HEDG Working Paper 08/15

# The Effects of Financial Incentives on Quality of Care: The Case of Diabetes

Anthony Scott
Stefanie Schurer
Paul H. Jensen
Peter Sivey

July 2008

ISSN 1751-1976

http://www.york.ac.uk/res/herc/research/hedg/wp.htm

# The Effects of Financial Incentives on Quality of Care: The Case of Diabetes*

Anthony Scott, Stefanie Schurer† Paul H. Jensen, and Peter Sivey

July 11, 2008

**Keywords**: Pay-for-performance, multivariate probit models, health care systems.
**JEL Classification**: I11, C21.

*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.

†Address for correspondence: Stefanie Schurer, Melbourne Institute of Applied Economic and Social Research, The University of Melbourne, Melbourne, 3010, Victoria, Australia, sschurer@unimelb.edu.au, Tel: +61-3-83442190, and Fax: Tel: +61-3-83442111.

**Abstract**

Australia introduced an incentive payment scheme for general practitioners to ensure systematic and high quality care in chronic disease management. There is little empirical evidence and ambiguous theoretical guidance on which effects to expect on the quality of care. This paper evaluates the impact of the payment incentives on quality of care in diabetes, as measured by the probability of ordering an HbA1c test. The empirical analysis is conducted with a unique data set and a multivariate probit model to control for the simultaneous self-selection process of practices into the payment scheme and larger practices. The study finds that the incentive reform had a positive effect on quality of care in diabetes management and that participation in the scheme is facilitated by the support of Divisions of General Practice.

# 1 INTRODUCTION

A recent development in the remuneration of health care providers, and of doctors in particular, is pay-for-performance. Remuneration is conditional on achieving pre-specified levels of activity, standards of care, or behaviours in specific disease areas based on guidelines of best practice. These systems are often being superimposed on existing remuneration schemes, creating more complex blended payment systems that aim to re-orient the payment scheme away from structures and processes towards health outcomes.

Even though there is a growing literature on performance-based pay in health care (Gaynor *et al.*, 2001; Chalkley and Tilley, 2006; Lippi Bruni *et al.*, 2007; Simoens and Giuffrida, 2004; Shen *et al.*, 2004), it is difficult to assess the possible effects of financial incentives on the actual behaviour of general practitioners (GPs). Some argue that in complex areas of service provision pay-for-performance should not be used, as the potential to distort behaviour away from an organisation's broader goals, which may lead to unintended consequences, is high (Prendergast, 1999; Mooney and Ryan, 1993). Payment incentives may crowd-out pre-existing intrinsic motivation and therefore lead to lower levels of performance (Frey and Jegen, 2001). Providers may have been performing well in absence of the reform or providers may manipulate data or even submit fraudulent claims to absorb rents from the incentive payments.

Empirical evidence on possibly the most elaborate pay-for -performance scheme, the Quality and Outcomes Framework (QoF) in the UK (Roland, 2004), suggested a high level of quality of care during the first year (Doran *et al.*, 2006). However, Doran *et al.* (2006) and Gravelle *et al.* (2008) report evidence of gaming and concerns that many GPs were already meeting the standards of care in the QoF, implying the payment reform may not have changed quality of care.

Australia introduced a pay-for-performance scheme for GP remuneration in chronic disease management in 1999. The Practice Incentive Program (PIP), superimposed on the existing fee-for-service payment system, includes capitation payments for practice infrastructure and pay-for-performance to encourage higher quality of care in diabetes, asthma, mental health and

cervical screening.

To date, there has been no systematic assessment of the effects of the PIP reform on quality of care. We therefore study the effects of this scheme on the quality of care in diabetes, a major chronic disease with clear and evidence-based indicators of quality of care. Quality of care is measured by whether an HbA1c test was ordered during a consultation. We then compare the average proportion of HbA1c tests ordered by GPs that have joined the PIP scheme (treatment group) with those ordered by GPs that have not joined (control group).

The paper adds to the literature in several ways. First, a unique data-set is used that includes detailed information on the content of GP consultations, and therefore, actual clinical behaviour is observed rather than fee claims. Data is available for several consultations conducted by one GP, and therefore theoretically, GP-specific heterogeneity can be controlled for. Second, the endogenous nature of general practices' participation in the PIP and the empirically observed simultaneous growth trend of larger general practices are handled by applying a trivariate probit model. We merge longitudinal data on the activities and characteristics of each GP's Regional Primary Organisation (Division of General Practice) with our data-set. These data are used as exclusion restrictions to identify the three simultaneous equations of HbA1c tests, PIP participation and practice size. Whether the use of instruments is necessary to identify the effects of interest is also examined. Wilde (2000) suggests that exclusion restrictions are not necessary under certain conditions and thus we add to the debate of the appropriate specification of multivariate probit models.

The main result of our paper is that the incentive reform had a positive effect on quality of care in diabetes management. Practices located in Divisions which provided support services were more likely to join the PIP scheme, and therefore, increase quality of care. This effect, however, is only made observable through the appropriate choice of exclusion restrictions and by controlling for the correlation of unobservable factors that positively influence the probability of increasing quality of care and negatively influence the probability of joining the incentive reform.

The remainder of the paper is structured as follows: Section 2 explains

the institutional design of the incentive payments. Section 3 presents the theoretical model that guides hypothesis formulation for the empirical tests. Section 4 outlines the empirical framework and Section 5 describes the data used. In Section 6 we present a descriptive analysis of the data (6.1) and estimation results from the pooled and multivariate probits (6.2). Section 7 discusses the results in light of the existing literature.

# 2 INSTITUTIONAL BACKGROUND

In Australia, the national universal insurance scheme, Medicare, is funded through general taxation. Medical services, including GP and specialist services, are directly funded by the federal government through uncapped fee-for-service.

GPs act as gatekeepers to specialist care. Medicare funding of GP services is based on the Medicare Fee Schedule. GPs can charge patients what the market will bear - there are no price controls. Patients can then claim a fixed rebate (the schedule fee) from Medicare, thus facing a co-payment equal to the difference between the rebate and the price charged by the GP. Alternatively, GPs can claim the schedule fee directly from Medicare and only charge patients the co-payment. GPs can also 'bulk-bill' and do not charge the patient a co-payment and receive 100% of the rebate directly from Medicare as full payment. The fee schedule for GPs is based on payment for each consultation provided, with the fee rising across four different levels of complexity and consultation length. Fees range between AUS $ 15 and AUS $ 91.70 in 2008. The fee structure provides incentives for a high throughput of patients and shorter consultations, which in turn leads to higher costs because of inappropriate prescribing and referral and a poor quality of care (Scott and Shiell, 1997; Beilby, 2003). There is no patient registration and there are no restrictions on where GPs may establish their practice, except for doctors emigrating to Australia.

In addition to the Medicare system of fee-for-service reimbursement, blended payments for GPs were introduced in 1999 through the Practice Incentive Program (PIP). In 2003 this program contributed to almost 10% of GPs' income (Harris *et al.*, 2004). The aim was to move away from the fee-

for-service model towards a system of remuneration linked to the provision of quality of care rather than volume. Practices joining the PIP scheme were required to be accredited from 1 January 2001. Practices that participated in the PIP scheme before that date were required to be fully accredited by 1 January 2002. Practices that participated in the PIP scheme for the first time from 1 January 2001 onwards, must have been registered for accreditation by the date of joining and needed to be accredited within 12 months of joining. The accreditation process involves both administrative and financial burden. GP practices have to pay a fixed fee to accreditation agencies and comply with a range of organisational restructuring requirements. Accreditation takes place on the basis of the standards of the Royal Australian College of General Practitioners and is conducted by two independent not-for-profit organisations which were established in 1997 and 1999. Once accredited, the status is valid for three years.

PIP practices can receive capitation payments for achieving a minimum level of practice infrastructure including payments for having IT infrastructure, 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 so practices can provide follow-up care. General practices in remote areas receive an additional loading of 15% to 50% of the total PIP remuneration depending on the geographical size of the region of the practice location and the remoteness of the practice.

From November 2001, PIP practices who were signed on for the chronic disease payments were also eligible to claim Service Incentive Payments (SIPs) in the areas of asthma, diabetes, mental health and cervical screening. This is an additional fee for the completion of a defined annual cycle of care (a sequence of visits) based on evidence-based clinical guidelines. The first pay-out of SIPs for eligible practitioners were made in February 2002 for the completion of a cycle of care that was recorded from 1 November 2001 onwards.

In May 2003, Outcome Payments (OPs) were also introduced that provided additional remuneration for completing a cycle of care for a certain percentage of the population in each disease area.

The incentives in the PIP made it financially worthwhile for some practices to become accredited, and it is worth noting that at the time the PIP and accreditation were introduced, the proportion of solo general practices in Australia began to fall markedly. The fixed set up and administrative costs of accreditation, joining the PIP scheme, and claiming SIP and OP could more easily be borne by larger general practices, and this may have encouraged practices to become larger in order to benefit from economies of scale and scope. The decision to become accredited, join the PIP, and increase practice size were likely to have been made simultaneously for many general practices.

In addition, general practices were being encouraged and supported to join the PIP scheme and become accredited by their local Divisions of General Practice. Divisions provided information, resources and advice to local practices about accreditation, joining the PIP scheme and claiming SIPs. Divisions are geographically based primary care organisations, established in 1992. There are currently 115 across Australia with a membership of close to 98 % of all GPs (Hordacre *et al.*, 2006). A key change since the late 1990s is the use of Divisions by the Federal government and other agencies to deliver funding for specific health care services and schemes (Australian Division of General Practice, 2005).

For diabetes, the sign-on capitation payment is paid to practices who have a register of diabetes patients and a recall system. In 2008, this is AUS $ 1 per patient, about AUS $ 1,000 for the average GP. These capitation payments were introduced in August 1999.

The SIP is paid on completion of an annual cycle of care and can be claimed only at the last encounter of the cycle. In 2008, this payment was AUS $ 40 per completed cycle of care.

Claims for the OP can be made only by practices where 20 % of diabetes patients have completed an annual cycle of care and at least a minimum of 2 % of all their patients are patients with diabetes. For diabetes, this payment is AUS $ 20 per patient in 2008.

To claim a cycle of care, a GP needs to perform the following services for a diabetes patient (Diabetes Australia, 2007, p. 28):

- 6 monthly: measurement of height, weight and BMI, blood pressure and foot examination.
- 12 monthly: order an HbA1c test (also called glyctest), a lipid profile test, and a urine microalbumin test.
- 24 monthly: ensure one eye-examination.
- All Diabetes 2 patients must initially see an ophthalmologist or optometrist (and then every two years).
- Educate patient on diet and diabetes.
- Offer assistance with smoking cessation.
- Review medication.
- Referral to a diabetes educator (initially).
- Referral to podiatrist's help if neuropathy, peripheral vascular disease, foot abnormality or calluses was diagnosed.

# 3 THEORETICAL FRAMEWORK

We examine the effect of PIP on quality of care in diabetes. The incentives in the PIP for increasing quality of care include the combined effects of capitation payments, a Service Incentive Payment (SIP) and an Outcomes Payment (OP). The evaluation problem we face is due to the voluntary nature of participating in the PIP scheme and the possibility that, even if a practice joined the scheme, a GP within that practice may not increase quality of care. In this respect, two decisions have to be modelled: (i) the decision of the practice to join the PIP scheme; and (ii) the decision of an individual GP to provide and finish a cycle of care and, therefore, to increase quality of care.

A general practice $g$ will join the PIP if the utility derived from joining is strictly greater than the utility from not joining: $U_g^{PIP} > U_g^{NOPIP}$. The

utility from joining is[1]:

$$U_g^{PIP} = U\left[H\left(n^c\left(m_g, e_g\right)\right), Y\big((pm_g + n_g p^{CAP} + n^c\left(m_g, e_g\right) p^{SIP} + p^T\left(\frac{n^c\left(m_g, e_g\right)}{n_g}, t\right) + Y_r\big) - C(m_g, \gamma^{SIP} e_g, F_{PIP})\right], \quad (1)$$

and the utility from not joining the PIP is:

$$U_g^{NOPIP} = U\left[H\left(n^c\left(m_g, e_g\right)\right), Y\left(pm_g\right) - C(m_g, e_g, 0)\right] \quad (2)$$

The general practice's utility is assumed to be a function of the health gains ($H$) of each patient who obtains quality care ($n^c$), and income $Y(.)$ net of costs $C(.)$.

Independent of joining the PIP, income for practices depends on the number of consultations ($m$) conducted and the total price charged to each patient for each consultation ($p$)[2]. Income for practices participating in the PIP scheme is supplemented by capitation payments ($p_{CAP}$) per patient ($n$), by SIPs ($p_{SIP}$) that depend on the number of completed cycles of care ($n^c$), and OPs ($p_T$) if the number of patients receiving a completed cycle of care reaches a target percentage of all eligible patients. In simple terms, a PIP practice can increase its average revenue per consultation if it joins the PIP and if its GPs are able to claim SIPs and OPs. Rural practices receive extra revenues $Y_r$ due to the rural loading.

Independent of the PIP status, the practice's cost function depends on effort ($e$) and the number of consultations ($m$). Practice costs are higher in the PIP scheme since joining PIP incurs fixed financial and administrative costs ($F_{PIP}$) related to accreditation. For large practices the cost per employed GP is smaller, since the capitation payment is based on the number of patients seen, but the accreditation cost is independent of practice size. On the other hand, practices who claim SIP need to make an extra effort in improving quality and in ensuring that their patients attend scheduled consultations for completion of the cycle. We introduce an extra parameter

[1]The notation is defined in Table I.

[2]This price includes the Medicare fee rebate that is fixed for all consultations and the patient co-payment which is variable.

Table I. Notation

| Variable | Definition |
|---|---|
| $n$ | Diabetes patients |
| $n^c$ | Cycles of care associated with quality of care |
| $m$ | Consultations |
| $e$ | GPs effort in producing cycles of care |
| $H$ | Health of patients |
| $Y$ | Doctor's/practice's income |
| $C$ | Doctor's/practice's costs |
| $p$ | Medicare FFS payment |
| $p^{CAP}$ | PIP capitation payment |
| $p^T$ | PIP outcome target payment |
| $p^{SIP}$ | SIP payment |
| $F^{PIP}$ | Fixed cost of joining PIP scheme |
| $i$ | Subscripts GPs |
| $g$ | Subscripts general practice |
| $Y_r$ | Rural loading |
| $\gamma^{SIP}$ | Extra effort required for GPs to claim SIPs |

$\gamma^{SIP} > 1$ in the cost function to represent the extra effort.

GPs also derive utility directly from their patients' health status ($H$). What the literature now calls 'intrinsic motivation' (see Francois and Vlassopoulus (2007) for a survey) may be the main driver for providing quality-enhancing care.

Linking the direct arguments of the practice utility function through $m_g$ and $n^c$ we assume patient health is an increasing function of the number of cycles of care, which is in turn an increasing function of the number of patients and effort, $H\left(n^c(m_g, e_g)\right)$. Effort and consultations link cycles of care to the cost function.

A general practice decides to join the PIP scheme, if for optimal values of $m_g$ and $e_g$ the marginal benefit, through patient health and revenue is greater than the marginal cost, in terms of the utility cost of effort and the monetary cost of participation and consultations. Practices not joining the PIP have no monetary incentive for quality-enhancing effort, only an incentive via the patient's health.

We can see from (1) that practices will be more likely to join the PIP the higher are any of the PIP payments ($p^{CAP}, p^{SIP}, p^T$), the lower is the target threshold $t$ and the less costly it is to produce cycles of care ($n^c(.)$). Equally, practices will be less likely to join PIP, the higher is the fixed cost $F^{PIP}$ and the extra effort $\gamma_{SIP}$.

The second decision concerns GPs within practices who have joined the PIP to complete the quality-enhancing cycles of care in diabetes. In our setting, finishing a cycle of care and claiming the SIP is analogous to improving quality of care. An individual GP $i$ chooses to increase quality of care if $U_i^{SIPCLAIM} > U_i^{NOSIPCLAIM}$, whereby the utility from GP increasing quality is defined as:

$$U_i^{Sipclaim} = U\left[H\left(n^c\left(m_i, e_i\right)\right), Y\left(pm_i + n_g p^{CAP} + n^c\left(m_i, e_i\right) p^{SIP} + p^T\left(\frac{n^c\left(m_i, e_i\right)}{n_g}, t\right)\right) - C(m_i, \gamma^{SIP} e_i)\right], \quad (3)$$

and the utility derived from not claiming SIP is defined as:

$$U_i^{Nosipclaim} = U\left[H\left(n^c\left(m_i, e_i\right)\right), Y\left(pm_i + n_g p^{CAP} + p^T\left(\frac{\sum_i n^c\left(m_i, e_i\right)}{n_g}, t\right)\right) - C(m_i, e_i)\right]. \quad (4)$$

The SIP payment $p^{SIP}$ and Medicare reimbursement $p$ are paid to individual GPs based on their cycles of care (SIPs) and consultations. Strictly speaking, the capitation $n_g p^{CAP}$ and the target payment is shared between GPs of the same practice and, thus, should be divided by the number of GPs in the practice. For simplicity, we abstract from the income sharing rule[3].

In a reduced form representation, the optimal number of completed cycles of care $n^{c*}$ can be expressed as:

$$n^{c*} = n^{c*}(p, p^{CAP}, p^{SIP}, p^T, \gamma^{SIP}, F_{PIP}). \quad (5)$$

It is a function of the price paid for each consultation by Medicare ($p$),

[3]In the case of a shared income rule for the Outcome Payment (OP), individual GPs would have an incentive to free-ride on the work of other GPs. Such a scenario is one explanation for a lower up-take of SIP claiming.

of the prices for the capitation payment ($p^{CAP}$), for the SIP claim ($p^{SIP}$), and for the target payment ($p^T$), of the quality enhancing parameter ($\gamma^{SIP}$) and of the fixed cost to join the PIP ($F_{PIP}$). The functional form of Eq. (5) depends on the functional form of the utility function, of the health production function, and of the revenue and cost functions in the PIP and SIP claiming equations.

In sum, the optimal level of quality of care depends on whether a practice joins the PIP. PIP can increase or decrease practice revenue depending on size of associated costs, and these in turn depend negatively on the parameter $\gamma_{SIP}$ and the fixed costs $F^{PIP}$, and positively on the revenues $p^{CAP}$, $p^{SIP}$ $p^T$. Some practices may choose to join while others won't. Practices that manage to decrease quality-enhancing effort $\gamma_{SIP}$ are more likely to increase quality of care and claim SIP. This could happen, for instance, through the activities of Divisions of General Practice.

# 4 EMPIRICAL SPECIFICATION

In a first step, we estimate the individual GP's decision to provide a cycle of care following Eq. (5). Let's assume the general practitioner $i$ derives utility from increasing the quality of care in consultation $j$, but this latent utility $Y_{ij}^*$ is unobserved:

$$Y_{ij}^* = \alpha + \beta_1 PIP_i + \gamma' X_i + \gamma' X_j + \delta' T_t + u_{ij} \tag{6}$$

We observe, however, a discrete proxy with $Y_{ij} = 1$ if the difference in utility from increasing quality of care, compared to not increasing quality of care, is positive ($Y_{ij}^* > 0$), and $Y_{ij} = 0$ otherwise. The discrete indicator used is whether the GP has ordered an HbA1c test for diabetes. $PIP_i$ is a binary variable that takes the value 1 if the GP works in a practice which has joined the PIP, and 0 otherwise. The parameter $\beta_1$ measures the direction and size of the effect of PIP on quality of care provided compared to GPs working in practices who are not in the PIP scheme. The row vector $X_i$ contains a set of observed GP and practice characteristics that vary across GPs. The row vector $X_j$ contains variables measuring the characteristics of the patient at

the encounter. $T_t$ is a vector of dummy variables for each year capturing the time trend in quality of care. All other unobservable factors are relegated to the error term $u_{ij}$. Assuming the error term to be normally distributed $u_{ij} \sim N(0,1)$ yields the probability of ordering an HbA1c test:

$$\Pr(Y_{ij} = 1) = \Phi(\alpha + \beta_1 PIP_i + \gamma' X_i + \gamma' X_j + \delta' T_t), \tag{7}$$

where $\Phi$ denotes the cumulative standard normal distribution function. Parameter estimates are obtained by Maximum Likelihood.

The treatment indicator $PIP_i$ cannot be assumed to be exogenous, since joining the PIP scheme depends on the decision of the general practice in which the GP works, which may also be strongly correlated with the decision to provide quality of care by the individual GP.

Selection into treatment depends on the GP's preference for different practice styles, special interests in the specific disease or differences in the amount and quality of knowledge held by each GP. Some of these may be partly captured by the GP's experience, age and gender. There may also be unobserved differences in practice cost structures, e.g. practice size, that influence the decision to join the PIP, but that also influence the probability of treating diabetes patients. Estimates of $\beta_1$ and other parameters are likely to be biased in Eq. (6).

We address this identification problem with a system of equations estimated with a multivariate probit model[4]. The method accounts for correlations between unobservables. Instead of relying on mere functional form assumptions, exclusion restrictions should be used to improve the identification of each simultaneous equation (Maddala, 1983; Monfardini and Radice, 2008). There are three endogenous decisions that are relevant, all of which

[4]We could also exploit the panel structure of the data (multiple consultations per GP) to control for unobservable GP-invariant effects. Random effects models assuming a linear correlation between the unobserved and time-varying observed variables proposed by Mundlak (1978) cannot be used in this setting. The data we use for the empirical analysis provides time-varying data only for patient characteristics. We assume, however, that unobservable factors influencing both the decision to join the PIP scheme and to provide quality of care are related to practice characteristics. We have tested random effects models, but results do not differ substantially from a pooled version. Fixed effects models can also not be exploited, because we have no variation in the treatment effect within observations.

are likely to have a common set of unobserved factors leading to correlation amongst the error terms. These are the decisions to: provide quality of care in consultations (Eq. (8)), to join the PIP (Eq. (9)), and to be in a larger practice (Eq. (10)).

$$Y_{ij}^{*} = \alpha_1 + \beta_1' X_{1ij} + \beta_{PIP} PIP_{ij} + \beta_S S_{ij} + u_{1ij}, \quad (8)$$

$$PIP_{ij}^{*} = \alpha_2 + \beta_2' X_{2ij} + \pi' I_{2ij} + u_{2ij}, \quad (9)$$

$$S_{ij}^{*} = \alpha_3 + \beta_3' X_{3ij} + \pi' I_{3ij} + u_{3ij}. \quad (10)$$

We allow $X_{1ij} = X_{2ij} = X_{3ij}$, so that each equation includes the same set of control variables. The row vectors $I_{2ij}$ and $I_{3ij}$ capture the instruments used. The error terms $u_{rij}$, for $r \in \{1, 2, 3\}$ are assumed to be distributed as multivariate normal, each with a mean of zero, and variance-covariance matrix V, where V has values of 1 on the leading diagonal and correlations $\rho_{rk} = \rho_{kr}$ as off-diagonal elements. The log-likelihood function for a sample of N independent observations:

$$L = \sum_i w_i \log \Phi_3(\mu_i; \Omega), \quad (11)$$

in which $w_i$ is an optional weight, $\Phi_3(\mu_i; \Omega)$ is a standard trivariate normal cdf, $\mu_i = (K_{i1}\beta_1' X_{1ij}, K_{i2}\beta_2' X_{2ij}, K_{i3}\beta_3' X_{3ij})$ with $K_{i1} = 2Y_{ij}$, and so on. The log-likelihood function depends on the trivariate standard normal distribution function $\Phi_3(.)$ and it is evaluated using the Geweke-Hajivassiliou-Keane (GHK) smooth recursive conditioning simulator. This exploits the fact that a multivariate normal distribution function can be expressed as the product of sequentially conditioned univariate normal distribution functions, which can be easily and accurately evaluated.

According to Wilde (2000), exclusion restrictions for each equation are not needed, if there is at least one exogenous regressor in each equation and if there is sufficient variation in the data. Similar to Balia and Jones (2008) and Schneider and Schneider (2006), we test this hypothesis by re-estimating the three equations without instruments.

Whether or not the instrumental variables are strong and valid are tested

with the help of a Wald test suggested in Rashad and Kaestner (2004, p. 497). The validity of the instruments can only be tested in an over-identified system of equations. The idea is to test a just-identified trivariate model, by including one instrumental variable into the PIP status (GP size) equation and include all the remaining instrumental variables into the HbA1c test equation. Then we test the null hypothesis that all remaining instrumental variables in the HbA1c test equation are jointly statistically insignificant.

Exogeneity of $PIP_{ij}$ and $S_{ij}$ in Eq. (8) can be tested with a t- or a likelihood ratio test on the statistical significance of the correlation coefficient $\rho$.

$$H_0 : \rho_{21} = 0 \ \& \ H_0 : \rho_{32} = 0$$
$$H_1 : \rho_{21} \neq 0 \ \& \ H_1 : \rho_{32} \neq 0.$$

However, a failure to reject $H_0$ should be interpreted with caution. Monfardini and Radice (2008) have shown in a Monte Carlo simulation that tests of $\rho$ tend to display high over-rejection patterns, and in certain cases over-rejection remains serious even in larger sample sizes. The same study has also shown that, even though under correct distributional assumptions, the lack of availability of a valid instrument will make inference on the degree of exogeneity based on the correlation coefficient unreliable and, under misspecified error terms, the model is performing poorly.

# 5 DATA

For the empirical analysis we use data from the *Bettering the Evaluation and Care of Health (BEACH)* study conducted by the Family Medicine Research Centre of the University of Sydney in collaboration with the Australian Institute of Health and Welfare (Britt *et al.*, 2005). Each year since 1998/99, data on 100 consecutive GP encounters from a random sample of 1,000 GPs has been collected. Each encounter contains data on up to four problems treated, and for each problem includes data on drugs prescribed, treatments conducted, referrals written, and pathology and imaging ordered. In addition, a variety of information is provided about the patient, and the GP's

and practice's characteristics.

Since this paper is focussed on diabetes, we use a sub-set of data where at least one of the four problems managed is diabetes. The following variables available in the data-set are used to control for patient, practice, GP and encounter characteristics:

1. Patient characteristics: age, gender, non-English speaking background, Aboriginal or Torres Strait islander (ATSI), concession card status, and whether the patient has type I or type II diabetes;

2. Encounter characteristics: up to four problems managed at the encounter (proxy for co-morbidities), whether the problem is new, and year of encounter;

3. GP characteristics: age (age group 45 years and younger, 46 to 59 years, 60 years and older), gender, number of sessions worked per week (proxy for full- and part-time employment), and country of graduation;

4. Practice characteristics: number of GPs in the practice (solo GP, 2 to 4 GPs, 5 and more GPs (the latter being used as indicator of a large practice)), computer use, whether accredited, practice postcode, practice SEIFA[5] (proxy for socio-economic disadvantage), the geographic remoteness (city, inner region, outer region)[6], and the state in which the practice is located.

In addition, we linked the practice information with information on the Division of General Practice which the practice belongs to using the practice postcode. This enables us to obtain data on the activities and small area characteristics of each Division from the Annual Survey of Divisions of General Practice (Hordacre *et al.*, 2006). Variables include Division activities and characteristics.

[5]SEIFA stands for Socio-Economic Index for Areas and is based on the postcode of the practice.

[6]Remoteness is a measure with discrete indicators, categorized on the basis of the index ARIA. *ARIA* is a continuous measure scaled between 0 and 12 that calculates remoteness as accessibility to service centres based on road distances. The location of the practice is available in five broad categories (1) Major cities; (2) Inner regional areas; (3) Outer regional areas ; (4) Remote areas: (5) Very remote areas. Since we have only a small number of practices located in the remote and very remote areas, we bundle them under outer regional areas.

There are a number of tasks in the guidelines that indicate higher quality of care in diabetes. Although one could use treatment intensity as a dependent variable (i.e. the total number of guideline-recommended tasks) this may not in itself be indicative of quality of care, as a well controlled diabetic receiving high quality care should require fewer tests and investigations.

The main indicator of quality used in the analysis is whether a glycated haemoglobin (HbA1c, HbA1, fructosamine) blood test, which is often abbreviated by glyctest, was conducted during the consultation. Guidelines state that it should be measured at least once every 12 months (Diabetes Australia, 2007). Blood glucose monitoring enables appropriate lifestyle and medication adjustment and monitors long term glycaemic control. The United Kingdom Prospective Diabetes Study (UKPDS) showed reduced incidence and progression of diabetes related complications in subjects with a low HbAlc (UKPDS 33, 1998).

Whether a GP works in a PIP practice is measured using data on accreditation status of the practice and the practice's use of information technology (IT). Such a definition is based on the assumption that all practices who are accredited join the PIP scheme, and furthermore that the use of IT is required to register and recall diabetic patients and claim the diabetes sign-on payment. These practices are also therefore eligible to claim the SIP and OP. Although practices could become accredited from 1998, a question on accreditation was only included in the BEACH GP questionnaire from 2001 onwards, after the PIP scheme was introduced. Our data allow us to compare only the activities of the treatment group with the activities of a 'contemporaneous' control group (practices who were not accredited and didn't use IT at the same point in time).

The activities and characteristics of the Division are used as instrumental variables for identifying the equations in the trivariate probit model. Previous evidence using Division-level panel data has shown that Divisions' activities supported GPs in improving the quality of care in diabetes, through education and information, but they have not influenced SIP claims for diabetes. However, Divisions have had an impact on practice infrastructure and so may have influenced the decisions of practices to become accredited, join the PIP, and also increase the average size of practices (Scott and Coote,

2007). Thus, the activities and characteristics of Divisions are unlikely to be correlated with the error term in the HbA1c equation, but are likely to be partially correlated with PIP status and practice size.

The set of instrumental variables used for identifying the PIP status equation are the *number of staff employed by the Division per practice*, indicator variables that reflect *whether the Division is active in asthma programs*, *number of SIP services provided for cervical screening*, and the *number of practicing GPs*. All of these measures refer to data at the Division level.

The *number of staff employed by the Division per practice* reflects the potential administrative support which a Division can provide to practices to become accredited, to join the PIP scheme, and to claim SIP services. The number of staff will not have any independent influence on the clinical decision-making of a GP, as they are not involved in the actual encounter. *Whether the Division is active in Asthma programs*, where asthma is one of the four disease areas managed under the PIP scheme, reflects the fact that Divisions active in one area of disease management are likely to be active in another area of disease management (e.g diabetes). Although a Division, which is active in diabetes management, may influence indirectly the number of HbA1c tests undertaken by GPs (through a recall/reminder system based on the diabetes patient register system managed at the Division level), activity in asthma cannot have any effect on clinical decision making of the practice in diabetes. The *number of SIP services provided for cervical screening* indicates peer-effects within the Division. The number of SIP services provided in any other disease area should be highly correlated with the number of SIP services provided in diabetes. The more practices are active in the PIP scheme, the more likely it is that other practices get involved too, as word-to-mouth information about the benefits of the program may circulate. How many SIP services are being claimed, however, will have no influence on the proportion of HbA1c tests, as the disease areas are independent of each other. Last, the *number of practicing GPs* within the Division's catchment will have an influence on the number of medical services provided, which ultimately has an influence on the competition among practices within the Division. Competition among practices will have an influence on practice organisation and practice size, but is unlikely to influence independently the

proportion of HbA1c tests conducted during the encounter.

The set of instrumental variables used for identifying the large practice size equation are the *proportion of female GPs*, the *proportion of solo practices*, and the *number of practicing GPs*, which are all measured at the Division level. Practice organisation and style manifest themselves in the number of female GPs employed by the practice. Whereas the gender of the GP may have an influence also on the probability of ordering an HbA1c test (gender-specific difference in treatment styles), the proportion of female GPs active in the Division should not influence the decision to order an HbA1c test, once the gender of the GP has been controlled for. A similar argument holds for the proportion of solo practices within the Division. Last, for reasons mentioned above, the *number of practicing GPs* within the Division should have an effect on the practice size, as the degree of competition within a Division should have an influence on practice organisation but not on clinical decision-making.

# 6 RESULTS

## 6.1 Descriptive analysis

In total, we use a sample from 2000/2001 to 2006 of around 13,000 GP-encounter observations, in which diabetes was one of the four problems managed during the encounter. In about 21 % of consultations of our sample an HbA1c test has been ordered. About 27 % of the sample belong to the treatment group, i.e. those practices identified to participate in the PIP scheme. Summary statistics of all control variables used is provided in Table III in the Appendix.

Trends in the proportion of encounters in which an HbA1c test was ordered are illustrated in Fig. 1. We compare the changes for practices that belong to the treatment group after the Service Incentive Payment (SIP) had been introduced (*PIP 2001-06*), those that did not join the PIP scheme after SIP had been introduced (*No PIP 2001-06*), and all practices before the SIP component was introduced (*All 1998-01*). Crucial reform dates are indicated by light gray, vertical solid lines in the year 1999 (August), when

the PIP scheme and capitation payments were introduced, in the year 2001 (November), when the SIP was introduced, and in the year 2003 (May), when the Outcome Payment (OP) was introduced.

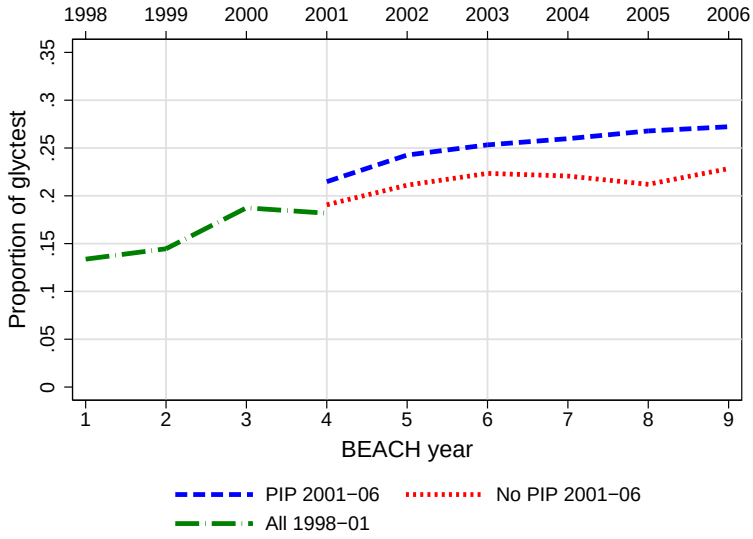


Figure 1. Proportion of consultations in which glyctest was conducted

There is an increasing trend in the proportion of HbA1c tests ordered since 1998, both before the PIP scheme was introduced in 1999 and afterwards. For both control and treatment group, this trend is continued after 2001, but it is slightly stronger for the treatment group (thick, dashed line) after 2003 than for the control group (thin, dotted line)[7].

A similar growth trend can be observed for the proportion of large practices in the sample since 2000. A practice is considered large if it employs five or more GPs. The average growth trend is mainly driven by practices that have joined the PIP program. In fact, the proportion of large practices among the control group has declined steadily since 2002. Fig. 2 illustrates these trends.

[7]The jump in average values in 2001 is due to the fact that in this year we have observations for the group *All 1998-2001* (until 31 October) and for both the treatment and control group (from 1 November onwards).

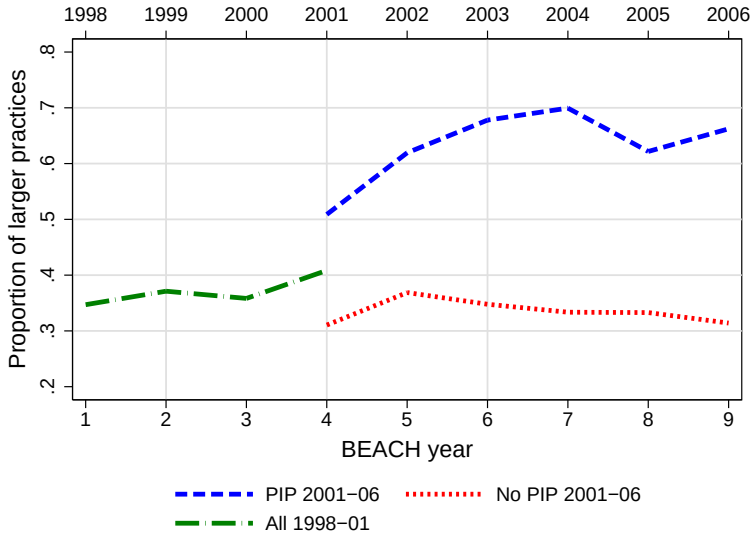


Figure 2. Proportion of large practices ($\geq$ 5 GPs)

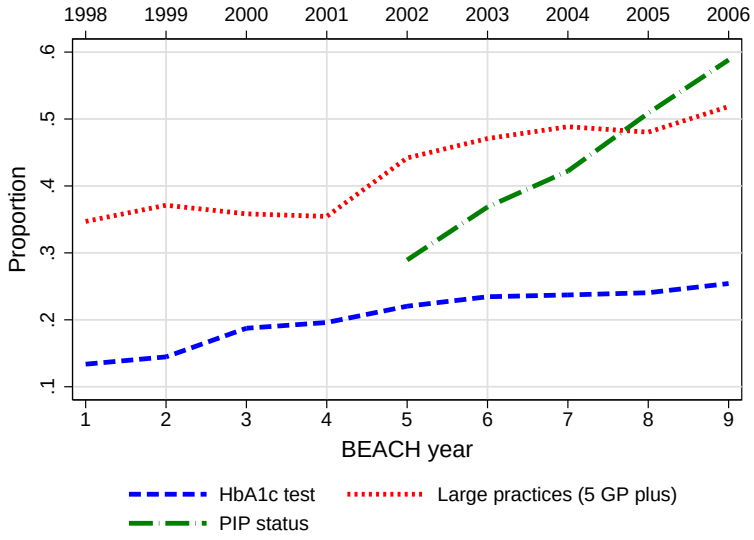


Figure 3. Growth trends between 1998 to 2006

Fig. 3 compares the growth trends of the proportion of large practices, HbA1c tests, and practices that have joined the PIP scheme. It is evident that all three variables show a significant increase over the years. The in-

crease in the uptake of the PIP scheme is steepest: almost 60 % of general practitioners belong to the treatment group in the final year of our sample. This graph illustrates well the possibility that growth trends of the three factors may be determined jointly.

### 6.2 Estimation results

Table II reports the coefficients and marginal effects of joining the PIP scheme on the probability of ordering an HbA1c test. All models control for the same set of independent variables. Model (1) assumes no self-selection into treatment. The effect is not significant, both economically and statistically[8]. Full results are reported in Table IV in the Appendix.

In contrast, the trivariate probit model yields a statistically significant treatment effect (*PIP 2001-2006*) at a 1 % level. In terms of magnitude, the marginal effect is 15 %, suggesting that a GP working in a practice that joined the PIP program is 15 % more likely to order a test than a comparable GP in a practice that has not joined.

Estimation results are obtained using 200 draws to simulate the log-likelihood function[9]. Theoretical results suggest to use at least the square root of the number of observations as the number of draws, which would be in our case at least 110 draws. Cappellari and Jenkins (2006) mention that the coefficients and/or their standard errors will ultimately change with the number of draws (and therefore, as we believe, the marginal effects).

Self-selection occurs in the sense that practices which are located in a Division that delivers an asthma program and whose number of staff per practice is large are more likely to join the PIP scheme. This latter observation suggests that it is the work and effort of the Division staff in supporting

[8]We have tested for the possibility that the treatment effect is conditional on the timing, the practice size or the remoteness of the practice. None of the three hypotheses, however, can be confirmed. The treatment effect is statistically insignificant for all outlined interaction effects.

[9]The default is 5 draws in the STATA program mvprobit written by Lorenzo Cappellari and Stephen Jenkins. The authors point out that it is the responsibility of the researcher to choose the correct number of draws. Under standard conditions, the MSL estimator is consistent as the number of observations and the number of draws tend to infinity and is asymptotically equivalent to the true maximum likelihood estimator as the ratio of the square root of the sample size to the number of draws tends to zero. Thus, other things equal, the more draws, the better.

Table II. Results Pooled and trivariate probit

| | Pooled | Trivariate probit | | |
|---|---|---|---|---|
| | probit | PIP status | HbA1c test | GP size |
| *Treatment: coefficient* | | | | |
| PIP 2001-2006 | 0.023 | | 0.515 | |
| | (0.033) | | (0.166)*** | |
| *Treatment: marginal effect* | | | | |
| PIP 2001-2006 | 0.007 | | 0.151 | |
| | (0.010) | | (0.049)*** | |
| *Instruments used in PIP status eq.* | | | | |
| Division active in asthma program | | 0.252 | | |
| | | (0.155) | | |
| Num of staff in Division per practice | | 0.572 | | |
| | | (0.245)** | | |
| Number of SIP services (cerv screen) | | 0.002 | | |
| | | (0.001)*** | | |
| *Instruments used in GP size eq.* | | | | |
| Proportion of solo GPs in Division | | | | -0.852 |
| | | | | (0.264)*** |
| Proportion of female GPs in Division | | | | 0.713 |
| | | | | (0.398)* |
| *Instrument used in PIP status & GP size eq.* | | | | |
| Number of GPs active in Division | | -0.004 | | -0.001 |
| | | (0.001)*** | | (0.001)*** |
| *Correlation of unobservables between equation* | | | | |
| $\rho_{21}$ | | | -0.258 | |
| | | | (0.109)** | |
| $\rho_{31}$ | | | -0.550 | |
| | | | (0.170)*** | |
| $\rho_{32}$ | | | -0.032 | |
| | | | (0.135) | |

The marginal effect for the treatment effect is computed with the *mvpred* command and the delta method. All equations include the same set of control variables as used in the full pooled probit specification. * 10 %, ** 5 %, *** 1 % significance level.

practices to SIP claim that influences the practice decision to join the PIP scheme, rather than pure capacity constraint arguments.

In addition, practices are also more likely to joint the PIP scheme in a Division in which the number of SIP services provided in cervical screening are high. This suggests that practices are more likely to join the scheme if its application is more prevalent and perceived as effective, as positive information about the nature of the scheme influences expected costs and benefits. Practices are also less likely to join the PIP if the number of practitioners active in the Division is large.

The instruments used in the PIP status equation are tested to be both valid and strong. All possible permutations of testing for the validity of the instruments with the help of a Wald test indicate that we fail to reject the

null hypothesis that three of the four instruments are jointly insignificant in the HbA1c test equation, while leaving the remaining instrument in the PIP status equation[10].

Larger practices (5 and more GPs employed) are more likely to be located in Divisions in which the total number of active GPs and the proportion of solo GPs are low. They are more likely to be located in Divisions in which the proportion of female GPs is high. All instruments have no independent statistically significant influence on the probability to order an HbA1c test[11]. These results are summarised in Table V in the Appendix.

The error terms are negatively correlated between the PIP status and the HbA1c equation and between the PIP status and the GP size equation. Both correlation coefficients are statistical significant at the 5 % and 1 % level, rejecting the hypothesis of exogeneity of the two variables PIP status and GP size in the HbA1c test equation[12]. The negative correlation suggests that unobservables which are negatively influencing the probability of joining the PIP scheme are positively influencing the probability of conducting the HbA1c test, and thus, increasing quality of care.

The unobservables influencing the probability of a large practice size and the probability to order an HbA1c test are not correlated. The treatment effect, resulting from a bivariate probit specification that accounts only for the correlation of unobservables between the PIP status and the HbA1c test equation and which uses the exact log-likelihood, is 20 % and statistically significant at the 1 % level (results omitted).

Another intriguing result from our analysis is that once we drop the instruments from the trivariate (bivariate) probit, an approach suggested by Wilde (2000), we find no statistically significant effect of the PIP reform on quality of care. The magnitude of the marginal effect (standard error) is .098 (.069) and the p-value = 0.156. In addition, the correlation between the error-term of all three (two) equations is small and statistically insignificant

[10]For all possible permutations the p-value lies between 0.330 or 0.808.

[11]For all possible permutations the p-value lies between 0.542 and 0.642.

[12]Statistical significance of the correlation in the trivariate probit model depends on the choice of the instruments chosen. In a bivariate probit model however the correlation of unobserved factors of the PIP status and the HbA1c equation is highly statistically significant and independent of the choice of the instruments.

(all p-values are greater than 0.200). One explanation could be that our data does not provide enough variation to identify the equations exclusively on the basis of functional form assumptions.

# 7 DISCUSSION

Pay-for-performance is being increasingly used to change clinical behaviour. The study finds that the incentive reform had a positive effect on quality of care in diabetes management. The magnitude of the effect lies between a 15 % and 20 % difference in the probability of ordering an HbA1c test since the reform was introduced.

Practices self-selected into the treatment due to the fact that not all practices faced the same conditions of joining in and claiming. The activities of the Divisions of General Practice played a major role in the uptake and therefore on the effects of the reform on quality of care.

The strong influence of Divisions' activities highlights the importance of a supporting regional infrastructure when a complex payment system is introduced to remunerate GPs. This confirms other evidence of the effect of Divisions on diabetes care provided by GPs. For example, before the introduction of the new incentive payment, some Divisions of General Practice had already established programs with some elements of systematic care, including Division-based registers, recall systems and GP feedback (Harris *et al.*, 2002). Practices that were linked to these shared-care diabetes registers reviewed patients more often and were generally more likely to conduct tests recommended in the evidence-based guidelines than GPs not using registers (DeDomenico *et al.*, 2005).

Difficulties in uptake have been discussed in the general literature on the implementation of highly complex payment systems (Robinson *et al.*, 2004). For the PIP, apart from an initial enthusiastic uptake when the PIP scheme was introduced, the response has levelled off in later years (Harris *et al.*, 2004). Proudfoot *et al.* (2007) found that the implementation of the new payment systems for chronic disease management was perceived as the most difficult task out of all tasks faced by GPs. In the case of claiming services that completed a cycle of care, for instance, it was mainly

the extent of administrative burden that posed problems. This included the lack of support staff that would be needed to make the claim which is more problematic for solo and/or rural practices that cannot hire extra staff (DeDomenico *et al.*, 2005). Furthermore, some GPs, even though they provided the service, did not claim it or they claimed it via another care plan. Many GPs seem to be not aware of which services they can claim. A government review of 'red tape' in 2004 led to a number of changes in PIP (and other schemes) from 2005/6 that made it less costly for GPs to claim a range of fees.

In addition to a lack of organisational capacity in some practices, the level of remuneration may not have been sufficient to encourage more practices to join the PIP and become accredited, and once in the PIP, the incentive payments may have been too low. At around 10% of total GP income, this may have been insufficient for some practices, especially if they could increase earnings in other less costly ways.

There are a number of issues about the modelling that are important to emphasise. Since our data were not a panel nor was it possible to observe treatment status before the reform, this study could not directly examine changes in the behaviour of the same GPs over time in a difference-in-difference framework. Our implicit comparisons of different GPs in each year rely on the fact that each cross-section is a random sample of all GPs. More importantly, the trivariate probit model allowed us to control for the correlation of unobservable factors between the decision to join the PIP, the decision to undertake an HbA1c test, and the decision to be a large practice. One can speculate on what these unobserved factors might be, such as intrinsic motivation or unobserved marginal costs, but this can only be confirmed by more detailed data which does not yet exist. The model also allowed identification of the effect of PIP on the probability of ordering an HbA1c test through the use of strong and valid instrumental variables.

Due to sample size considerations, it was not possible to examine the effect of PIP over time. Usually one expects an initial strong effect and then a levelling off of the effect of the incentive. The marginal effect reported here is therefore an average between 2001/2 and 2006.

The results imply that the PIP is having a moderate effect on quality of

care in diabetes and should therefore be continued as long as policy makers believe that these additional benefits are worth the costs of the scheme. Further research is evaluating the effects of the PIP on other incentivised areas of care (asthma, mental health and cervical screening) which would give a more complete picture of the benefits of the scheme.

## References


Australian Division of General Practice. 2005. Dynamic Divisions. *Corporate Publication* **3**.

Balia S, Jones A. 2008. Mortality, lifestyle, and socio-economic status. *Journal of Health Economics* **27**(1): 1–26.

Beilby J. 2003. *Attendance Item Restructure Working Group. Final Report.* Royal Australian College of General Practitioners.

Britt H, Miller G, Knox S, *et al.* 2005. General practice activity in australia 2004-2005. *General Practice Series AIWH* (18).

Cappellari L, Jenkins SP. 2006. Calculation of multivariate normal probabilities by simulation, with applications to maximum simulated likelihood estimation. *ISER Working Paper* **2006-16**.

Chalkley M, Tilley C. 2006. Treatment intensity and provider remuneration: dentists in the British National Health Service. *Health Economics* **15**: 933–946.

DeDomenico M, Yu C, Taggart J, *et al.* 2005. Barriers & enablers to the uptake of NIDP diabetes service incentive payments in general practice. *Centre for General Practice Integration Studies, School of Public Health & Community Medicine, UNSW* .

Diabetes Australia. 2007. Diabetes management in general practice. *Annual Report* **2007/08**.

Doran T, Fullwood C, Gravelle H, *et al.* 2006. Pay-for-performance programs in family practices in the United Kingdom. *New England Journal of Medicine* **355**: 375–384.

Francois P, Vlassopoulus M. 2007. Pro-social motivation and delivery of social services. *July 5* **mimeo**.

Frey BS, Jegen R. 2001. Motivation crowding theory: A survey of empirical evidence. *Journal of Economic Surveys* **15**: 589–611.

Gaynor M, Rebitzer JB, Taylor LJ. 2001. Incentives in HMOs. *NBER Working Paper* **8522**.

Gravelle H, Sutton M, Ma A. 2008. Doctor behaviour under a pay for performance contract: Further evidence from the quality and outcomes framework. *CHE Research Paper* **34**.

Harris M, Georgiou A, Powell-Davies P. 2004. Diabetes service incentive payments by division 2001/2002 to 2003/2004. *Report of National Division Diabetes Program Division Diabetes & CVD Quality Imrovement Program Centre for GP Integration Studies, School of Public Health & Community Medicine, UNSW* **November**.

Harris M, Ruscoe DPW, Infante F, *et al.* 2002. Quality of care provided by general practitioners using or not using Division-based diabetes registers. *Medical Journal of Australia* **177**: 250–252.

Hordacre A, Kalucy MKE, Moretti C. 2006. *Making the Connections. Report of the 2004-2005 Annual Survey of Divisions of General Practice*. Primary Health Care Research & Information Service, Department of General Practice, Flinders University, Adelaide.

Lippi Bruni M, Nobilio L, Ugolini C. 2007. Economic incentives in general practice: The impact of pay for participation programs on diabetes care. *University of Bologna Economics Working Paper* **607**.

Maddala G. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, Cambridge.

Monfardini C, Radice R. 2008. Testing exogeneity in the bivariate probit model: A Monte Carlo study. *Oxford Bulletin of Economics and Statistics* **70**(2): 0305–9049.

Mooney G, Ryan M. 1993. Agency in health care: getting beyond first principles. *Journal of Health Economics* **12**: 125–135.

Mundlak Y. 1978. On the pooling of time series and cross section data. *Econometrica* **46**(1): 69–85.

Prendergast C. 1999. The provision of incentives in firms. *Journal of Economic Literature* **37**: 7–63.

Proudfoot J, Infante F, Holton C, *et al.* 2007. Organisational capacity and chronic disease care. An Australian general practice perspective. *Australian Family Physician* **36**(3): 286–288.

Rashad I, Kaestner R. 2004. Teenage sex, drugs and alcohol use: problems identifying the cause of risky behaviour. *Journal of Health Economics* **23**: 493–503.

Robinson JC, Shortell SM, Li R, *et al.* 2004. The alignement and blending of payment incentives within physician organizations. *Health Services Research* **39**(5): 1589–1606.

Roland M. 2004. Linking physicians' pay to the quality of care-a major experiment in the United Kingdom. *The New England Journal of Medicine* **351**(14): 1448–1454.

Schneider U, Schneider B. 2006. The effects of education and working hours on health: A multivariate probit approach. *University of Bayreuth Discussion Paper* **05-06**.

Scott A, Coote W. 2007. Whither divisions of general practice? An empirical and policy analysis of the impact of Divisions within the Australian health care system. *Medical Journal of Australia* **187**(2): 95–99.

Scott A, Shiell AM. 1997. Do fee descriptors influence treatment choices in general practice? A multilevel discrete choice model. *Journal of Health Economics* **16**(3): 323–342.

Shen J, Andersen R, Brook R, *et al.* 2004. The effects of payment method on clinical decision-making. *Medical Care* **42**(3): 297–302.

Simoens S, Giuffrida A. 2004. The impact of physician payment methods on raising efficiency of the healthcare system. *Applied Health Economics Health Policy* **3**(1): 39–46.

UKPDS 33 UKPDS. 1998. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes. *Lancet* **352**(9131): 837–853.

Wilde J. 2000. Identification of multiple equation probit models with endogenous dummy regressors. *Economics Letters* **69**(3): 309–312.

Table III. Summary statistics

| Variable | Mean | Std. Dev. | Min. | Max. | N |
|---|---|---|---|---|---|
| PIP 2001-06 | 0.271 | 0.445 | 0 | 1 | 24172 |
| Practice accredited | 0.676 | 0.468 | 0 | 1 | 21248 |
| Full computer use at practice | 0.319 | 0.466 | 0 | 1 | 24172 |
| Percentage of HbA1c tests | 0.208 | 0.406 | 0 | 1 | 24161 |
| Age of GP in years | 51.607 | 10.805 | 25 | 86 | 24066 |
| GP age less than 46 years | 0.307 | 0.461 | 0 | 1 | 24172 |
| GP age 46 to 59 years | 0.464 | 0.499 | 0 | 1 | 24172 |
| GP age 60 years & more | 0.229 | 0.42 | 0 | 1 | 24172 |
| GP female | 0.273 | 0.446 | 0 | 1 | 24172 |
| Practice SEIFA | 5.922 | 3.218 | 1 | 11 | 23950 |
| Graduated in Australia | 0.727 | 0.445 | 0 | 1 | 24172 |
| Graduated in OECD country | 0.115 | 0.319 | 0 | 1 | 24155 |
| Graduated in non OECD country | 0.157 | 0.364 | 0 | 1 | 24155 |
| Fellow of RACGP | 0.332 | 0.471 | 0 | 1 | 24054 |
| Inner region | 0.204 | 0.403 | 0 | 1 | 24155 |
| Large cities | 0.665 | 0.472 | 0 | 1 | 24155 |
| Outer region | 0.131 | 0.337 | 0 | 1 | 24155 |
| GP: solo | 0.173 | 0.378 | 0 | 1 | 23962 |
| GP: 2 to 4 | 0.394 | 0.489 | 0 | 1 | 23962 |
| GP: 5 and more | 0.434 | 0.496 | 0 | 1 | 23962 |
| NSW | 0.371 | 0.483 | 0 | 1 | 24155 |
| ACT | 0.012 | 0.108 | 0 | 1 | 24155 |
| NT | 0.013 | 0.113 | 0 | 1 | 24155 |
| QLD | 0.167 | 0.373 | 0 | 1 | 24155 |
| SA | 0.086 | 0.281 | 0 | 1 | 24155 |
| TAS | 0.026 | 0.16 | 0 | 1 | 24155 |
| VIC | 0.241 | 0.428 | 0 | 1 | 24155 |
| WA | 0.083 | 0.276 | 0 | 1 | 24155 |
| Year 2002/2003 | 0.112 | 0.315 | 0 | 1 | 24172 |
| Year 2003/2004 | 0.12 | 0.325 | 0 | 1 | 24172 |
| Year 2004/2005 | 0.111 | 0.314 | 0 | 1 | 24172 |
| Year 2005/2006 | 0.129 | 0.335 | 0 | 1 | 24172 |
| Year 2006/2007 | 0.12 | 0.325 | 0 | 1 | 24172 |
| Age 1-14 | 0.004 | 0.065 | 0 | 1 | 24172 |
| Age 15-24 | 0.01 | 0.099 | 0 | 1 | 24172 |
| Age 25-44 | 0.081 | 0.274 | 0 | 1 | 24172 |
| Age 45-64 | 0.384 | 0.486 | 0 | 1 | 24172 |
| Age 65-74 | 0.28 | 0.449 | 0 | 1 | 24172 |
| Age 75 plus | 0.24 | 0.427 | 0 | 1 | 24172 |
| Patient female | 0.491 | 0.5 | 0 | 1 | 24172 |
| ATSI | 0.031 | 0.173 | 0 | 1 | 24172 |
| Non-English speaking | 0.151 | 0.358 | 0 | 1 | 23265 |
| Diabetes type 1 | 0.088 | 0.283 | 0 | 1 | 24172 |
| Diabetes type 2 | 0.913 | 0.282 | 0 | 1 | 24172 |
| Health care card | 0.73 | 0.444 | 0 | 1 | 21107 |
| Number of other problems | 1.307 | 0.961 | 0 | 3 | 24172 |
| Division asthma program | 0.884 | 0.175 | 0 | 1 | 24152 |
| Number of SIP services provided | 189.61 | 135.667 | 13 | 671.778 | 24099 |
| Ratio Division staff per practice | 0.181 | 0.148 | 0.036 | 1.162 | 24152 |
| Number of GPs active in Division | 98.722 | 67.091 | 6.571 | 293.2 | 24152 |
| Percentage of female GPs within Division | 0.314 | 0.099 | 0.031 | 0.651 | 24152 |
| Percentage of solo practices within Division | 0.398 | 0.135 | 0.05 | 0.733 | 24152 |

Table IV. Full results: pooled and trivariate probit

| | Pooled probit | Trivariate probit | | |
|---|---|---|---|---|
| | HbA1c | PIP status | HbA1c | GP size |
| GP age less than 46 years | 0.111 | 0.061 | 0.071 | 0.342 |
| | (0.035)*** | (0.077) | (0.045) | (0.063)*** |
| GP age 60 years & more | -0.194 | -0.280 | -0.122 | -0.312 |
| | (0.043)*** | (0.091)*** | (0.053)** | (0.072)*** |
| Sessions worked p week | 0.030 | -0.119 | 0.047 | 0.001 |
| | (0.038) | (0.062)* | (0.040) | (0.065) |
| GP female | 0.009 | 0.037 | 0.008 | -0.081 |
| | (0.008) | (0.014)** | (0.011) | (0.014)*** |
| PRSEIFA | 0.152 | 0.022 | 0.000 | 0.014 |
| | (0.039)*** | (0.010)** | (0.006) | (0.010) |
| Inner region | 0.194 | 0.419 | 0.042 | -0.006 |
| | (0.050)*** | (0.088)*** | (0.054) | (0.080) |
| Outer region | 0.004 | 0.456 | 0.100 | -0.447 |
| | (0.005) | (0.118)*** | (0.074) | (0.107)*** |
| Graduated in OECD country | -0.053 | 0.027 | -0.048 | -0.068 |
| | (0.045) | (0.084) | (0.045) | (0.085) |
| Graduated in NON OECD country | -0.185 | 0.033 | -0.165 | -0.293 |
| | (0.050)*** | (0.085) | (0.056)*** | (0.080)*** |
| Large practice | 0.017 | 1.550 | -0.052 | |
| | (0.032) | (0.234)*** | (0.240) | |
| ACT | 0.129 | -0.363 | 0.155 | 0.114 |
| | (0.124) | (0.206)* | (0.137) | (0.207) |
| NT | 0.171 | -0.332 | 0.221 | 0.036 |
| | (0.172) | (0.328) | (0.177) | (0.271) |
| QLD | 0.035 | 0.304 | -0.031 | 0.005 |
| | (0.042) | (0.082)*** | (0.048) | (0.087) |
| SA | -0.042 | 0.280 | -0.117 | 0.348 |
| | (0.056) | (0.121)** | (0.068)* | (0.111)*** |
| TAS | -0.266 | 0.213 | -0.301 | 0.134 |
| | (0.136)* | (0.162) | (0.137)** | (0.158) |
| VIC | -0.034 | 0.074 | -0.089 | 0.160 |
| | (0.043) | (0.085) | (0.051)* | (0.085)* |
| WA | -0.007 | -0.085 | -0.038 | 0.112 |
| | (0.057) | (0.100) | (0.061) | (0.111) |
| Year 2003/2004 | 0.059 | 0.153 | 0.026 | 0.123 |
| | (0.050) | (0.088)* | (0.053) | (0.086) |
| Year 2004/2005 | 0.059 | 0.283 | 0.000 | 0.176 |
| | (0.048) | (0.094)*** | (0.053) | (0.087)** |
| Year 2005/2006 | 0.090 | 0.550 | -0.012 | 0.145 |
| | (0.048)* | (0.099)*** | (0.060) | (0.087)* |
| Year 2006 | 0.113 | 0.685 | -0.022 | 0.197 |
| | (0.049)** | (0.106)*** | (0.067) | (0.085)** |
| Age-group 1 to 14 | -0.288 | 0.064 | -0.291 | 0.130 |
| | (0.261) | (0.239) | (0.260) | (0.205) |
| Age-group 15 to 24 | 0.147 | -0.221 | 0.168 | -0.051 |
| | (0.125) | (0.140) | (0.125) | (0.128) |
| Age-group 25 to 44 | 0.094 | -0.199 | 0.118 | 0.043 |
| | (0.051)* | (0.054)*** | (0.052)** | (0.054) |
| Age-group 45 to 64 | 0.043 | -0.082 | 0.054 | 0.001 |
| | (0.031) | (0.033)** | (0.031)* | (0.033) |
| Age-group 75 plus | -0.084 | -0.051 | -0.075 | 0.001 |
| | (0.034)** | (0.036) | (0.034)** | (0.034) |
| Patient female | -0.016 | -0.009 | -0.016 | 0.009 |
| | (0.025) | (0.026) | (0.025) | (0.025) |
| ATSI | -0.166 | -0.382 | -0.089 | -0.079 |
| | (0.073)** | (0.156)** | (0.077) | (0.162) |
| Non-English background | -0.065 | -0.117 | -0.021 | -0.227 |
| | (0.043) | (0.068)* | (0.046) | (0.062)*** |
| Type 1 Diabetes | -0.343 | -0.032 | -0.337 | 0.059 |
| | (0.050)*** | (0.049) | (0.051)*** | (0.050) |
| Health care c | -0.127 | -0.065 | -0.113 | -0.012 |
| | (0.029)*** | (0.034)* | (0.029)*** | (0.034) |
| Other problems | 0.046 | 0.033 | 0.040 | -0.004 |
| | (0.014)*** | (0.017)* | (0.014)*** | (0.017) |
| Constant | -0.841 | -2.031 | -0.872 | 0.685 |
| | (0.102)*** | (0.244)*** | (0.171)*** | (0.261)*** |

Table IV reports the coefficients of all control variables used in each equation of the trivariate probit model. * 10 %, ** 5 % , *** 1 % significance level.

Table V. Validity test of instruments

| Eqs. PIP status & HbA1c test | | | | Eqs. GP size & HbA1c test | | |
|---|---|---|---|---|---|---|
| Single instrument in PIP eq.: | | | | Single instrument in GP size eq.: | | |
| asthma | SIPservice | StaffPrac | nGPs | nGPs | soloGPs | femGPs |
| *p-value of joint $H_0$: all other instruments are jointly insign in HbA1c eq.* | | | | | | |
| .331 | .330 | .808 | .373 | .542 | .642 | .584 |
| *p-value of single instrument in PIP/GP size eq.* | | | | | | |
| .052 | .626 | .001 | .015 | .001 | .000 | .012 |
| *p-value of treatment effect* | | | | | | |
| .046 | .068 | .003 | .056 | .411 | .400 | .409 |

Table V reports the p-values for testing the hypothesis that the three (two) of the four (three) instruments used in the PIP status (GP size) equation are jointly not statistically significant in the HbA1c equation, while the remaining instrument is used to identify the PIP status (GP size) equation. THe p-value of the test that the single remaining instrument is statistically not significant in the PIP status (GP size) equation is reported in the second line. The p-value of the hypothesis that the treatment effect is not statistically significant under these scenarios is reported in the third line.