The effect of a performance incentive scheme on the allocation of time across multiple tasks: evidence from physicians.

Peter Sivey, Anthony Scott, Stefanie Schurer

Melbourne Institute of Applied Economic and Social Research
The University of Melbourne

Abstract
Despite substantial evidence of the unintended consequences of performance incentives in the presence of multitasking and information asymmetries, their use for physicians is growing. The aim of this paper is to examine the effects of a pay for performance scheme for primary care physicians in Australia on their allocation of time between tasks in remunerated and unremunerated health problems. We examine whether physicians substitute tasks, increase costs, or become more efficient at undertaking tasks as a result of the incentive scheme. A unique dataset of almost 60,000 physician-patient consultations from almost 4,000 physicians is used. A cost function approach is used to estimate the marginal time-costs and marginal rates of substitution between tasks in consultations, whilst controlling for physician fixed effects and patient characteristics. Preliminary results show that the incentive scheme is associated with lower marginal costs per task for both remunerated and unremunerated health problems, suggesting that GPs produce consultations more efficiently under the incentive scheme. This may be due to both an increase in IT the accompanied the incentive scheme, and also reductions in informational asymmetries between doctors and patients through more frequent patient visits.

Contact: a.scott@unimelb.edu.au

JEL codes: J44, J33, M52, I12

Acknowledgements
This study was funded by an Australian Research Council Discovery Grant (DP0771005). Thanks go to Helena Britt, Lisa Valenti, and Clare Bayram at the Family Medicine Research Centre at the University of Sydney for providing and assisting with the data. Thanks to participants The Australian Health Economics Society Conference in 2008, the International Health Economics Association in 2009, and the Centre for Applied Economic Research Health Economics Workshop, UNSW, 2010.

Introduction

There is general agreement that the use of strong selective performance incentives in jobs that have multiple and complex tasks and objectives may not lead to efficient outcomes (Holmstrom and Milgrom, 1991; Baker, 1992). In the presence of variation in the principal's ability to measure, monitor and reward multiple objectives, performance incentives that focus only on a subset of poorly measured objectives are likely to re-direct agents' efforts towards the remunerated performance metrics and away from other objectives, such that the overall welfare effects are uncertain. Gaming may also be an issue as agents use private information to meet performance targets that may result in unintended consequences. Much of the literature suggests that incentive schemes should be 'low powered' in the presence of multitasking and where some areas of performance cannot be measured, or are measured with error. The argument for low powered incentive schemes is further strengthened in jobs with high intrinsic or pro-social motivation, such as teaching and medicine, where strong incentives may 'crowd out' intrinsic motivation (Frey, 2001; Mooney and Ryan, 1993). Many cite the above as reasons why high powered incentives schemes are rarely observed in practice, and much of the literature focuses on the optimal mix of remuneration to encourage efficient time allocation across tasks and outcomes (Ma and McGuire, 1997; Eggleston, 2005; Preyra and Pink, 2001;McGuire, 2000).

Despite the potential problems of using incentive schemes in complex jobs, a number of recent efforts at encouraging efficiency in health care have focused on using financial incentives to reward for increased quality of care through physician pay for performance. There are hundreds of pay for performance schemes operated by private health insurers in the US, many of which use tournament-based pay. Most schemes in the US are relatively low powered schemes with less than 10% of revenue derived from the incentive scheme, with the remainder coming from fee-for-service, capitation or salary. In the UK, a relatively high-powered scheme has been introduced where performance based pay is up to 50% of the revenue for some physicians, and is in place for almost all 30,000 general practitioners (GPs), who 'earn' points across 146 quality indicators across 11 disease areas. In Australia, an incentive scheme to improve quality of care has been in place since 2001 where the figure is around 10% of revenue, in addition to the system of fee-for-service.

These schemes often focus on what can be measured and include processes of care (e.g. the recording smoking status in medical notes) and 'indicators' of quality of care and health outcomes, such as high immunization and cervical screening rates, or clinical measures (e.g. blood glucose levels in diabetes) that have been shown in other research such as randomized clinical trials to be positively and strongly related to improved health status.

There is a large literature that evaluates and discusses these schemes, but only a handful of empirical studies exist with strong study designs such as randomized controlled trials (Gosden et al., 2000; Scott et al., 2010; Christianson et al, 2008). There is also much anecdotal evidence and theoretical discussion of the unintended consequences due to multi-tasking, such as the trade-off between cost and quality of care, though only a few empirical studies have examined this in the context of physician behaviour (Dumont et al., 2000; Gravelle et al., 2010; Sutton et al., 2009). In the 'high pwoered' UK scheme for GPs, Gravelle provides evidence of gaming, where GPs are allowed to exclude patients from the denominator of the performance measure. Also from the UK scheme, Sutton shows evidence from the UK scheme that there may be positive spillovers to unremunerated disease areas where there are complementarities in production. For example, the monitoring and recording of blood pressure may encourage doctors to record other important risk factors, such as smoking status. More generally, GPs who improve quality of care quality of care in one disease area, may improve in other disease areas as well if there are complementarities in production. Dumont et al (2007) show evidence of the effect of a change in the remuneration scheme on the allocation of physician time amongst labour supply decision, including time spent on administration and patient care, though this was a performance incentive scheme.

The aim of this paper is to examine whether GPs in an Australian incentive scheme change their allocation of time between tasks in remunerated and unremunerated health problems in consultations with patients who present with multiple health problems. Agency and asymmetry of information in the physician-patient consultation is a key source of market failure in health care, yet there remain few empirical studies examining the nature of consultations (Scott and Vick, 1998; Tai-Seale et al, 2008). In particular, we examine the extent to which the incentive scheme influences substitution between tasks, increases costs, or enables GPs to produce tasks more efficiently. A unique and rich dataset is used where GPs record tasks (i.e. prescribing, tests ordered, counselling, referrals) in each patient consultation, and a detailed set of patient and GP characteristics. The time costs of the consultation (consultation length) are modelled as a function of tasks (outputs) using a multiproduct cost function with a flexible functional form and with GP fixed effects to control for differences in unobserved characteristics of the GP's production function. We compare the marginal time costs of outputs and marginal rates of substitution between outputs for GPs in the incentive scheme with those not in the scheme. Since joining the incentive scheme is voluntary, we model selection into the scheme using a simple two-stage instrumental variable approach.

We add to the literature by examining time allocation within GP consultations for patients presenting with multiple health problems. We find evidence that GPs in the incentive scheme had a lower marginal cost per task for remunerated health problems, and so became more efficient in producing tasks. There is also evidence of positive spillovers for unremunerated health problems, which also had lower marginal costs per task. There are likely to be strong complemetarities in production in this context.

Institutional context

Medicare is Australia's national tax-financed health insurance scheme, which provides fully subsidised hospital treatment in public hospitals and subsidised visits to medical practitioners and pharmaceuticals. GPs in Australia charge patients a fee for each consultation subject to what the market will bear, and patients can claim back a fixed amount of the fee according to the Medicare Benefits Schedule. GPs can 'bulk-bill' if they do not charge patients and claim the schedule fee directly from Medicare. The majority of GPs are in private practice.

The Federal Government of Australia launched the Practice Incentive Program (PIP) in 1999 to augment the existing fee-for-service payment system. Around two-thirds of primary care practices had joined the PIP in 2008, with some variation over time. Joining depends on the expected costs and benefits to the practice, and practices who join may be systematically different to those who do not join. In addition to fee-for-service for each visit, PIP practices receive capitation payments for achieving a minimum level of practice infrastructure including payments for having IT for internet access and prescription software, being a teaching practice, and providing after hours care. They can also receive capitation sign-on payments for asthma, diabetes, mental health and cervical screening in exchange for maintaining disease registers. From November 2001, GPs in PIP practices which received sign-on payments were also eligible to claim Service Incentive Payments (SIPs) in these same four disease areas. This is an additional fee for the completion of a defined annual cycle of care (a sequence of visits), with activities in a cycle based on evidence-based clinical guidelines, or as in the case of cervical screening, the performance of a pap-smear to women who are aged between 20 and 69 and who have not been screened in the past four years. In May 2003, Outcome Payments (OPs) were introduced that paid additional income to the GP if a target level of cycles of care was reached (20% of all diabetes patients in a practice and 50% of all eligible patients in a practice are screened). Practices in remote areas received an additional loading of between 15% and 50% of total PIP payments. In 2003, PIP payments represented approximately 9% of a GP's average income (Commonwealth of Australia 2003). Table A1 in the appendix outlines the details of the payments in each disease area.

In-kind support to primary care practices in joining the PIP, which includes the need to become accredited and have IT infrastructure, was provided by the 120 or so Divisions of General Practice, which are regional primary care organizations covering all of Australia. They are run by primary care GPs and receive around 50% of their funding used to conduct specific programs to enhance primary care service delivery from the Federal Government.

Scott et al (2009) and Schurer et al (2010) evaluated the effect of the incentive scheme on quality of care in the four main disease areas using a bivariate probit model that accounted for selection into the scheme. They find the strongest effects for diabetes and weaker effects for asthma, mental health and cervical screening. Any unintended consequences of the scheme have not yet been evaluated.

Model and Hypotheses

Assume a principal-agent relationship in a consultation where the agent (GP) chooses a vector of tasks (*Y*) each of which have a time-cost to the GP. In a consultation with a patient, tasks may include the decision to provide advice on lifestyle, to order a test, to ask about medical history, or to prescribe medication. GPs perform the optimal mix of tasks to maximize a function of their own and of their patient's utility, subject to a time-cost constraint. Patient's utility is an argument in the GP's utility function. Each task is associated with a specific health problem *h*, with a number of possible tasks performed per health problem. Assume the patient presents with two health problems, one of which (Y_1) is included in the incentive scheme - the remunerated health problem - whilst the other (Y_2) is not - the unremunerated health problem. We assume that GPs minimize time costs by choosing the optimal mix of tasks for each consultation. We model the GPs' costs in the consultation as $C(Y_1, Y_2)$, where C is the length of the consultation which is a function of the number of tasks performed for each health problem. For GPs in the incentive scheme each task in Y_1 attracts additional remuneration over and above the fee for the consultation, whilst for those not in the incentive scheme the GP only receives the fee for the consultation.

There are a number of ways that the incentive scheme may affect GP behaviour within consultations. These are summarised in Figure 1, where: $\partial C^{PIP} / \partial Y_1^{PIP} - \partial C / \partial Y_1$ is the difference in the marginal time-cost of remunerated tasks between GPs in the PIP and GPs not in the PIP, and; $\partial C^{PIP} / \partial Y_2^{PIP} - \partial C / \partial Y_2$ is the difference in the marginal time-cost of unremunerated tasks between GPs in the PIP and GPs not in the PIP.

Figure 1. The effect of the incentive scheme on marginal time-costs

$\partial C^{PIP} / \partial Y_2^{PIP} - \partial C / \partial Y_2$

		+	=	-
$\partial C^{PIP} / \partial Y_1^{PIP} - \partial C / \partial Y_1$	+	A	B	C
	=	D	E	F
	-	G	H	I

First, the financial incentive may encourage GPs in the PIP to spend more time performing tasks for the remunerated health problem compared to GPs not in the PIP: the first row in the table: the marginal time-cost will be higher. In addition, the marginal time cost of tasks for the unremunerated health problem may also change. There may be positive spillovers in that the marginal time costs may increase, reflecting complementarities in the use of time between Y_1 and Y_2 (cell A). Cell B is where there is no change in the marginal cost of Y_2 . There may also have been substitution of time (cell C), such that the marginal time cost of unremunerated tasks will fall for those in the incentive scheme. This would be shown by a lower rate of marginal rate of substitution between the time per task of GPs in the incentive program compared to those not in the program.. Where time spent on tasks is perfectly substitutable and independent, then only Y_1 will be performed and Y_2 will be zero. For example, if a patient presents with diabetes and a common cold, then the GP may ignore the common cold and focus only on the diabetes. However, tasks are unlikely to be fully independent in this context as they are related to the same patient who may think that ignoring the cold is poor quality care. A GP sensitive to patients' preferences may therefore at the margin increase the time per task for diabetes and reduce the time per task for the common cold. Given that there is little GPs can do to cure the common cold, then this shift in the allocation of time may also be welfare enhancing. If the patient presented with diabetes and has a family history of breast cancer, then any substitution in time per task towards diabetes may not be welfare enhancing if the GP fails to recommend genetic counselling or a mammogram for breast cancer.

It is worth also noting that in cell A and B (and C depending on whether MRS is positive or negative), total consultation length could also increase. This is possible only if consultation length is a non-binding constraint, such that the GP has spare capacity during the day, which is unlikely given trends in falling hours of work, or if the GP increases his queues or turns other patients away. An unintended consequence could therefore be to reduce access for other patients, which may or may not be welfare reducing depending on their forgone health gains

Cell E is where the incentive scheme has no effect: there is no difference in marginal cost for Y_1 and Y_2 between GPs in the incentive scheme and those not in the incentive scheme. Cells D and F are unlikely to occur.

Further possible responses are shown in the final row of Figure 1: cells G to I. They show that the marginal costs may be lower for GPs in the incentive scheme. This is due to both changes in the technology of production, and reductions in informational assymetries in the GP-patient relationship. Other elements of the PIP, in addition to the incentive payments, included financial and administrative support for implementing new IT systems, patient registers and reminder/recall systems. This may mean that GPs are able to complete tasks more efficiently and so reduce the marginal time-cost of the remunerated tasks. In addition, the incentive scheme leads to more frequent visit for each patient as they complete cycles of care and are more regularly monitored - repeated consultations may reduce informational asymmetries between GPs and patients as each task does not have to be accompanied by lengthy explanation of alternatives, or medical history, or in persuading the patient to accept their treatment recommendations. Both of these factors may reduce the marginal time-cost of Y_1 for GPs in the incentive scheme.

If there are complementarities in production (cell I), then the marginal time cost of the unremunerated tasks may also fall, increasing the MRS $\partial Y_1 / \partial Y_2$. Better technology and more information exchange through more frequent visits are therefore also likely to influence other aspects of the patients care. Overall consultation length may be shorter for GPs in the incentive scheme. This may lead to increases in throughput overall, but a potential unintended consequence is reduced access for other patients as the frequency of visits for patients with remunerated health problems increase, and with a fixed GP capacity (hours worked per week), this means that fewer patients with unremunerated health problems can be seen.

Empirical model

We use a hybrid Diewert multiproduct cost function (Hall 1973) with two outputs (the number of GP tasks in health problem Y_1 and health problem Y_2) and a single input (time) which has a unit price. The time-cost to the GP of each consultation is both the single input and the measure of cost. The form of the cost function with one input is:

$$C = \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_{ij} \left(Y_i Y_j \right)^{1/2}$$

and with two tasks (m=2), this simplifies to:

$$C = B_1 Y_1 + B_2 Y_2 + B_{12} (Y_1 Y_2)^{1/2}$$

where $B_1 = \alpha_{11}$, $B_2 = \alpha_{22}$, $B_{12} = \alpha_{12} + \alpha_{21}$ and we can calculate a marginal time-cost for each task:

$$\frac{\partial C}{\partial Y_1} = B_1 + \frac{1}{2} B_{12} (Y_1 Y_2)^{-1/2} Y_2$$

$$\frac{\partial C}{\partial Y_2} = B_2 + \frac{1}{2} B_{12} (Y_1 Y_2)^{-1/2} Y_1$$

The ratio of these marginal time-costs provides the marginal rate of substitution between tasks in terms of time:

$$\frac{\partial Y_1}{\partial Y_2} = \frac{\partial C}{\partial Y_2} \bigg/ \frac{\partial C}{\partial Y_1}$$

The approach outlined so far includes only costs and outputs. Studies analyzing hospital costs (Granneman 1986, Scott and Parkin 1995, Carey 1997) have used control variables to capture observable heterogeneity in hospital costs. We follow this approach and include control variables in the right-hand side of the cost equation:

$$C = B_1 Y_1 + B_2 Y_2 + B_{12}(Y_1 Y_2)^{1/2} + X\beta$$

Effect of the PIP reform

We allow the PIP to influence all of the coefficients through the inclusion of interaction terms and to have an independent effect on consultation time:

$$C = B_1 Y_1 + B_2 Y_2 + B_{12}(Y_1 Y_2)^{1/2} + B_1^{PIP} Y_1 PIP + B_2^{PIP} Y_2 PIP + B_{12}^{PIP}(Y_1 Y_2)^{1/2} PIP$$
$$+\gamma PIP + X\beta$$

where ‘PIP’ is the indicator variable (=1) for the GP being in the Practice Incentive Program. We therefore examine whether the marginal time costs and marginal rates of substitution are different for those in the PIP compared to those not in the PIP. The overall effect of the PIP on total time costs (consultation length) can also be examined.

GP’s self-select into the PIP scheme and so the PIP variable and interaction terms in the time-cost function may be endogenous. We use an instrumental variable technique that is a modification of two-stage least squares where the first stage is a probit model predicting PIP membership from the exogenous explanatory variables in the cost function ($X\beta$) and some instruments Z:

$$PIP = \Phi\left(X\beta + Z\omega\right) + u$$

In the second stage, we insert the predicted probability of PIP ($\widehat{PIP}$) into the cost function:

$$C = B_1Y_1 + B_2Y_2 + B_{12}(Y_1Y_2)^{1/2} + B_1^{PIP}Y_1\widehat{PIP} + B_2^{PIP}Y_2\widehat{PIP} + B_{12}^{PIP}(Y_1Y_2)^{1/2}\widehat{PIP}$$
$$+\gamma\widehat{PIP} + X\beta$$

The intuition of this approach is that the $\widehat{PIP}$ includes only the 'exogenous' variation in PIP that is explained by $X\beta + Z\omega$. So by including PIP and its interactions with the number of tasks we allow the B^{PIP} coefficients to measure the 'exogenous' effect of the PIP on the B′s in the cost function.

METHODS

Data Source

Our analysis uses data from the *Bettering the Evaluation and Care of Health (BEACH)* study conducted by the Family Medicine Research Centre of the University of Sydney (Britt et al, 2009). Data are collected on 100 consecutive GP encounters (consultations) from a different random sample of 1,000 GPs each year, between 1998/9 and 2006/7. We therefore pool each of these yearly cross sections. Note the data are not a panel. Each encounter contains information on up to four health problems treated and for each problem, information on drugs prescribed, treatments conducted, whether counselling was provided, referral, and pathology tests and imaging ordered by the GP. The time the consultation begins and ends is also recorded on the encounter form. Problems managed are coded with the International Classification of Primary Care (ICPC-2) (WONCA 2005). The encounter form also collects data also provide information on the patient's age, gender, postcode, non-English speaking background, and Aboriginal or Torres Strait Islander status. A separate survey is also sent to participating GPs asking about age, gender, qualifications, and a range of practice characteristics, including whether the practice is accredited and the use of IT. Practice postcodes were used to merge the data with information on each Division of General Practice (Hordacre et al. 2006). These are geographically based primary care organisations covering all of Australia (111 in 2009) run by GPs, which provide various types of support to primary care services in their area.

Participation in the PIP

The data do not allow us to directly observe whether a practice participates in the PIP program. However, we do observe two practice characteristics that are necessary for participating and claiming incentive payments; whether the practice is accredited and the use of IT. A necessary condition to join the PIP is to become accredited against the standards of the Royal Australian College of General Practitioners (RACGP). In addition, around 85% of practices in the PIP claim capitation payments for the use of IT, and around 92% of practices in the PIP claim sign-on payments for diabetes, asthma and mental health which required establishing a patient register which has the ability to provide a recall and reminder service. Those in the PIP were therefore defined using a proxy variable created from practice characteristics reported in the BEACH GP survey: 'Accredited practices that use information technology (IT) for internet, prescribing, and medical records'.

To identify the effect of the PIP, we use an instrumental variables approach. As instruments we use 'The number of staff employed by a Division divided by the number of practices in the Division' and 'Prevalence rates of diabetes (or asthma) per 1,000 population'. The instruments are taken from the Annual Survey of Divisions of General Practice (Hordacre et al. 2006).

The number of Division staff per practice measures the 'intensity' of division support for general practices. Previous research has shown that the Divisions of General Practice influence general practices' participation in the PIP through practice infrastructure support, including by providing practice nurses to assist with claims and through information technology support (DeDominico et al 2005 Georgiou et al 2004; Scott and Coote 2007, 2009). Importantly, there is significant variation in the degree to which divisions provide this support (Scott and Coote 2007, 2009, Hordacre et al 2006, 2008). This is unlikely to have any impact on consultation length. Prevalence rates are a good indicator of population needs for claimable cycles of care and therefore influence the decision to join the PIP, but do not directly influence the quality of care provided in a specific encounter, which is more a function of patient and doctor characteristics.

As we have two instruments available for modeling each disease area, we are able to conduct an over-identification test that allows us to judge whether our instruments are valid. We test separately whether each instrument is significant in the outcome equation (1), while including the other in the PIP status equation (2).

Results

We use a sub-set of the BEACH including only consultations on female patients aged between 20 and 70, pooled between 2002/3 and 2006/7. Females are higher users of health care than males comprising over 60% of all encounters, and were more likely to have more than one problem managed compared to males (Britt et al., 2009).

The 'PIP status' variable shows that 51% of consultations are by GPs in the PIP according to our proxy measure (Table 1). The average consultation is almost 16 minutes long with a substantial standard deviation of 8.7 minutes. We can see that there is an average of almost one task for each remunerated health problem (nearly 1 per consult) compared to 0.52 tasks per consultation for unremunerated health problems. The number of tasks for remunerated health problems has a large standard deviation compared to those for remunerated health problems.

Table 2 shows the first stage probit equation which is used to predict PIP membership. The two instrumental variables are jointly significant though their individual significance is stronger for Division staff than it is for diabetes prevalence, which is insignificant. GPs aged between 45 and 60 are more likely to join the PIP than those under 45, whilst those over 60 are less likely to join. GPs in practices located in more affluent areas are more likely to join, as are GPs outside of major cities. GPs in larger practices are also more likely to join. There are also geographical differences in PIP membership. Relative to NSW, GPs in ACT are less likely to join, whilst those in Queensland, South Australia, and Victoria are more likely to join. There are also time trends in PIP membership, with more joining over time.

The regression results from the cost function are shown in Table 3. All signs are similar across models and are as expected, except for GP age and health care card, though the size of coefficients vary across models. Coefficients are in minutes. An increase in the number of tasks for both remunerated and unremunerated areas are positively related to consultation length, as are their interactions. There are strong non-linear effects of patient age, with consultation length increasing above the age of 40 years old. Aboriginal and Torres Strait Islander patients have longer consultations. The effect of GP age varies across models, but female GPs have consultation lengths over 2 minutes longer than male GPs. Consultation length is shorter for GPs who work fewer sessions, for GPs who graduated from non-OECD countries, and longer for GPs in practices in more affluent areas, and outside of major cities. There are no differences across states, and the time dummies are not significant in the OLS but are in the OLS+IV model.

Table 4 presents a joint test of significance of the PIP status variable and its interactions with remunerated and unremunerated tasks. The test statistic is not significant for the models with fixed GP effects. This suggests that 'within' variation in tasks performed by GPs across multiple consultations may not be sufficient to identify the effect of the incentive scheme on the time-cost function. The preliminary results also suggest that the instruments we use fail the overidentification tests. The identification strategy needs further work.

Preliminary results in relation to the hypotheses set out in Figure 1 are shown in Table 5. Under all specifications, marginal costs per task fall for remunerated health problems , and also fall for unremunerated health problems except in the OLS+FE model, where there is a small increase. The absolute reductions in the marginal time costs are largest for the remunerated health problems. The percentage reductions are largest for the remunerated health problems in the IV models, translating into an increase in the marginal rate of substitution shown in the last column.

The results correspond with our hypothesis that the incentive scheme was associated with an increase in the productive efficiency of consultations, because of either a higher level of IT support and/or reduced informational asymmetries. Although the effect of PIP is not statistically significant in the models with fixed-effects, the direction and order of magnitude of the estimated effect on marginal time-per-task similar. In the IV models without FE, there is evidence that overall consultation lengths are shorter for GPs in the incentive scheme compared to those not in the scheme.

Discussion

The preliminary results suggest that the Practice Incentives Program appears to reduce the time-per task for GPs for both remunerated and unremunerated health problems in the same consultation. Time per task is reduced more for the tasks for remunerated health problems. These results conform broadly with the hypothesis that GPs are able to reduce their time-per-task for remunerated disease areas tasks due to the efficiency–promoting elements of the incentive scheme, such as the use of information technology and repeated consultations (cycles of care). There seem to also be spillover effects of the incentive scheme to the tasks of unremunerated health problems. This further suggests that the use of IT and reduction of informational asymmetries are complementary in the production of tasks across different health problems of the same patient.

These results adds to the growing literature on the unintended consequences of pay-for-performance schemes and the importance of institutional frameworks and the role of administrative/technological support which accompany pay-for performance programs, and which may lead to positive spillovers, rather than 'undesirable' consequences due to substitution. These results are similar to Sutton et al, (2009) who also found evidence of positive spillovers and complementatires in production for the UK incentive scheme for GPs in the context of recording information about disease risk factors.

One interpretation of these effects is that our data cover a period sometime after the incentives were first introduced, and so may give a window into the longer term effects of such schemes. One would expect that immediately after the introduction of such a scheme that marginal costs per task would rise as there is a focus on the remunerated health problems, and so more tasks would be undertaken and more time spent on each task. However, over time as patients visit more frequently, their chronic diseases may become more manageable and so fewer tasks will need to be performed - though patients will still require monitoring - and those that are performed are done so more efficiently. Our data may therefore reflect the period sometime after the incentive scheme has established itself.

The results presented are preliminary, and further econometric testing needs to be conducted to establish the robustness of our results, particularly the IV strategy used, where standard errors in the cost function still need to be adjusted to reflect our use of a predicted variable. Nevertheless, the general result of lower marginal costs per task does seem to be robust across all current specifications of the model.

The data does not allow us to examine the welfare effects of multitasking. For example, though a higher throughput because of shorter consultations would be considered to be an improvement in efficiency, this does not consider effects on health outcomes or other aspects of care valued by patients. Furthermore, the observed effects may be because the incentive scheme encourages more frequent visits (cycles of care) for patients with remunerated health problems. With a fixed GP capacity in terms of hours worked per year, this means that other patients with unremunerated health problems have to wait longer for an appointment, and so access to care is reduced. Further research is required on how incentive schemes change the allocation of physician time across patients, as anecdotal reports of ‘patient dumping’ and selection are common, with some empirical evidence to back this up (Gravelle et al, 2010).

References

Baker (2002) The use of performance measures in incentive contracting. American Economic Review 415-420.

Baker G (1992) Incentive contracts and performance measurement. Journal of Political Economy, 43, 598-614.

Britt H, Miller GC, Charles J, Henderson J, Bayram C, Valenti L, Pan Y, Harrison C, Fahridin S & O'Halloran J. (2009) General practice activity in Australia 1999-00 to 2008-09: 10 year data tables General practice series no. 26. Australian Institute of health and Welfare: Canberra.

Christiansen JB., Leatherman S., Sutherland K. (2008) Lessons from evaluations of purchaser pay for performance programs: a review of the evidence. Medical Care Research and Review 65;5S.

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

Frey BS, Jegen R. (2001) Motivation Crowding Theory. Journal of Economic Surveys 15(5):589-611.

Gosden T, Forland F, Kristiansen I, Sutton M, Leese B, Giuffrida A, et al. (2000) Capitation, salary, fee-for-service and mixed systems of payment: effects on the behaviour of primary care GPs. Cochrane Database of Systematic Reviews.

Gravelle H, Sutton M, Ma A. (2010) Doctor behaviour under a pay for performance contract: treating, cheating and case finding? Economic Journal (forthcoming)

Holmstrom B., Milgrom P. (1991) Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. Journal of Law, Economics, & Organization, 7, 24-

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

McGuire T (2000) *Physician agency*. In: Culyer AJ, Newhouse JP. Handbook of Health Economics, Volume 1A. Elsevier.

Mooney G, Ryan M (1993) Agency in health care: getting beyond first principles. *Journal of Health Economics*, **12**, 125-238.

Preyra C., Pink G (2001) Balancing incentives in the compensation contracts of nonprofit hospital CEOs. Journal of Health Economics 20;4:509-525.

Scott A., Vick S. (1999) Patients, doctors and contracts: an application of principal-agent theory to the doctor-patient relationship. Scottish Journal of Political Economy 46:111-134.

Schurer S, Sivey P, Scott A, Jensen P. (2010) Evaluating the Effects of Pay-for-Performance, Information Technology, and Accreditation on Health Care Quality in the Presence of Treatment Selection Bias. Health Services Research (forthcoming)

Scott A, Schurer S, Jensen PH, Sivey P. (2009) The effects of an incentive program on quality of care in diabetes management. Health Economics. 18:1091-108.

Sutton M, Elder R, Guthrie B, Watt G. (2009) Record rewards: the effects of targeted quality incentives on the recording of risk factors by primary care providers. Health Econ. 2009 (accepted for publication).

Scott A and Parkin D (1995) Investigating Hospital Efficiency in the New NHS: The Role of the Translog Cost Function. Health Economics, 4: 467-478

Tai-Seale M, Mc Guire T and Zhang, W (2007) Time Allocation in Primary Care Office Visits. Health Services Research 42 (5) pp1871-1894

Hall R E (1974) The Speci.cation of Technology with Several Kinds of Output. Journal of Political Economy 81 pp878-892

Carey, K (1997) A Panel Data Design for Estimation of Hospital Costs Functions Review of Economics and Statistics pp443-453

Caves D W, Christensen L R and Tretheway, M W (1980) Flexible Cost Functions for Multiproduct Firms. Review of Economics and Statistics 62 (3) pp477-481

Grannemann T W , Brown, R S and Pauly M V (1986) Estimating Hospital Costs: A Multiple-Output Analysis Journal of Health Economics 5 pp107-127

Dumont E, Fortina B, Jacquemet N, Shearer B. (2008) Physicians' multitasking and incentives: Empirical evidence from a natural experiment. Journal of Health Economics 27; 1436–1450

Appendix

Table A1. Quality Indicators From the Four Disease Areas Eligible for Incentive Payments in the Practice Incentive Program (PIP)

Disease Area	**Activity**	**Service Incentive Payment (SIP)**	**Indicator Used for Analysis**
Diabetes Targets patients with Type I and Type II diabetes.	Annual cycle of care within 12 months, whereby a cycle of care comprises 12 quality items such as blood tests, eye and foot examinations, dietary and life-style advices.	Sign-on payment: 1,000 A$ per FTE GP, 40 A$ per completed cycle per patient, Outcomes Payment of 5 A$ per quarter for each diabetes patient if 20% target level is reached.	HbA1c test
Asthma 3+ Visit Plan Targets patients with medium and severe asthma.	At least three asthma related consultations within the past four weeks to four months, whereby the cycle of care involves documenting the diagnosis, reviewing the use of medication and asthma devices, providing a written asthma plan and self-management education, and a review of the fulfillment of the asthma plan. This requirement was replaced in 2005 with a requirement of two visits within 12 months.	Sign-on payment: 250 A$ per FTE GP, 100 A$ per completed cycle per patient.	Spirometry test
Cervical Screening Targets female patients between 20 and 69 who have not been screened in the past four years.	Order or conduct a pap-smear.	Sign-on payment: 250 A$ per FTE GP, 35 A$ for every screened female patient. Outcomes Payment of 3 A$ per eligible patient per year to practices where at least 50 per cent of eligible patients are screened in a 30-month reference period.	Pap-smear
Mental Health 3 Step Mental Health Process as part of the Better Outcomes in Mental Health Care Program. For GPs who have undergone specific training.	Process requires GPs to: undertake initial mental health assessments of their patients; develop written mental health plans; and review the progress of their patients against the mental health plans. Reviews will require a recall consultation to review progress and make amendments, if necessary within 4 weeks to 6 months.	Sign-on payment: 150 A$ per FTE GP, 150 A$ for each completed cycle.	Psychological or psychiatric counseling

Table 1: Descriptive Statistics

Variable	Mean	Std. Dev.	Min	Max
PIP Status	0.514	0.500	0	1
C: Consultation length (minutes)	15.897	8.691	1	155
Number of Problems managed	1.573	0.796	1	4
Y_1: Number of tasks per encounter for remunerated health problem	0.094	0.321	0	4
Y_2: Number of tasks per encounter for unremunerated health problem	0.521	0.746	0	6
Division staff/GP practice	0.161	0.112	0	1
Division diabetes prevalence	170.074	10.254	0	1
GP age >45	0.386	0.487	0	1
GP age >60	0.152	0.359	0	1
GP Female	0.464	0.499	0	1
GP number of sessions worked per week	7.904	2.424	0	1
Practice SEIFA	6.678	3.232	0	1
Inner regional	0.206	0.404	0	1
Outer regional	0.075	0.264	0	1
Graduated overseas: OECD	0.097	0.296	0	1
Graduated overseas: non-OECD	0.137	0.344	0	1
Practice size: 2 to 4 GPs	0.350	0.477	0	1
Practice size: 5 plus GPs	0.552	0.497	0	1
ACT	0.020	0.141	0	1
NT	0.000	0.000	0	1
QLD	0.189	0.391	0	1
SA	0.078	0.269	0	1
TAS	0.029	0.167	0	1
VIC	0.230	0.421	0	1
WA	0.086	0.280	0	1
Year - 2004	0.181	0.385	0	1
Year - 2005	0.204	0.403	0	1
Year - 2006	0.210	0.407	0	1
Aboriginal	0.010	0.102	0	1
Non-english speaking	0.092	0.289	0	1
Health care card	0.387	0.487	0	1
Patient age	44.673	14.113	0	1

Table 2: First stage probit model of PIP status:

	Coeff.	S.E.
Division staff/GP practice	0.344	0.139
Division diabetes prevalence	0.003	0.002
GP age >45	0.045	0.020
GP age >60	-0.099	0.025
GP Female	-0.033	0.020
GP sessional	0.000	0.004
Practice SEIFA	0.006	0.003
Inner regional	0.151	0.029
Outer regional	0.155	0.042
Graduated overseas: OECD	0.017	0.030
Graduated overseas: non-OECD	-0.037	0.027
Practice size: 2 to 4 GPs	0.228	0.031
Practice size: 5 plus GPs	0.430	0.027
ACT	-0.131	0.062
QLD	0.140	0.025
SA	0.076	0.036
TAS	0.013	0.054
VIC	0.107	0.024
WA	0.002	0.034
Year – 2004	0.111	0.023
Year - 2005	0.176	0.022
Year - 2006	0.212	0.023
Pseudo R^2	0.141	
Obs	63529	

Table 3: Full regression results for three alternative model specifications.

Variable	OLS: No IV or FE's Coeff	S.E.	FE, No IV's Coeff	S.E.	IV, No FE's Coeff	S.E.	IV and FE's Coeff	S.E.
PIP	0.295	0.146			-6.166	2.100		
Y1	5.042	0.258	4.415	0.226	5.821	0.503	4.814	0.422
Y1*PIP	-0.304	0.356	-0.068	0.310	-1.719	0.827	-0.738	0.710
Y2	2.055	0.125	1.594	0.083	2.100	0.213	1.541	0.155
Y2*PIP	-0.444	0.163	0.064	0.117	-0.568	0.370	0.180	0.284
(Y1*Y2)1/2	1.299	0.402	0.878	0.367	1.809	0.765	1.188	0.692
(Y1*Y2)1/2*PIP	-1.126	0.533	-0.876	0.484	-2.153	1.321	-1.572	1.195
Age 25 to 29	0.509	0.161	0.435	0.151	0.489	0.153	0.401	0.141
Age 30 to 34	0.454	0.157	0.236	0.147	0.471	0.152	0.240	0.140
Age 35 to 39	0.841	0.165	0.483	0.152	0.885	0.161	0.499	0.146
Age 40 to 44	1.142	0.165	0.745	0.153	1.158	0.161	0.721	0.148
Age 45 to 49	1.314	0.164	0.965	0.151	1.365	0.161	0.974	0.145
Age 50 to 54	1.553	0.167	0.963	0.152	1.625	0.164	0.984	0.145
Age 55 to 59	1.195	0.162	0.748	0.150	1.267	0.161	0.778	0.145
Age 60 to 64	1.187	0.171	0.657	0.158	1.181	0.168	0.637	0.149
Age 65 to 69	0.914	0.170	0.255	0.158	0.923	0.167	0.238	0.151
Aboriginal	2.298	0.524	0.259	0.519	0.915	0.485	0.017	0.374
Non-English speaking	-0.295	0.182	-0.047	0.153	-0.432	0.186	-0.064	0.153
Health care card	-0.208	0.099	0.110	0.084	-0.163	0.097	0.097	0.082
GP age >45	-0.194	0.158			0.051	0.176		
GP age >60	0.438	0.228			-0.058	0.295		
GP Female	2.025	0.189			1.835	0.198		
GP sessional	-0.184	0.036			-0.191	0.036		
Practice SEIFA	0.144	0.026			0.174	0.027		
Inner regional	0.164	0.183			1.536	0.461		
Outer regional	0.577	0.260			1.874	0.549		
Graduated overseas: OECD	-0.374	0.222			-0.297	0.222		
Graduated overseas: non-OECD	-1.046	0.199			-1.258	0.213		
Practice size: 2 to 4 GPs	-0.704	0.316			0.454	0.504		
Practice size: 5 plus GPs	-0.520	0.323			2.004	0.878		
ACT	-0.080	0.479			-0.867	0.538		
QLD	-0.348	0.204			0.489	0.328		
SA	0.149	0.312			0.760	0.372		
TAS	0.357	0.494			0.524	0.496		
VIC	-0.683	0.195			-0.020	0.289		
WA	-0.674	0.259			-0.659	0.264		
Year – 2004/5	0.214	0.200			0.873	0.302		
Year – 2005/6	-0.046	0.193			0.971	0.387		
Year – 2006/7	0.327	0.194			1.616	0.445		
Constant	13.819	0.534	14.027	0.112	13.913	0.527	13.999	0.109
R^2	0.1016		0.0720		0.1049		0.0734	
obs	58992		58992		58992		58992	

Table 4: Econometric Tests

Test	Test statistic	p-value
Effect of PIP $B_1^{PIP} = B_2^{PIP} = B_{12}^{PIP} = \gamma^{PIP} = 0$		
Model 1: OLS	4.16	0.002***
Model 2: FE, No IV	1.74	0.156
Model 3: IV, No FE	5.02	0.001***
Model 4: IV & FE	1.81	0.142
Exogeneity (Model 1 = Model 2)		
Overidentifying restrictions	Diabetes rate: 2.74	0.027**
(Model 3)	Division staff: 1.99	0.093*
Instrument strength		

Table 5. Marginal time-costs and marginal rates of substitution

		(∂C)/(∂Y1)	(∂C)/(∂Y2)	(∂Y1)/(∂Y2)
OLS, no IV or FE's	Non-PIP	6.57	2.33	0.354
	PIP	4.94	1.64	0.332
	Difference[1]	***-1.63***	***-0.69***	***-0.022***
	% difference	***-24.8%***	***-29.6%***	***-6.2%***
OLS, FE	Non-PIP	5.44	1.78	0.327
	PIP	4.34	1.66	0.382
	Difference[1]	***-1.10***	***+0.13***	***0.055***
	% difference	***-20.2%***	***+7.3%***	***17.0%***
IV, No FE's	Non-PIP	7.94	2.48	0.312
	PIP	3.7	1.46	0.395
	Difference[1]	***-4.24***	***-1.02***	***0.083***
	% difference	***-53.4%***	***-41.1%***	***26.6%***
IV and FE's	Non-PIP	6.2	1.79	0.289
	PIP	3.6	1.64	0.456
	Difference[1]	***-2.6***	***-0.15***	***0.167***
	% difference	***-41.9%***	***-8.4%***	***57.8%***

1. $\partial C^{PIP} \big/ \partial Y_i^{PIP} - \partial C \big/ \partial Y_i$