delta(x_1+x_2) \geq 0$ is that $t_2 \leq t_1$. First, note that $-u_{22}\,\Delta p_1 \geq 0$ by the model assumption. Second, $t_1\,\Delta p_2 - t_2\,\Delta p_1 = ab(t_1k_2p_2 - t_2k_1p_1)$. From the first-order conditions, we have that $p_1 = p_2t_1/t_2 + u_2t_1/t_2u_C$. Therefore, $t_1k_2p_2 - t_2k_1p_1$ can be expressed as $t_1p_2(k_2 - k_1) - k_1t_1u_2/u_C$. This term is positive because in the FHG model $k_2 = 0.2 > k_1 = 0.1$. Therefore, $t_1\,\Delta p_2 - t_2\,\Delta p_1 > 0$. Lastly, provided that $t_2 \leq t_1$, we have that $\Delta(x_1+x_2) \geq 0$.

[26] For the relative importance of these factors in explaining labour supply of Canadian family physicians, see Crossley et al. (2006).

[27] For example, the FFS physicians can increase their income in the range of 10–20 percent by switching to the FHG model.

Comparative statics when the after-hours is binding

The first-order necessary conditions for the corner solution to the problem to the problem described by (1)-(4) are:

$$u_C p_1 - t_1 u_l = 0$$

$$u_C p_2 - t_2 u_l + u_2 + \lambda = 0$$

$$x_2 - x_{\min} = 0$$

Totally differentiating these conditions and rearranging, we have

$$\begin{bmatrix} u_{cc}p_1^2 + u_{ll}t_1^2 & u_{cc}p_1p_2 + u_{ll}t_1t_2 & 0 \\ u_{cc}p_1p_2 + u_{ll}t_1t_2 & u_{cc}p_2^2 + u_{ll}t_2^2 + u_{22} & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} dx_1 \\ dx_2 \\ d\lambda \end{bmatrix}$$

$$= - \begin{bmatrix} u_{cc}p_1x_1 + u_C & u_{cc}p_1x_2 & u_{cc}p_1 & 0 \\ u_{cc}p_2x_1 & u_{cc}p_1x_1 + u_c & u_{cc}p_2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} dp_1 \\ dp_2 \\ dm \\ dx_{\min} \end{bmatrix}$$

Let J denote the determinant of the matrix of coefficients associated with changes in the endogenous variables. By the second-order condition, $J > 0$.

With negligible income effects ($u_{cc} \to 0$), by Cramer's rule we have:

$$\frac{\partial x_1}{\partial p_1} = u_c/J > 0$$

$$\frac{\partial x_2}{\partial p_2} = -u_{ll}t_1^2/J > 0$$

$$\frac{\partial x_i}{\partial p_j} = 0$$

$$\frac{\partial x_1}{\partial x_{\min}} = u_{ll}t_1t_2/J < 0$$

$$\frac{\partial x_2}{\partial x_{\min}} = -u_{ll}t_1^2/J > 0$$

The total change in x_1 and x_2 is then given by:

$$\Delta(x_1 + x_2) \propto u_C\,\Delta p_1 - u_{ll}t_1^2\,\Delta p_2 + u_{ll}(t_1t_2 - t_1^2)\,\Delta x_{\min}$$

where the factor of proportionality is J.

The first two terms on the right-hand side are non-negative. The third term is also non-negative, provided $t_2 \le t_1$. Therefore, we have that $\Delta(x_1 + x_2) \ge 0$.

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jhealeco.2010.10.005.

References

Campbell, S., Reeves, D., Kontopantelis, E., Middleton, E., Sibbald, B., Roland, M., 2007. Quality of primary care in England with the introduction of pay for performance. New England Journal of Medicine 357 (2), 181–190.

Crossley, T.F., Hurley, J., Jeon, S., 2006. Physician labour supply in Canada: a cohort analysis. McMaster University. Department of Economics, Working Paper Series, 2006-02.

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

Devlin, R.A., Sarma, S., 2008. Do physician remuneration schemes matter? The case of Canadian family physicians. Journal of Health Economics 25 (7), 1168–1181.

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.

Li, J., Hurley, J., DeCicca, P., Buckley, G., 2010. Physician response to pay-for-performance—evidence from a natural experiment. In: Presentation at the Canadian Health Economics Study Group meeting, Montreal, May 26-27.

Leuven, E., Sianesi, B., 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.

McGuire, T., Pauly, M., 1991. Physician response to fee changes with multiple payers. Journal of Health Economics 10, 385–410.

McGuire, T., 2000. In: Culyer, A.L., Newhouse, J.P. (Eds.), Physician Agency. Handbook of Health Economics, vol. 1. Amsterdam, North-Holland.

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, P.R., Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70 (1), 41–55.

Rosenbaum, P.R., Rubin, D.B., 1985. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. The American Statistician 39 (1), 33–38.

Rosenthal, M.B., 2010. Impact of a medical home intervention on utilization and spending. In: Presentation at the American Economic Association meeting, Atlanta, January 3-5.

Rosser, W.W., Colwill, J.M., Kasperski, J., Wilson, L., 2010. Patient-centered medical homes in Ontario. New England Journal of Medicine 362 (3), e7(1)–e7(3).

Russell, G.M., Dahrouge, S., Hogg, W., Geneau, R., Muldoon, L., Tuna, M., 2009. Managing chronic disease in ontario primary care: the impact of organizational factors. Annals of Family Medicine 7, 309–318.

Scott, A., 2000. In: Culyer, A.L., Newhouse, J.P. (Eds.), Economics of General Practice. Handbook of Health Economics, vol. 1. Amsterdam, North-Holland.

Smith, P.C., York, N., 2004. Quality incentives: the case of U.K. general practitioners. Health Affairs 23 (3), 112–118.

Sutton, M., Elder, R., Guthrie, B., Watt, G., 2010. Record rewards: the effects of targeted quality incentives on the recording of risk factors by primary care providers. Health Economics 19, 1–13.

Wooldridge, J.M., 2005. Fixed-effects and related estimators for correlated random-coefficient and treatment-effect panel data models. The Review of Economics and Statistics 87 (2), 385–390.