Physicians' Multitasking and Incentives: Empirical Evidence from a Natural Experiment*

Etienne Dumont†, Bernard Fortin†, Nicolas Jacquemet‡, Bruce Shearer§

June 11, 2008

Abstract

We analyse how physicians respond to contractual changes and incentives within a multitasking environment. In 1999 the Quebec government (Canada) introduced an optional mixed compensation system, combining a fixed *per diem* with a partial (relative to the traditional fee-for-service system) fee for services provided. We combine panel survey and administrative data on Quebec physicians to evaluate the impact of this change in incentives on their practice choices. We highlight the differentiated impact of incentives on various dimensions of physician behaviour by considering a wide range of labour supply variables: time spent on seeing patients, time devoted to teaching, administrative tasks or research, as well as the volume of clinical services and average time per clinical service. Our results show that, on average, the reform induced physicians who changed from FFS to MC to reduce their volume of (billable) services by 6.15% and to reduce their hours of work spent on seeing patients by 2.57%. Their average time spent per service increased by 3.58%, suggesting a potential quality-quantity substitution. Also the reform induced these physicians to increase their time spent on teaching and administrative duties (tasks not remunerated under the fee-for-service system) by 7.9%.

Keywords: physician payment mechanisms; multitasking; mixed-payment systems; incentive contracts; labour supply; self-selection; panel estimation.

JEL classification: I10, J22.

*The authors thank the *Collège des médecins du Québec* for making its survey data available and the *Régie d'assurance-maladie du Québec* and Marc-André Fournier for the construction of the database. They are also grateful to André-Pierre Contandriopoulos and Philippe Joncas for helpful discussions, and to Yvon Brunelle, Anthony J. Culyer (the editor), Rose-Anne Devlin, Pierre-Thomas Léger, and two anonymous referees for their very useful comments on an earlier draft. They also acknowledge financial support from the Canadian Institute of Health Research, the Canada Research Chair in the Economics of Social Policies and Human Resources and the Fonds québécois de la recherche sur la société et la culture.

†Département d'économique, Université Laval, CIRPÉE and CIRANO; ✉etidumont@hotmail.com; bernard.fortin@ecn.ulaval.

‡University Paris 1 Panthéon–Sorbonne and Paris School of Economics; ✉nicolas.jacquemet@univ-paris1.fr.

§Département d'économique, Université Laval, CIRPÉE, CIRANO and IZA; ✉Bruce.Shearer@ecn.ulaval.ca.

1 Introduction

The labour-supply behaviour of physicians is an important determinant of the performance of any health-care system. Physicians make decisions on both the extensive margin, affecting their time spent at work, and the intensive margin, affecting the number of services provided. What is more, physician services are provided within a multitasking environment (Ma, 1994; Ma and McGuire, 1997) – decisions over the number of patients treated, the time spent seeing patients and the time devoted to teaching and to administrative tasks affect the quantity as well as the quality of health care supplied. Consequently, employers of physician services (in both the private and public sector) must be concerned with the efficient allocation of physician time and effort across different tasks.

One policy instrument which is available to address these issues is compensation design. Economists have written extensively on payment systems and their effects on individual behaviour and performance (*e.g.*, Hart and Holmstrom, 1987; Milgrom and Roberts, 1992). Within the health-economics literature, many authors have highlighted the different incentives in commonly used payment systems. For example, capitation contracts can reduce the utilization of health services (Pauly, 1990), yet can lead to the under-provision of necessary services (Blomqvist, 1991) and to the cream-skimming of patients (Newhouse, 1996).[1]

Empirical evidence suggests that physicians do respond to incentives. Several studies show that fee-for-service physicians (who receive a fee for each clinical service provided) perform more clinical services than do physicians paid a salary or capitation contract; examples include: Gaynor and Gertler (1995); Gruber and Owings (1996); Barro and Beaulieu (2003).[2] Yet, the interpretation of these results is unclear. While the increase in services can be beneficial to patients waiting for treatment, such high-powered incentive contracts, that closely link remuneration to quantity of care, may cause physicians to see too many patients and to neglect other tasks such as the quality of care and non-clinical activities such as teaching and administrative duties.[3] Theoretical work suggests that low-powered incentive contracts can be optimal when some tasks are unobservable or observable only at very high costs (*e.g.*, Holmstrom and Milgrom, 1991; Baker, 1992). These contracts are often referred to as "mixed-compensation" contracts in the physician compensation literature (Eggleston, 2005). They usually combine an upfront payment (per unit of time or per patient) with a partial fee-for-service. Such contracts reduce the opportunity cost of spending time with patients and performing administrative tasks. Their potential benefit is to encourage the provision of aspects of services (such as the quality of care) which are valued by the patient but are unobservable to them and the (public or private) insurer (Ma and McGuire, 1997). As shown by these authors, the effect of these contracts depends on whether the quality of care and observable services are complements or substitutes. As long as they are substitutes, increasing the upfront payment for the quality of care (*e.g.*, introducing a *per diem*) and a lower marginal payment for the quantity of services, will encourage the physician to provide a better quality of care. Also, if time spent performing

[1] Under capitation contracts physicians receive a fixed payment per patient (per period), independent of the services they provide.

[2] See Léger (2007) for a recent review of the empirical evidence on physician compensation systems.

[3] Another concern is induced demand which can be exacerbated under high-powered incentive contracts. See, for example, Gruber and Owings (1996).

non-clinical services such as administrative duties is remunerated under the upfront payment, an increase in this parameter may stimulate these activities.

To date, empirical work on multi-tasking issues has been limited by data constraints.[4] Most data sets contain only narrow measures of physician behaviour; *e.g.*, the number of patients seen by a physician (Gaynor and Gertler, 1995), the hospitalization rate of patients (Hillman et al., 1989) or the rate of antibiotic prescription (Hutchinson and Foley, 1999). Yet, evaluating compensation contracts within a multi-tasking context requires measuring physician behaviour along multiple dimensions. This paper presents a step in that direction. We construct a panel of survey and administrative data to analyse the multi-dimensional response of physicians to changes in their contracts (and incentives).

Our data come from the province of Quebec, Canada. They provide detailed information on the labour-supply behaviour of specialist physicians working in the province over a five-year period (1996-2000), including the number of services each physician provided, the weeks and hours each physician worked (disaggregated into several activities), individual earnings, as well as personal characteristics. From these data, it is also possible to construct a proxy for the quality of care provided by physicians: the average time per clinical service.[5] Importantly, our sampling period covers a major policy change with respect to specialist physician pay. Quebec specialists were traditionally paid a fee-for-service (FFS) contract. In 1999, the government introduced a mixed compensation system (MC) as an alternative to the FFS system. Physicians who are paid under MC receive a base wage (or *per diem*), independent of the number of services provided, and a reduced, or prorated, (vis-à-vis the FFS contract) fee-for-service. We define the incentive ratio as the prorated fee-for-service, paid under MC, divided by the original fee-for-service, paid under FFS. Incentive ratios less than one imply reduced incentives. The incentive ratio averaged 41% in our sample[6], suggesting that the MC system substantially altered physician incentives.

We exploit this contractual change to measure the effect of incentives on various aspects of physician labour supply: services provided, hours devoted to seeing patients, hours devoted to teaching and to administrative tasks and income. It is noteworthy that the MC system is optional and applies only to specialist physicians when working in health establishments (mainly hospitals). By 2000, 31.03% of specialists were paid, at least in part, under the MC system. As discussed in Fortin et al. (2008), we expect the reform to induce physicians who switch from FFS to MC to reduce their volume of clinical services but also to increase their quality of care and the time they spend on non clinical activities remunerated under the *per diem*.

Attributing changes in observed physicians' behaviour to changes in the compensation policy requires controlling for a competing hypothesis that can explain the observed changes; namely that different types of physicians are attracted to different compensation policies - a phenomenon known as *adverse selection* in the theoretical literature and *self selection* in the empirical literature (Chiappori and Salanié, 2003). Thus, one would expect the physicians who have a low productivity (in terms of quantity of care provided)

[4]Paarsch and Shearer (2000) treats multi-tasking issue within a manual-labour context.

[5]Ma and McGuire (1997) suggests the use of this variable as a proxy for the intensity or quality of treatment provided by the physician. Of course, the measure is imperfect: it does not distinguish between time spent with patients and time spent between patients, nor does it provide any information on its actual effect on the health of patients.

[6]The average incentive ratio is the weighted average fraction of FFS prices paid for clinical services when performed under MC (the weights being the relative importance of each service in 1998).

to be more inclined to choose the MC system, *ceteris paribus*. The longitudinal aspect of our data allows us to observe individual physicians under both the FFS and the MC payment systems. We use this information to isolate changes in behaviour resulting from the change in the compensation system from self-selection. More precisely, we estimate fixed effect regression models which can be regarded as a generalization of a difference-in-differences approach. Panel models with censored data are also used when necessary.

Our results suggest that physicians did react to the change in incentives resulting from the switch to MC, reducing the (billable) volume of services provided by 6.15%. There is also evidence that the change had important effects on other dimensions of physician labour supply and addressed, at least partially, multi-tasking problems associated with FFS: the time spent per service increased by 3.58% under the MC system and the hours devoted to activities not remunerated under FFS (administrative and teaching) increased by 7.9%. These results are qualitatively robust to various specifications of our empirical model and sample composition, although there is evidence that the reaction to the reform varied across gender and specialty.

The rest of the paper is organized as follows. The next section describes the institutional details governing physician compensation in Quebec. Section 3 describes the data, the variables, the treatment and control groups used in our empirical analysis. It also presents descriptive statistics. Section 4 explains our empirical strategy, while Section 5 discusses the results. Section 6 provides a series of robustness checks to our empirical specification and the last section concludes.

2 Context

Health care falls under the jurisdiction of provincial governments under the Canadian constitution. While the federal government does exert some influence on conformity to national standards through the Canada Health Act, policies remain largely a provincial responsibility. Within the province of Quebec, the vast majority of specialist physicians have been paid under a FFS compensation system until 1999.[7] The fees paid are service specific, accounting for the difficulty and time intensiveness of the service, and apply to practices both in the public sector and in private clinics, although the latter are administered in the private sector. They are periodically negotiated between the Quebec Government and the *Fédération des médecins spécialistes du Québec*.

2.1 The Mixed Compensation scheme

In September 1999, the Quebec Government introduced the MC system for 19 out of 31 specialties. The scheme became accessible to other specialties later in 2000. Adoption of the MC system was optional. In hospitals, specialists work within departments, made up of physicians performing similar tasks. Upon the introduction of the MC system, each department voted on its adoption, switching to the MC system only if the vote passed unanimously.

[7] FFS compensation represented about 80% of specialists' earnings in 1999. The remaining 20% come from sessional hours in health care establishments (almost 10%), earnings in laboratories (9%) and a very small part from salary (1%).

The MC scheme combines a *per diem*, paid independently of the number of services provided, with a prorated FFS payment. The incentive ratio (the prorated fee divided by the original FFS fee) averaged 41%, substantially altering the monetary incentives to supply services. Note as well, the incentive ratio varied across services and specialties. At the extreme, some services were deemed *non-billable* under the MC system, receiving no fee-for-service (*e.g.*, most follow-up visits).

A number of additional conditions bear mention. First, the *per diem* only applies to certain activities, principally time spent on administration, teaching and seeing patients, the most notable exclusion being time spent on research. Research activities are typically paid by the hospital where they take place, or through government grants. Second, the MC system is available only for activities completed in health establishments (mainly hospitals). Hours of work and services provided within private clinics continue to be paid under the FFS system.[8] Third, services provided within hospital emergency rooms by non emergency-room physicians are excluded from the MC system. Finally, there are restrictions on the number of *per diems* a physician can claim and the time-period during which he/she can claim them. Half *per diems* are claimed on a 3.5-hour basis.[9] The maximum number of half *per diems* that a physician can claim during a two-week period is 28 and these can only be claimed Monday to Friday between 7AM and 5PM. Once the maximum number of *per diems* is reached, or when a physician works outside the *per-diem* claimable hours, he/she is paid on the FFS basis.

In introducing the reform the government targeted three general goals. First, the government sought to improve the quality of health care, increasing the amount of time physicians spend with each patient. Second, the government sought to improve efficiency in the quantity of health care, reducing the marginal gain to seeing patients unnecessarily. Finally, the government sought to induce physicians to increase the amount of time devoted to teaching and administrative duties and to reduce inequities between physicians devoting their time to these duties and those who do not.

2.2 Expected effects of the reform

The multitasking nature of the supply of health services, whereby physicians can adjust both the quantity and quality of services, creates non-linearities in the physicians' budget constraint. This is due to the fact that the marginal benefit of an hour worked depends on the effort (or services) supplied – the opportunity cost of leisure is not constant. This complicates the comparative-static analysis of physician labour-supply behaviour. Fortin et al. (2008) provide an extensive theoretical analysis of the reform, focusing on the quality/quantity trade-off in the supply of health services. Their results highlight the general ambiguity of comparative static results within this setting. Nevertheless, economic intuition suggests that the reduction in the fee-for-service under MC would induce physicians who switch to MC to reduce their volume of clinical services (and clinical hours of work) and to increase the time they spend per clinical service (quality of care). The introduction of the *per diem* which remunerates

[8] More precisely, physicians who demonstrate that there is a strong complementarity between private clinic and hospital activities can choose to receive a *per diem* when working in a private clinic. However, in this case, they would not receive any FFS payment for services performed in the clinic. Therefore most physicians choose to be paid on the FFS basis when working in a private clinic.

[9] During the period under study, the half *per diem* was set equal to 300 CAN$.

time spent on administrative and teaching duties can also be expected to increase time spent on these activities. Fortin et al. (2008) derive sufficient conditions under which the reform will induce these effects as well as inducing physicians who switch to reduce their total (clinical and non-clinical) hours of work. However, while these effects are intuitive, the sufficient conditions supporting them are not; they rely on a complex interaction between the non-linearities in the budget constraint and the indifference curves. The ambiguities that are inherent in such non-linear models give added importance to empirical work analyzing the effects of the reform.[10]

3 Data and Variables

3.1 Data

Our data contains information on the practice patterns and individual characteristics of physicians practising in Quebec between 1996 and 2002.[11] These data come from two sources. First, each year, the *Collège des médecins du Québec* (CMQ) conducts a survey of its members. This survey contains information on individual characteristics such as specialty, age and gender, as well as establishment and geographic characteristics. The survey also contains time-use information, namely, time spent at work, time devoted to seeing patients in establishments or private clinics, time devoted to teaching, administrative duties and research activities (all measured as hours per working week, averaged over the whole year).

The fact that we use self-reported measures of physicians' use of their time raises the question of measurement errors and its consequences for the empirical analysis. Two potential problems arise. First, we may be using imprecise measures of the hours actually worked by physicians. This will lead to consistent but less precise estimates of the coefficients of the model, something shared by most empirical research on labour supply (*e.g.*, Blundell and MaCurdy, 1999). A second, more serious, difficulty will occur if the measurement errors in the self-reported data are correlated with the explanatory variables. In particular, if switching to MC induces a distortion in the way hours are reported, the estimates of the impact of the reform will be inconsistent, unduly attributing the change in reported hours (for a given level of actual hours of work) to the reform itself. However this problem is unlikely to occur here since our time-use data were collected by a professional organization (CMQ) which is not in any way responsible for physician compensation. Although physicians may have an incentive to misrepresent their hours of work to the organisation that administers their pay (the *Régie de l'assurance-maladie du Québec*, RAMQ), they have no incentive to do so to the CMQ.

[10] The model in Fortin et al. (2008) was developed to analyse this particular reform. It allows a much richer choice set for physicians than in Ma and McGuire (1997), including services performed, hours worked and on-the-job leisure. This added complexity explains the lack of intuition over the results.

[11] We ignore data from 2001 and 2002 in our analysis since there was an increase in fees paid to physicians (in the order of 11%, on average) at the beginning of 2001. Given we are not estimating a structural model, adding these data would not likely add to our analysis of the effect of the reform but would rather identify the difference in reaction to the changes in fees between the treatment and the control groups. While of interest in itself, we prefer to leave such an analysis for future research.

Our second source of data is the RAMQ. This source provides very precise information on the practice intensity of physicians – our data come directly from the administrative files used to compensate physicians. In particular, we observe the income and the number of services provided on a quarterly basis for each physician. The data from the annual surveys and from the RAMQ were matched on the basis of an anonymous number attributed to each physician. This also allowed us to keep track of each physician in our sample across time periods.

3.2 Variables

Typically, physicians provide a variety of services, each remunerated at (possibly) different rates. To construct a measure of the volume of services supplied, we aggregated services together to generate an annual quantity index of services (weighted by the basic fees in 2000).[12] Recall that the government designated some services as non-billable under the MC system. Since our data come from the billing records of physicians, these services are unobservable when they are performed under MC. We have therefore decided to exclude them from our measure of clinical services for all physicians over the sampling period, concentrating solely on billable services. Including non-billable services would have created an upward bias in the impact of the reform on the volume of services since they disappear after the reform. As long as the reform does not induce physicians under MC to substitute too much across various clinical services, our measure of the effect (in %) of the reform on the volume of total clinical services will not be strongly affected by their exclusion. If the reform induces MC physicians to substitute billable for non-billable clinical services, our results can be interpreted as a lower bound of its effect (both in % and in absolute value) on the volume of total clinical services.

Our income variable differs from the gain the physician derives from practice due to two further institutional factors: income ceilings and regional-practice compensating differentials. Two types of income ceilings were active during the sampling period. The first is a ceiling on the net income a specialist could earn in a private clinic during a given semester. The net income on which the ceiling applies is defined as 65% of income earned by the physician in private clinics and allows the physician to net out expenses. Second, up until 1999, the government also imposed a ceiling on gross income (independent of where it was earned). In both cases, once the relevant income attains the ceiling, payments to physicians are reduced by 75% from their regular levels.[13] Regionally differentiated compensation rates apply to the resulting income. A premium was offered to younger physicians settled in some designated regions during the period under study (see Bolduc et al., 1996, for more detail).

3.3 Definition of Sample, Treatment, and Control Groups

Since we focus on a comparison between the FFS and MC schemes, we dropped from our sample all physicians who received less than 100% of their earned income under these schemes over the sample pe-

[12] The formula for calculating the volume of services for a physician i at year t is given by $\Sigma_{j=1}^{m} p_j^{2000} S_{ij}^t$, where m is the number of different services, p_j^{2000} is the basic fee for service j in 2000, and S_{ij}^t is the number of services j performed by physician i at time t.

[13] We show in Section 6.4 that our results should not be affected by this rule.

riod. Physicians in specialties, such as geriatrics, psychiatry and public health, who receive an important part of their income from salary were therefore excluded. We also excluded physicians who are never observed under the FFS system. This excludes some young physicians who begin their career during the sample period and are only observed under the MC system.

In order to analyse the effect of the reform, we divided our sample into treatment and control groups. Our definition of these groups centered around the physician's participation in the MC system once it became available in the year 2000. A natural criterion would have been to condition treatment on being remunerated under MC. However, due to the particular setting of the reform (see Section 2), changing from FFS to MC does not exclude practicing under FFS (*e.g.*, during particular hours of the day, in emergency rooms or in private clinics). We thus defined the treatment group as those physicians who were compensated *in part* under MC in every quarter of 2000. Similarly, the control group is defined as those physicians being remunerated exclusively by FFS in every quarter of 2000. Physicians who did not match these criteria were excluded from the sample.[14] Such a restrictive definition of the treatment and control groups is aimed at building as homogeneous groups as possible.[15] This same objective also lead us to exclude some observations from the study. First, the following laboratory specialties were excluded: anatomopathology, medical biochemistry, medical microbiology and nuclear medicine. Moreover, the MC scheme was introduced later in 2000 for cardiac and vascular surgery, internal medicine, otorhinolaryngology, physiatry, radiology and urology. Therefore these specialties were also dropped from our analysis. Finally, ophthalmology was removed from our sample since very few physicians of this specialty chose MC. Their participation rate to MC was only 2% in 2000. We are, therefore, left with the following specialties in our sample : anesthesiology, cardiology, dermatology, endocrinology, gastroenterology, general surgery, gynecology and obstetrics, nephrology, neurosurgery, neurology, orthopaedics, pediatrics, pneumology, radio-oncology, and rheumatology. The final sample contains 2 120 physicians, down from 8 549 specialist physicians in the initial database.[16]

3.4 Descriptive Statistics

Table 1 provides information concerning the number of physicians in each specialty in our sample as well as the participation rates in the MC scheme (the proportion of physicians who switched to MC) in 1999 and 2000. Of the 2 120 physicians in our sample, 423 are in the treatment group (they switched to MC) and 1 697 are in the control group (they remained in FFS). There are 1 655 males and 464 females. Cardiology (260 physicians), General surgery (259 physicians) and Pediatrics (253 physicians) are the largest specialties, while Radiation Oncology (30 physicians), Neurosurgery (31 physicians) and Rheumatology (44 physicians) are the smallest specialties.

[14] To account for changes in practice behaviour in 1999 for the physicians who changed from FFS to MC in the last quarter of 1999, we introduced a dummy variable identifying these treated individuals in the regressions for clinical services and income since those variables are measured over calendar years. By contrast, time-use variables do not overlap with the reform in 1999 since the CMQ surveys are conducted in July while the reform was introduced in September.

[15] In Section 6.1, we show our results are quite robust with respect to other definitions.

[16] The specialities included in our sample accounts for close to 50% of total specialities, over the period covered by our initial data base (see the first column of Table 1).

Table 1: Mixed Compensation System characteristics in 1999 and 2000

	Percentage of total specialists	MC participation rate 1999	MC participation rate 2000	Number of physicians	MC Incentive Ratio Excluding NBS	MC Incentive Ratio Including NBS
Anesthesiology	7.61	23.12	32.75	229	31.01	30.90
Cardiology	4.18	2.20	6.92	260	44.32	34.45
General Surgery	6.77	14.29	18.15	259	48.10	35.33
Orthopedic Surgery	4.02	6.96	9.24	184	45.87	38.84
Dermatology	2.26	16.39	17.99	139	53.53	32.98
Gastroenterology	1.99	1.82	2.59	116	75.00	31.24
Obstr. + Gynec.	5.17	4.08	6.39	219	40.04	37.14
Pneumology	2.12	5.19	5.95	84	79.00	18.60
Neurosurgery	0.84	75.86	83.87	31	70.66	52.99
Neurology	2.51	9.09	10.83	120	28.60	24.01
Pediatrics	7.46	27.88	43.48	253	30.00	14.27
Radiation Oncology	0.75	47.62	70.00	30	79.00	54.00
Nephrology	1.50	4.41	5.19	77	34.00	21.43
Endocrinology	1.67	22.73	24.00	75	28.00	20.30
Rheumatology	0.91	41.46	61.36	44	45.23	36.12
Total	49.77	14.04	19.95	2120	41.19	30.80

Note: For each specialty, the first column reports the proportion (in %) of all specialists in the initial data base (1996–2002). The two next columns report the proportion (in %) of physicians who chose MC in our sample. The number of physicians for each specialty in 1999 and 2000 is reported in the fourth column. The last two columns provide the average incentive ratio (in %) associated with MC – namely the weighted average fraction of FFS prices paid for clinical services when performed under MC (the weights being the relative importance of each service in 1998)– calculated either over only those services for which the prorated fee is positive, hence excluding non-billable Services (*fourth column*) over all services (*last column*), whether they are billable or not. The *MC Incentive Ratio* is the ratio of the prorated fee-for-service, paid under MC, to the original fee-for-service, paid under FFS.

The participation rate to MC strongly varies across specialties (from 2.59% for gastroenterology to 83.87% for neurosurgery in 2000). Two reasons may explain this result. First, the average incentive ratios vary from one specialty to another. For example, in neurosurgery, services are compensated at a relatively high incentive ratio (including non-billable services) of 53%. Similarly, in radiation oncology where the MC participation rate are 70%, the average incentive ratio is 54%. Thus, there is a positive correlation between the level of MC participation rate and the MC incentive ratio. Second, some specialties are known to have a relatively low productivity in terms of quantity of services. This is the case for pediatrics since children need special care. The *per diem* may induce them to choose the MC system even though their incentive ratio is quite small (14.27%). This may partly explain their relatively high MC participation rate of 43.5%.

Table 2 shows summary statistics on practice variables for the treatment and control groups, before and after the reform. We note that before the reform, the control group provided more services on average (132 430 CAN$ *vs* 112 540 CAN$), had a larger income (212 740 CAN$ *vs* 178 840 CAN$) and worked slightly fewer hours per week (54 *vs* 55.28) than the treatment group. These statistics indicate that physicians who chose to remain under FFS were, on average, higher productivity physicians, in terms of the volume of services. Consequently, the control-group physicians would have suffered more

Table 2: Descriptive statistics : Practice variables by group

	Control group				Treatment group			
	Before		After		Before		After	
	Mean	St. Error	Mean	St. Error	Mean	St. Error	Mean	St. Error
Volume	132.43	69.08	136.27	74.82	112.54	52.50	109.86	54.49
Income	212.74	78.83	214.50	84.93	178.84	68.12	200.76	63.64
Hours at work	54.00	14.04	52.29	14.04	55.28	12.65	52.38	12.82
Clinical hours	44.70	15.41	44.43	15.44	42.23	14.54	41.45	13.85
_____ Hospital	31.02	18.79	30.59	19.48	32.53	17.87	32.83	17.37
___ Private clinic	13.68	15.05	13.83	15.29	9.70	14.24	8.62	13.36
Non Clinical hours	6.20	7.92	4.88	7.65	8.30	8.83	7.34	8.50
_ Administration	3.60	6.51	2.06	6.16	4.19	6.66	2.57	5.99
_____ Teaching	2.59	4.04	2.82	4.26	4.11	5.20	4.78	5.72
Research hours	3.10	8.60	2.98	8.06	4.74	9.26	3.58	7.55

Notes: The cells display the average level of the practice variable in *row* and its st. deviation, for the control (*left-hand side columns*) and treatment (*right-hand side columns*) groups, before (*first sub-columns*) and after (*second sub-columns*) reform. The Before Period is 1996-1999 (the annual CMQ surveys are conducted in July while the reform was introduced in September 1999) and the After Period is 2000. There are 1 697 physicians in the control group and 423 in the treatment group. The period of observation is 1996-2000, leading to N=9 238 observations. *Variables*: Volume and income are measured in thousands (2000) CAN$; hours of work variables are measured on a weekly basis.

from the reduction of fees under the MC scheme than the treatment group. This is consistent with the presence of a self-selection problem. Also, before the reform, physicians in the treatment group worked less in private clinics (generally not subject to mixed compensation) and they performed more non-clinical activities such as administrative and teaching tasks (compensated under the MC *per diem*).

Table 3 shows the makeup of the treatment and control groups in terms of the personal characteristics of the physicians in our sample. The male physicians represent 81% of the control group (before the reform) while representing 72% of the treatment group, indicating that the MC participation rate is smaller for males than for females. Similarly francophone physicians represent 79% of the control group and 86% of the treatment group, which indicates that the MC participation rate is higher for francophone than for non-francophone physicians. Also, the average age of physicians in the treatment group is smaller than that of the control group. This indicates that the MC participation rate decreases with age. Moreover, the MC participation rate is more important in Metropolitan regions (CMA with or without teaching university) than in smaller areas.

The statistics on volume of service, income and time-use variables in Table 2 can be directly used to compute simple difference-in-difference (DD) estimates of the impact of the reform. However, no control variables are taken into account in such estimates. Given the presence of a potential self-selection bias, we will focus on econometric estimates.

Table 3: Descriptive statistics : Personal Characteristics by group

	Control group				Treatment group			
	Before		After		Before		After	
	Mean	St. Error	Mean	St. Error	Mean	St. Error	Mean	St. Error
Sex (Male=1)	0.81	0.39	0.80	0.40	0.72	0.45	0.69	0.46
Language (French=1)	0.79	0.41	0.79	0.41	0.86	0.35	0.87	0.34
Age	5.19	2.09	5.63	2.11	4.66	1.96	5.01	1.94
CMAs with teaching university	0.73	0.45	0.72	0.45	0.78	0.41	0.77	0.42
CMAs without teaching university	0.06	0.25	0.07	0.25	0.08	0.27	0.09	0.29
Census Agglomeration	0.13	0.33	0.11	0.31	0.11	0.31	0.11	0.31
Other areas	0.08	0.28	0.10	0.31	0.03	0.18	0.03	0.17
Regional Hospital	0.15	0.36	0.14	0.35	0.11	0.31	0.10	0.31
Supra Regional Hospital	0.24	0.43	0.24	0.43	0.15	0.35	0.15	0.36
Local and regional Hospital	0.50	0.50	0.50	0.50	0.72	0.45	0.71	0.46
Other Hospital	0.10	0.31	0.12	0.33	0.03	0.16	0.10	0.31
Anesthesiology	0.09	0.28	0.09	0.29	0.15	0.36	0.18	0.38
Cardiology	0.14	0.35	0.14	0.35	0.05	0.21	0.04	0.20
Dermatology	0.07	0.25	0.07	0.25	0.07	0.25	0.06	0.24
Endocrinology	0.03	0.18	0.03	0.18	0.05	0.22	0.04	0.20
Gastroenterology	0.07	0.26	0.07	0.25	0.00	0.07	0.01	0.08
General surgery	0.12	0.33	0.12	0.33	0.11	0.32	0.11	0.31
Nephrology	0.04	0.20	0.04	0.20	0.01	0.10	0.01	0.10
Neurology	0.06	0.25	0.06	0.24	0.03	0.18	0.03	0.17
Neurosurgery	0.00	0.06	0.00	0.05	0.07	0.25	0.06	0.24
Obstetrics and gynecology	0.12	0.33	0.12	0.33	0.03	0.18	0.03	0.18
Orthopaedic surgery	0.09	0.29	0.10	0.30	0.05	0.21	0.04	0.20
Pediatrics	0.09	0.28	0.08	0.28	0.25	0.43	0.26	0.44
Pneumology	0.05	0.21	0.05	0.21	0.01	0.09	0.01	0.11
Radiation oncology	0.01	0.08	0.01	0.07	0.05	0.21	0.05	0.22
Rheumatology	0.01	0.10	0.01	0.10	0.07	0.26	0.06	0.24

Notes: Sample distribution over individual characteristics inside control (*left-hand side column*) and treatment (*right-hand side*) groups, before (*first sub-column*) and after (*second*) the reform. CMA: Census Metropolitan Areas. *Age*: There are 10 groups of five years interval of age, ranked from 1 (30 years old and less) to 10 (70 years old and more).

4 Estimation Strategy

As we seek to provide an empirical evaluation of the reform, our analysis is designed so as to measure the effect of the introduction of the MC system on physician behaviour. Our empirical strategy relies on a panel data model approach that can be seen as a generalization of the DD approach to the estimation of treatment effects. The panel dimension of our data is used to control for sample self-selection.[17] We also account for censoring, when required, by extending this framework to a pooled Tobit model.

4.1 A panel data model with sample selection

Let y_{it0} denote the supply of services associated with physician i's practice at time t under the FFS scheme. Similarly, y_{it1} denotes the supply of services by physician i at time t under the MC scheme. We denote, for the present, y as a generic measure of supply. In our empirical work we will consider specifications with the volume of clinical services, hours of work (disaggregated in various dimensions) and earnings as our dependent variable. We specify outcomes as a set of linear regression equations:

$$y_{it0} = \theta_t + \mathbf{x}_{it}\beta + c_i + u_{it0} \quad (1)$$

$$y_{it1} = \theta_t + \mathbf{x}_{it}\beta + \alpha + c_i + u_{it1}, \quad (2)$$

where the parameter θ_t denotes a time-varying intercept and $\mathbf{x}_{it}$ denotes a set of observable characteristics (such as age, specialty, gender, region, type of establishment and language). The disturbances are broken down into two components: c_i denotes an individual unobserved effect that does not vary over time while u_{it0} and u_{it1} represent temporary unobservable variables affecting the outcomes under the respective compensation systems. It is assumed that $E[u_{it0}|\theta, \mathbf{x}_i, c_i] = E[u_{it1}|\theta, \mathbf{x}_i, c_i] = 0$. Note that equations (1) and (2) represent the distribution of outcomes in the population. As such, (2) represents the (hypothetically) observed supply if all specialist physicians were to be paid under the MC scheme. The effect of MC on y is captured by α and is called the average treatment effect (ATE) in the literature (*e.g.*, Heckman and Vytlacil 2001). ATE estimates the average variation in the supply of services if all physicians change from FFS to MC as compared with no change at all. In this model, the assumption of a common time-varying intercept under each scheme (*i.e.,* same θ_t) is crucial to identifying the effect of MC on y. This *parallel trend* assumption will be tested for various supply of services equations in the econometric section.

Now let y_{it} denote the supply of services by physician i at time t under the *observed* payment scheme, *i.e.*, the scheme chosen by physician i at time t. Then, the observed outcomes are linked to the above equations through:

$$y_{it} = D_{it}y_{it1} + (1 - D_{it})y_{it0}, \quad (3)$$

[17] While the decision to adopt MC was not individual based, ethical concerns prohibited us from accessing information on the department in which an individual physician was working which would have allowed modeling the voting decision to adopt MC. We recognize that the individual effects in the empirical work could be somewhat different if the decision was made individual by individual rather than through this voting process. Note that the collective decision process reduces to an individual-based decision rule if physicians are mobile enough to choose their department based on similar preferences for MC.

where D_{it} is the indicator variable of the specialist i's decision to select MC at time t. Thus in this model, participation in the treatment group ($D_{it} = 1$) or to the control group ($D_{it} = 0$) is endogenous.[18]

Substituting (1) and (2) into (3) gives:

$$y_{it} = \theta_t + \mathbf{x}_{it}\beta + D_{it}\alpha + \epsilon_{it}, \tag{4}$$

where

$$\epsilon_{it} = c_i + u_{it0} + D_{it}(u_{it1} - u_{it0}). \tag{5}$$

As is well known (*e.g.,* Angrist and Imbens, 1999) standard regression methods applied to (4) generally fail to provide a consistent estimate of α due to a non-zero covariance between D_{it} and ϵ_{it}. The three components of the error term ϵ_{it} in (5) allow identification of the reasons for this difficulty. First, if physicians differ with respect to unobservable permanent elements (as represented by c_i) that affect their supply of services, and if, for instance, low productivity physicians (in terms of quantity of care) prefer the MC system, then self-selection leaves high productivity physicians under FFS and low productivity physicians under the MC system. A comparison of productivity levels will therefore confound the effects of the compensation system with the differences in unobserved productivity. Consequently, individuals may be more suited to one regime or the other – self-selection thus implies that $E[c_i|\theta, \mathbf{x}_i, D_{it} = 0] \neq 0$ and $E[c_i|\theta, \mathbf{x}_i, D_{it} = 1] \neq 0$.

A second problem occurs when physicians who decide to switch to MC experience a temporary change in the supply of services just prior to the change in the compensation system, something often referred to as *Ashenfelter's Dip* in the program-evaluation literature. An example would be if some physicians temporarily suffer from a health problem that reduced their productivity just before switching to MC. In this case, the *ex ante* comparison between groups no longer identifies permanent differences between individuals, *i.e.* $E[u_{it0} \mid \theta, \mathbf{x}_i, c_i, D_{it} = 1] \neq 0$. If the temporary divergence between groups vanishes in post-reform periods, this difference would then be unduly attributed to the treatment. Given our panel stretches over five years however, we are confident that we can control for any temporary changes in supply. We therefore disregard the risk of an *Ashenfelter's Dip* in the analysis.

A third problem occurs when the effects of the reform are not homogeneous, that is, when $E[u_{it1} - u_{it0}|\theta, \mathbf{x}_i, c_i, D_{it} = 1] \neq 0$. This expression represents the average unobservable change in the supply of services for the doctors who change from FFS to MC. Conditional on c_i and the explanatory variables, the OLS coefficient associated with D_{it} will therefore estimate $\tilde{\alpha} = \alpha + E[u_{it1} - u_{it0}|\theta, \mathbf{x}_i, c_i, D_{it} = 1]$. The parameter $\tilde{\alpha}$ is called the average treatment effect on the treated (ATT) in the literature (*e.g.* Heckman and Vytlacil, 2001). ATT is generally different from ATE (given here by α), the common impact of the reform, since it estimates the average impact of the reform only for those who actually choose to change to MC. Only when the population of physicians is homogeneous, ATT will be equal to ATE. In this

[18] In practice some physicians may work under both systems at the same time. Note also that, by moving from one health establishment to another, a physician could switch back and forth from MC to FFS. A department could also switch from one system to another by a voting decision. For simplicity, we attribute workers to the treatment group if part of their income was earned under MC in every quarter of the year (see the discussion in Section 3.3).

paper, we focus on the ATT's since the ATE's are not identified without imposing additional structure to the model.

We are thus left with a self-selection problem.[19] A number of approaches have been developed for controlling for this endogeneity, each of which requires different assumptions in order to be valid. In the health literature, the approach which has been most widely used is the instrumental variable method (*e.g.* Basu et al., 2007; Devlin and Sarma, 2008). However, given the longitudinal nature of our data, it is natural to use an alternative approach based on panel data regressions. Thus, we observe each physician in the selected sample before and after the introduction of the MC system. We can therefore exploit information on the supply of services prior to the introduction of the MC system to eliminate the individual effect c_i which is the source of the self-selection problem. More precisely, using (4) and (5) and the definition of $\tilde{\alpha}$, the model to be estimated becomes:

$$y_{it} = \theta_t + \mathbf{x}_{it}\beta + D_{it}\tilde{\alpha} + c_i + \eta_{it}, \tag{6}$$

where $\eta_{it} = u_{it0} + D_{it}(u_{it1} - u_{it0}) - D_{it}E[u_{it1} - u_{it0}|\theta, \mathbf{x}_i, c_i, D_{it} = 1]$, with $E(\eta_{it}|\theta, \mathbf{x}_i, c_i, D_{it}) = 0$. Our model assumes strict exogeneity of $\{\theta_t, \mathbf{x}_{it}, D_{it}\}$ for all t, conditional on the individual effect. Note that the model *does not* impose the absence of correlation between the individual effect and the explanatory variables. Eq. (6) will be estimated using a fixed effects (or within) transformation to eliminate the individual-specific effects. Also, the absence of self-selection will be tested using a Hausman test comparing the coefficients of the fixed effects (FE) model with those of the corresponding random effects model. The latter model is more restrictive since it *does* impose zero correlation between the observed explanatory variables and the individual effect.

When applying the FE model, the presence of heteroskedasticity of the error terms and serial correlation in η_{it} may give an improper covariance matrix estimator (see Bertrand et al., 2004). We provide robust standard error estimates that are valid in the presence of arbitrary heteroskedasticity and serial correlation in η_{it} when the length of the panel (here, equal to 5) is small (see Arellano, 1987).[20]

4.2 Extension to Censored Data

Some of the practice variables of interest in our study are left-censored in the sample (this is the case, *e.g.*, for teaching hours in 46% of the sample and for research hours in 69% of the sample). For those variables, the FE effects model estimated by OLS will be inconsistent. Moreover, a standard FE Tobit will also be inconsistent due to the so-called *incidental parameters problem*; *i.e.*, for a fixed length of the panel, the number of parameters to be estimated increases with the number of observations.

To accommodate this, we estimate the model using a *pooled Tobit*. Also, we add dummy variables to account for some observable time-invariant individual variables (*e.g.*, Treated=1 for physicians who

[19] Our model ignores the possibility of self-selection of patients into different provider types. In that case, we would expect the case-mix of physicians to change with the reform and our estimates could be biased. However, this problem is not likely to be important since the information concerning whether a physician is FFS or MC is confidential and cannot be released to the public by the RAMQ.

[20] To implement the Hausman test, we use a robust form of the statistic which accounts for this problem (see Wooldridge, 2001, p.291).

chose MC, sex, language, specialty – 14 dummies). Note that, being time-invariant, these variables are not identified under the FE model. This yields a new vector of explanatory variables, $\tilde{x}_{it}$. It is assumed that conditional on $\tilde{x}_{it}$, the error term of the corresponding latent model is distributed $\mathcal{N}(0, \sigma^2)$. A useful feature of the pooled Tobit is that the estimated parameters will be consistent and asymptotically normal even if the error terms are arbitrarily serially correlated. We provide estimates of the standard errors of coefficients robust to serial correlation of unrestricted form.[21] Unlike OLS estimates, the estimated impact of the reform differs from the estimated parameters in a Tobit model. This is computed as the average over the doctors who changed from FFS to MC of the estimate of $E(y_{it}|\theta, \tilde{\mathbf{x}}, D_{it} = 1) - E(y_{it}|\theta, \tilde{\mathbf{x}}, D_{it} = 0)$ at $t = 2000$ (period post-reform).

5 Results

The upper part of Table 4 provides results for all specialties in the sample, with the control group given by specialists who remained in the FFS scheme in 2000. For all models considered, tests of parallel trend do not reject the null hypothesis that, prior to the reform, time effects (θ_t) were the same in both control and treatment groups. This test is crucial for the identification of the impacts of the reform. Moreover, the Hausman test rejects the random effects model for all OLS models.

On average physicians who changed to MC reduced their volume of billable clinical services by 8 300 CAN$,[22] which represents a decrease of 6.15%. This decrease in the volume of services is accompanied by a reduction of 1.14 in weekly clinical hours. The reduction in the number of clinical hours is mainly due to the reduction of 0.81 in hours worked per week in private clinics.[23] This is the case since the reform had no significant impact on clinical time spent in hospitals.

There is also evidence that the reform addressed some of the multi-tasking problems associated with the FFS contract. First, note that, in relative terms, the reduction in the volume of services is larger than the reduction in clinical hours worked thus resulting in a 3.58% increase in the time spent per service. This result is therefore consistent with MC physicians substituting quality for quantity in their supplies of clinical services. Second, switching to the MC system caused physicians to spend more time (0.53 more hours per week) on non-clinical activities; *i.e.,* administrative tasks and teaching. While these activities are important in insuring the quality of health care (both in the present and future), they are not remunerated under the FFS system and are likely to be neglected. The measured increase, in the order of 7.92 %, suggests that the reform had positive effects in this direction as well. Notice, however, that we are less successful in decomposing the total increase in non-clinical activities into constituent parts, the effect on administrative activities being not significantly different from zero.

[21] An alternative to the pooled Tobit is the random effects Tobit. However, this model imposes strong restrictions on the variance matrix. Since this model imposes that the error terms are equicorrelated, the variance matrix will have common diagonal entries and common off-diagonal entries. In any case, econometric investigations showed that estimated coefficients from these two models do not differ very much.

[22] All results in dollars are in constant (2000) CAN$. The results for weeks of work are not reported because they are insignificant in all specifications.

[23] The estimated coefficients on hospital hours and hours worked in private clinics do not add up to the coefficient on total clinical hours because the latter was estimated by OLS and the former were estimated by pooled Tobits.

Table 4: Impact of mixed compensation on practice variables, all specialties

	Fixed effects OLS				Pooled Tobit					
	Volume	Income	Hours	Clinical hours/ week			Non-Clinical hours/ week			Research
	/ year	/ year	/ week	Tot.	Hosp.	Priv. Cl.	Tot.	Admin.	Teach.	hours/ week
(Control group: specialists paid only under the FFS scheme after reform)										
Treatment effect	-8.30***	17.29***	-1.09*	-1.14**	0.11	-1.52**	1.23***	1.01	0.60*	-2.43***
St. Error	1.55	2.48	0.58	0.57	0.60	0.74	0.45	0.67	0.34	0.68
Effect of the reform	-8.30***	17.29***	-1.09*	-1.14**	0.15	-0.81**	0.53*	-0.16	0.47**	-0.77***
St. Error	1.55	2.48	0.58	0.57	0.57	0.37	0.32	0.28	0.22	0.25
% effect of the reform	-6.15	8.05	-2.05	-2.57	0.45	-9.07	7.92	-7.89	10.74	-14.72
Log likelihood	-	-	-	-	-35 685	-25 151	-26 066	-19 402	-18 040	-14 353
Hausman test[a]	94.30	93.81	68.13	58.10	-	-	-	-	-	-
Test of parallel trend[b]	1.00	1.19	0.78	0.36	2.35	1.34	4.71	1.53	4.44	4.24
(Control group: general practitioners paid only under the FFS scheme)										
Treatment effect	0.27	-4.14***								
St. Error	1.65	2.32	0.56	0.54	0.63	0.66	0.52	0.80	0.37	0.83
Effect of the reform	-5.70***	23.38***	-1.96***	-1.64***	1.18**	-1.47***	1.12***	-0.41	0.50**	-1.09***
St. Error	1.65	2.32	0.56	0.54	0.60	0.40	0.33	0.30	0.22	0.28
% effect of the reform	-3.90	14.80	-4.24	-3.88	3.68	-16.58	18.06	-16.26	10.68	-18.59
Log likelihood	-	-	-	-	-33 207	-48 201	-29 124	-25 202	-10 689	-7 157
Hausman test[a]	165.58	92.61	46.68	140.47	-	-	-	-	-	-
Test of parallel trend[b]	3.77	0.73	12.86	9.41	24.04	2.08	70.15	0.78	8.70	9.80

Significance levels: * 10%, ** 5%, *** 1%.

Notes. Regression results based on varying control groups for the whole sample of physicians. In the fixed effects OLS models, we add dummies for years (4), regions (3), establishments (3 : "supraregional", "local and regional", and others) and age (8). Also, a control variable = 1 if the observation is also treated in 1999 has been introduced in the volume and income regressions. In the pooled Tobit models, we also add dummies for group (Treated = 1), sex (Male = 1), language (Francophone = 1) and specialties (14). These variables are unidentified in the fixed effects OLS models since they are time-invariant. The *Treatment effect* variable is set equal to 1 in 2000 if the observation is in the treatment group, to 0 otherwise.In the *Upper part*, FFS specialist physicians are used as control group for MC physicians. There are 1 697 physicians in the control group and 423 in the treatment group providing N=9 238 observations. In the *bottom part*, the treatment group remains the same but the control group is FFS general practitioners (2 662 physicians, N=13 124 observations).

[a]The statistic is $\chi^2(20)$ in the cases of volume and income and $\chi^2(19)$ in the cases of hours and clinical hours per week.

[b]The statistic allows to test the assumption that, before the reform, time changes have the same effect on both control and treatment groups. Years 1996 to 1998 (resp., 1996 to 1999) are used in the estimations of volume and income (resp., all other variables). In the fixed effects OLS models, the statistic is $F(2, 2\ 119)$ (specialists) or $F(2, 3\ 084)$ (GP) for volume and income and $F(3, 2\ 119)$ (specialists) or $F(3, 3\ 084)$ (GP) for all other variables. It is $\chi^2(3)$ (Wald test) in the pooled Tobit models.

We also note that the reform was not cheap – according to our results, physicians who chose the MC system benefited from an increase in annual income of 17 290 CAN$; *i.e.,* an increase of 8.05%. What is more, the impact of the reform caused a reduction of 1.09 total hours of work per week performed by MC physicians. As such, the reduction in the fees and the introduction of the *per diem* induced MC physicians to consume more leisure. Other studies (*e.g.*, Croxson et al., 2001), have also shown that changes in the payments received by doctors affect their number of working hours. Finally, we note that time spent on research activities, which are not remunerated neither by the FFS nor by the MC, decreased by 14.7%.

6 Robustness Checks

To validate our results, we performed a number of robustness tests. Our first step in this direction was to analyse a placebo reform, suppressing data from 1999 and 2000 and using 1998 as the post-reform period. The coefficients of the placebo variable (= 1 if treated in 1998) should not be significant since no true reform has been introduced during this year. If they were significant, then it would raise the possibility that the results of Table 4 are also significant by mere chance. The results[24] indicated that none of the placebo coefficients were significant, which is encouraging for the validity of our estimates of the impacts of the reform.

6.1 Changing the Control Group

Following Song and Manchester (2007), we consider how robust our results are to using a different control group: general practitioners who were paid exclusively on FFS contracts. As long as this control group is affected by the same economic shocks through time as are specialists who chose to be paid on a FFS basis (the basic control group), one should expect the estimated effect of the reform to be similar to the original one. Alternatively, the general-practitioner control group may be less susceptible to confounding forces such as general-equilibrium effects or the removal of income ceilings (see below). The bottom part of Table 4 presents the results. Generalists do not perform the same types of services as MC specialists. Therefore, for this control group, instead of using billable services on MC to construct a measure of volume of services, we used all services billed by general practitioners.

Our tests of parallel trend are rejected in this case for most time-use variables (except time spent on private clinic and on administrative duties). Thus, strictly speaking, our identification criterion fails for this sample. This may reflect the fact that general practitioners do not represent a natural control group for MC specialists. In spite of this, it is reassuring to see that the results obtained when using generalists as the control group are *qualitatively* similar to those obtained when using FFS specialists. Quantitatively there are some important differences though. Thus, using the alternative control group, our results show that the reform reduced the volume of clinical services performed by MC specialists by 5 700 CAN$ (or 3.9%) as compared with 8 300 CAN$ (or 6.15%) when using FSS specialists as the

[24]The corresponding results are available in a longer version of the paper (see Dumont et al., 2007) or directly from the authors on request.

control group. Also, the impact of the reform on total clinical hours per week of MC specialists is now 1.64, which is larger than the estimate obtained when FSS specialists are used as the control group (=1.14).

However, contrary to our results with FSS specialists as the control group, MC physicians do not seem to spend more time per service. Indeed, in percentage, the reductions in the volume (3.90%) and in clinical hours (3.88%) are quite the same. Again, the reduction in clinical hours is mainly due to the decrease of 1.47 hour per week in private clinics. However, contrary to the previous estimate, the number of hours in hospitals increases now by 1.18 hour per week, indicating the possibility of a substitution between private clinic and hospital hours. Note however that the statistic on the test of parallel trend is very large for hospital hours which might indicate a serious problem when computing the marginal effects of the reform. Non-clinical hours per week increase by 1.12 hours which is far more important than the growth in hours spent teaching (0.5 hours). Again, the large value of the parallel trend test statistic suggests a problem with these results. Note however that the latter estimate is quite similar to the one obtained when using MC specialists as the control group (=0.47). Finally, there is a difference of 6 090 CAN$ in the two estimates of the impact of the reform on the income of MC specialists. This gap might be attributable to an increase in the income of specialists in 2000 which the general practitioners did not enjoy.

In summary, tests of parallel trends suggest that FFS specialists are a better control group for MC specialists than are general practitioners. However, results using general practitioners as a control group qualitatively validate our previous results in spite of some quantitative differences: estimates of the effect of the reform increase for all variables except the volume of services. Moreover, the increase of hospital hours is now significant.

6.2 Heterogeneous Response

The next results in Table 5 provide an analysis of the reform by gender. They show that females (upper part of the Table) benefit more from MC than males (bottom part of the Table). First, females who changed from FFS to MC increased their income by 20 290 CAN$ in comparison with 15 480 CAN$ for males. Second, the reduction in billable clinical services is more important for females than for males. MC females reduce their volume of services by 8.22% as compared with 5.54% in the case of men (10 080 CAN$ *vs* 7 620 CAN$). Women seem to be financially penalized by the FFS system. They bill fewer services than men not only because they work fewer hours than men (51.03 against 52.6 for males) but also because, on average, they spend more time with their patients (2.76 against 3.14 for men). The higher increase in income for females thus partly reveals that the MC system has a stronger effect on their income because their "productivity" (in terms of quantity of services) is lower than that of males. The *per diem* benefits more to physicians with lower productivity and therefore induces a larger increase in females income. Third, the positive impact of the reform on the time spent per billable service is higher for women (=6.39%) than for men (=2.84%). This suggests that the quality-quantity substitution effect of the reform has been more important for women than for men.

Part of the results for women can also be explained by the behaviour of MC pediatricians since

Table 5: Impact of mixed compensation on practice variables, by gender

	Fixed effects OLS						Pooled Tobit			
	Volume	Income	Hours	Clinical hours/ week			Non-Clinical hours/ week			Research
	/ year	/ year	/ week	Tot.	Hosp.	Priv. Cl.	Tot.	Admin.	Teach.	hours/ week
Males (Control group: male specialists paid only under the FFS scheme after reform)										
Treatment effect	-7.62***	15.48***	-1.19	-1.19*	-0.41	-0.91	1.22**	1.13	0.54	-2.16**
St. Error	1.84	2.93	0.74	0.71	0.70	0.89	0.55	0.80	0.40	0.76
Effect of the reform	- 7.62***	15.48***	-1.19	-1.19*	-0.35	-0.44	0.54	-0.06	0.43	-0.74**
St. Error	1.84	2.93	0.74	0.71	0.67	0.41	0.40	0.35	0.27	0.30
% effect of the reform	-5.54	7.04	-2.22	-2.70	-1.03	-5.66	7.53	-2.65	9.42	-12.40
Log likelihood	-	-	-	-	-28 205	-19 867	-20 909	-15 851	-14 329	-11 934
Hausman test[a]	108.28	113.94	50.43	55.67	-	-	-	-	-	-
Test of parallel trend[b]	0.55	0.88	0.84	0.08	0.62	0.96	2.26	1.14	1.74	1.52
Females (Control group: female specialists paid only under the FFS scheme after reform)										
Treatment effect	-10.08***	20.29***	-0.79	-0.83	1.30	-3.37**	1.31	1.22	0.71	-2.62*
St. Error	27.96	4.62	0.96	0.98	1.18	1.35	0.83	1.24	0.67	1.50
Effect of the reform	-10.08***	20.29***	-0.79	-0.83	1.26	2.12 ***	0.65	-0.04	0.55	-0.75*
St. Error	27.96	4.62	0.96	0.98	1.10	0.82	0.55	0.50	0.40	0.44
% effect of the reform	-8.22	10.68	-1.54	-1.83	4.16	-17.71	11.94	-3.23	13.97	-21.40
Log likelihood	-	-	-	-	-7 410	-5 223	-5 094	-3 497	-3 658	-2 350
Hausman test[a]	23.54	29.55	35.71	22.32	-	-	-	-	-	-
Test of parallel trend[b]	0.55	0.88	0.89	1.36	4.12	4.75	6.53	3.61	8.36	14.16

Significance levels: * 10%, ** 5%, *** 1%.
Notes: Regression results by gender sub-samples. In the fixed effects OLS models, we add dummies for years (4), regions (3), establishments (3 : "supraregional", "local and regional", and others) and age (8). Also, a control variable = 1 if the observation is also treated in 1999 has been introduced in the volume and income regressions. In the pooled Tobit models, we also add dummies for group (Treated = 1), language (Francophone = 1) and specialties (14). These variables are unidentified in the fixed effects OLS models since they are time-invariant. The *Treatment effect* variable is set equal to 1 in 2000 if the observation is in the treatment group, to 0 otherwise. The *upper part* reports results for males physicians. There are 1 363 physicians in the control group and 292 in the treatment group, the number of observations is $N = 7\ 280$. The *bottom part* reports results for females physicians. There are 334 physicians in the control group and 131 in the treatment group, the number of observations is $N = 1\ 958$.
[a]The statistic is $\chi^2(20)$ in the cases of volume and income and $\chi^2(19)$ in the cases of hours and clinical hours per week.
[b]The statistic allows to test the assumption that, before the reform, time changes have the same effect on both control and treatment groups. Years 1996 to 1998 (resp., 1996 to 1999) are used in the estimations of volume and income (resp., all other variables). In the fixed effects OLS models, the statistic is $F(2, 1\ 654)$ (males) or $F(2, 464)$ (females) for volume and income and $F(3, 1\ 654)$ (males) or $F(3, 464)$ (females) for all other variables. It is $\chi^2(3)$ (Wald test) in the pooled Tobit models.

approximately one-half of the physicians in the female treatment group are pediatricians. The results for this specialty are presented in the upper part of Table 6. This decrease of 12 990 CAN$ (=12.8%) in their volume of services is larger than that for all specialties (=8 300 CAN$). This decrease is mostly explained by the reduction in time spent in private clinics (2.92 hours per week), which is almost equal to the reduction of 2.88 in total clinical hours worked. Prior to the reform, 53% of the volume of billable pediatric services were performed in private clinics. By reducing time spent in these clinics, the reform has also reduced the number of services provided in them. We also observe that the MC pediatricians did not increase the number of services provided in hospitals. Therefore, the fall in volume may only represent a reduction in the number of services billed in private clinics. The MC pediatricians also increased by 1.56 hours per week their time spent on non-clinical activities (administrative tasks and teaching) and reduced by 2.01 hours a week their time spent on research. All in all, they reduced their total hours of work by 1.87 hours per week.

The bottom part of Table 6 shows results for general surgery, another specialty which reacted strongly to the reform. MC general surgeons decreased their volume of services by 15 320 CAN$ (= 10.7%) while increasing their level of income by 18 370 CAN$ (= 8.14%). However, we cannot say much more about the effect of the reform on the use of their time since none of the coefficients measuring the impact of the reform is significant at the 5% level.

After removing these two specialties from the sample, the reduction in the volume of services performed by MC specialists – other than pediatricians and surgeons – is less important, but still evident:[25] it is 4 610 CAN$ instead of 8 300 CAN$ (or 3.32% as compared with 6.15%). On the other hand, the impact of MC on income remains very close to the previous estimates, corresponding to an increase of 17 950 CAN$ as compared with 17 290 CAN$ for all specialties. Moreover, it seems that the reduction in total clinical hours and in private clinic hours is mainly due to the behaviour of MC pediatricians since none of the coefficients associated with these two variables are significant at the 5% level for the other specialties. A similar result is obtained for research hours. In short, the reform had a more important impact on the allocation of time for pediatrics and general surgery than for other specialties. This result could partly be explained by the specific characteristics of these specialties.

6.3 General-Equilibrium Effects

As an additional check, we considered the possible presence of general-equilibrium effects (*e.g.*, Heckman et al., 1998). These would occur if, for example, the reduction in the volume of services on the part of the treatment group led to an offsetting increase in the volume of services among control-group physicians, perhaps in reaction to increased waiting periods on the part of patients. The fact that the average volume of services is higher among FFS physicians after the introduction of the reform is consistent with this explanation which would place doubts on the validity of our control group and our measures of the effect of the reform.

Following Lewis (1963), we exploit differences in the participation rate across specialties to investigate general-equilibrium effects. If these effects are important then specialties in which the participation rate

[25] The full results table is provided in Dumont et al. (2007).

Table 6: Impact of mixed compensation on practice variables, pediatrics and general surgery

	Fixed effects OLS				Pooled Tobit					
	Volume	Income	Hours	Clinical hours/ week			Non-Clinical hours/ week			Research
	/ year	/ year	/ week	Tot.	Hosp.	Priv. Cl.	Tot.	Admin.	Teach.	hours/ week
Pediatrics (Control group: pediatricians paid only under the FFS scheme after reform)										
Treatment effect	-12.99***	15.83***	-1.87*	-2.88**	0.34	-3.46*	3.92***	3.27*	0.48	-9.56**
St. Error	3.72	6.62	1.12	1.14	1.33	1.83	1.40	1.85	1.19	4.03
Effect of the reform	-12.99***	15.83***	-1.87*	-2.88**	0.49	-2.92**	1.56*	0.48	0.61	-2.01*
St. Error	3.72	6.62	1.12	1.14	1.20	1.48	0.80	0.68	0.56	1.06
% effect of the reform	-12.81	8.69	-3.75	-6.55	2.20	-15.88	27.89	25.14	13.88	-38.20
Log likelihood	-	-	-	-	-3 423	-3 757	-2 678	-2 078	-1 653	-1 147
Hausman test[a]	53.86	37.68	41.10	44.97	-	-	-	-	-	-
Test of parallel trend[b]	0.92	0.48	0.28	1.22	2.27	7.60	9.03	2.55	7.99	9.05
General Surgery (Control group: general surgeons paid only under the FFS scheme after reform)										
Treatment effect	-15.32***	18.37***	-2.70	-1.83	0.04	-4.68**	0.62	0.27	0.53	-1.36
St. Error	3.81	4.28	2.48	2.37	2.54	2.00	1.27	1.64	1.04	1.59
Effect of the reform	-15.32***	18.38***	-2.70	-1.83	0.06	-1.02	0.07	-0.83	0.38	-1.07
St. Error	3.81	4.38	2.48	2.37	2.52	0.83	0.95	0.75	0.66	0.77
% effect of the reform	-10.72	8.14	-4.81	-3.77	0.14	-31.47	0.81	-23.05	8.37	-21.53
Log likelihood	-	-	-	-	-4 509	-3 014	-3 093	-2 459	-1 916	-1 235
Hausman test[a]	39.41	8.07	169.10	26.50	-	-	-	-	-	-
Test of parallel trend[b]	1.46	0.98	1.13	0.71	4.81	0.96	1.19	2.08	2.49	2.08

Significance levels: * 10%, ** 5%, *** 1%.
Notes: Results from separate regressions for pediatricians and surgeons. In the fixed effects OLS models, we add dummies for years (4), regions (3), establishments (3 : "supraregional", "local and regional", and others) and age (8). Also, a control variable = 1 if the observation is also treated in 1999 has been introduced in the volume and income regressions. In the pooled Tobit models, we also add dummies for group (Treated = 1), sex (Male = 1) and language (Francophone = 1). These variables are unidentified in the fixed effects OLS models since they are time-invariant. The *Treatment effect* variable is set equal to 1 in 2000 if the observation is in the treatment group, to 0 otherwise. The *upper part* reports results for pediatricians. There are 143 physicians in the control group and 110 in the treatment group, the number of observations is $N = 1\ 072$. The *bottom part* reports results for surgeons. There are 212 physicians in the control group and 47 in the treatment group, the number of observations is $N = 1\ 109$.
[a] The statistic is $\chi^2(20)$ in the cases of volume and income and $\chi^2(19)$ in the cases of hours and clinical hours per week.
[b] The statistic allows to test the assumption that, before the reform, time changes have the same effect on both control and treatment groups. Years 1996 to 1998 (resp., 1996 to 1999) are used in the estimations of volume and income (resp., all other variables). In the fixed effects OLS models, the statistic is $F(2, 253)$ (pediatricians) or $F(2, 258)$ (surgeons) for volume and income and $F(3, 253)$ (pediatricians) or $F(3, 258)$ (surgeons) for all other variables. It is $\chi^2(3)$ (Wald test) in the pooled Tobit models.

is highest should exhibit the largest reaction among FFS physicians (the control group). What is more, the effect of the participation rate should be of opposite sign to the measured effect of the reform. We therefore regressed the different practice variables on the participation rate, restricting the sample to FFS physicians. In general, there is little evidence to support the presence of general-equilibrium effects.[26] The participation rate has no significant effect on the volume of services provided among FFS physicians. While it has a significant effect on hours spent seeing patients and on hours spent on administrative duties, in both cases the sign is inconsistent with general-equilibrium effects. In both cases the coefficient on participation rate has the same sign as the estimated effect of the reform. Given these results we conclude that general-equilibrium effects are not a major confounding element in our data.

6.4 Income Ceilings

Finally, we considered the effect of the removal of the gross income ceiling on physician behaviour. Recall, the government removed this income ceiling in January 1999. If this affected the treatment and control groups differently, then it would be a possible confounding effect in our estimates of the compensation system reform. However, there is little evidence to suggest that these ceilings (or their removal) had an important effect on observed physician behaviour. First, only a small number of specialists were actually constrained by the ceiling prior to its removal. In the last semester of 1998 only 5.5% of specialists had an income greater than or equal to the income ceiling. Second, there was no substantive change in our results when we eliminated physicians whose income were close to the ceiling (defined as having an income greater than or equal to the ceiling minus 10 000 CAN$) from our sample.[27] Given this, we are confident that any effect of the removal of the income ceilings on observed behaviour was minimal.

7 Conclusion

In this paper, we have investigated the labour-supply response of physicians to changes in their contracts. We have combined time-use surveys with administrative payroll information to construct a panel of labour-supply data on Quebec physicians. The comprehensive nature of these data allow us to analyse labour-supply response along many dimensions and address multi-tasking issues in response to a particular reform: the introduction of the mixed compensation system. Our results show that, on average, the reform induced physicians who changed from FFS to MC to reduce their volume of (billable) services by 6.15%. However, they increased their time spent per service by 3.58% and their time spent on administrative and teaching tasks (activities not remunerated under FFS) by 7.92%.

These results underline the importance of accounting for multitasking when evaluating compensation systems; they also point to directions for future research. The increase in time spent per service is suggestive of a quality-quantity substitution in the treatment of patients. This, coupled with the increase in time spent on teaching and administrative duties, suggests there were positive effects of the reform,

[26] The full results are available upon request.

[27] The full results are available upon request.

at least partially meeting government objectives and offsetting the reduction in services provided.[28] Of course, time spent per service is an imperfect measure of quality – physicians may simply be taking longer breaks between services, or spending more time with patients without affecting health outcomes. Further study and more extensive data will be needed to fully settle these issues. Comparative information on the health of patients treated by MC and FFS physicians would be useful in this regard. What is more, a complete welfare analysis would weight the increased time spent per patient (or improvement in health status of the patient) against extra time spent waiting to see a physician. Matching data on physician services to waiting times would be an important step in this direction.

Finally we note that our analysis is limited to reporting the effects of a particular reform – the introduction of the MC system. Yet, policy makers may want to predict the effects of other reforms, not observed within the current data. For example, the current reform was voluntary, allowing us to identify the effect of treatment on the treated (those that chose the MC system). It identifies the average treatment effect (the effect that would be observed if the government made MC mandatory for all physicians) only if those physicians who chose MC react to incentives in the same way as those who chose FFS. If there is heterogeneity in response, then measuring the effect of a mandatory reform requires predicting how physicians, who are only observed under FFS, would react under MC. This is not identified by the difference-in-difference estimator. Alternatively, the government may wish to predict the effects of alternative contracts. The fact that the income of physicians who switched to MC increased by 8% suggests that the reform might have been accomplished at a lower *per diem*. Yet, changes in the contract would, undoubtedly, affect participation, services provided and the effect of the reform. One approach to answering such questions would be to use our data to estimate structural parameters governing (optimal) labour-supply decisions. Within a structural model the behaviour under different contracts is a function of the same (utility function) parameters; identification of those parameters allows the prediction of behaviour under alternative policies and the evaluation the effects of (as yet) unobserved reforms. We leave this task for future research.

References

Angrist, J. D., Imbens, G. W. (1999). Comment on james j. heckman, "instrumental variables: A study of implicit behavioral assumptions used in making program evaluations". *Journal of Human Resources* 34 (4), 823–827.

Arellano, M. (1987). Computing robust standard errors for within-groups estimators. *Oxford Bulletin of Economics and Statistics* 49 (4), 431–434.

Baker, G. P. (1992). Incentive contracts and performance measurement. *Journal of Political Economy* 100 (3), 598–614.

[28] Another possible benefit of the reform is the reduction in induced demand. Yet, waiting lists to see specialists are very long in Quebec. Under these circumstances, it is difficult to believe that physicians feel it necessary to induce demand to meet income targets.

Barro, J., Beaulieu, N. (2003). Selection and improvement: Physician responses to financial incentives. *NBER WP* (10017).

Basu, A., Heckman, J. J., Navarro-Lozano, S., Urzua, S. (2007). Use of instrumental variables in the presence of heterogeneity and self-selection: An application to treatments of breast cancer patients. *Health Economics* 16 (11), 1133–1157.

Bertrand, M., Duflo, E., Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics* 119 (1), 249–275.

Blomqvist, A. (1991). The doctor as double agent: Information asymmetry, health insurance, and medical care. *Journal of Health Economics* 10 (4), 411–432.

Blundell, R., MaCurdy, T. (1999). Labour supply: areview of alternative approaches. In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics, Vol. 3A. Elsevier, Cambridge (MA), pp. 1559–1695.

Bolduc, D., Fortin, B., Fournier, M.-A. (1996). The effect of incentive policies on the practice location of doctors: A multinomial probit analysis. *Journal of Labor Economics* 14 (4), 703–732.

Chiappori, P.-A., Salanié, B. (2003). Testing contract theory: A survey of some recent work. In: Dewatripont, M., Hansen, L., Turnovsky, S. (Eds.), Advances in Economics and Econometrics, Eight World Congress. Cambridge University Press, Cambridge (MA), pp. 115–149.

Croxson, B., Caroll, P., Perkins, A. (2001). Do doctors respond to financial incentives ? uk family doctors and the fundholder scheme. *Journal of Public Economics* 79 (2), 375–398.

Devlin, R.-A., Sarma, S. (2008). Do physician remuneration schemes matter? the case of canadian family physicians. *Journal of Health Economics* Forthcoming.

Dumont, E., Fortin, B., Jacquemet, N., Shearer, B. (2007). Physicians' multitasking and incentives: Empirical evidence from a natural experiment. *IZA Discussion Paper* (3229).

Eggleston, K. (2005). Multitasking and mixed systems for provider payment. *Journal of Health Economics* 24 (1), 211–223.

Fortin, B., Jacquemet, N., Shearer, B. (2008). Policy analysis in the health-services market: accounting for quality and quantity. *Annales d'Economie et de Statistiques/The Annals of Economics and Statistics* Forthcoming.

Gaynor, M., Gertler, P. (1995). Moral hazard and risk spreading in partnerships. *Rand Journal of Economics* 26 (4), 591–613.

Gruber, J., Owings, M. (1996). Physician financial incentives and cesarean section delivery. *Rand Journal of Economics* 27 (1), 99–123.

Hart, O. D., Holmstrom, B. (1987). The theory of contracts. In: Bewley, T. (Ed.), Advances in economic Theory. Cambridge University Press, Cambrdige (MA).

Heckman, J., Vytlacil, E. (2001). Policy relevant treatment effects. *American Economic Review, Papers and Proceedings* 91 (2), 107–111.

Heckman, J. J., Lance, L., Christopher, T. (1998). General-equilibrium treatment effects: A study of tuition policy. *American Economic Review* 88 (2), 381–386.

Hillman, A. L., Pauly, M. V., Kerstein, J. J. (1989). How do financial incentives affect physicians' clinical decisions and the financial performance of health maintenance organizations? *New England Journal of Medicine* 321 (2), 86–92.

Holmstrom, B., Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization* 7 (3), 24–52.

Hutchinson, J. M., Foley, R. N. (1999). Method of physician remuneration and rates of antiobitic prescription. *Canadian Medical Association Journal* 160 (7), 1013–1017.

Lewis, H. G. (1963). *Unionism and Relative Wages in the United States.* University of Chicago Press, Chicago (Illinois).

Léger, P.-T. (2007). Physician payment mechanisms. In: Lu, M., Jonsson, E. (Eds.), Financing Health Care: New Ideas for a Changing Society. Wiley-VCH Press, Ch. 6.

Ma, C.-T. A. (1994). Health care payments systems: cost and quality incentives. *Journal of Economics & Management Strategy* 3 (1), 93–112.

Ma, C.-T. A., McGuire, T. G. (1997). Optimal health insurance and provider payment. *American Economic Review* 87 (4), 685–704.

Milgrom, P., Roberts, J. (1992). *Economics, Organization and Management.* Englewood Cliffs: Prentice Hall.

Newhouse, J. P. (1996). Reimbursing health plans and health providers: Efficiency in production versus selection. *Journal of Economic Literature* 34 (3), 1236–1263.

Paarsch, H., Shearer, B. (2000). Piece rates, fixed wages and incentive effects: Statistical evidence from payroll records. *International Economic Review* 41 (1), 59–92.

Pauly, M. V. (1990). The rational nonpurchase of long-term-care insurance. *Journal of Political Economy* 98 (1), 153–168.

Song, J. G., Manchester, J. (2007). New evidence on earnings and benefit claims following changes in the retirement earning test in 2000. *Journal of Public Economics* 91 (3-4), 669–700.

Wooldridge, J. M. (2001). *Econometric Analysis of Cross Section and Panel Data.* MIT Press.