FINANCIAL INCENTIVES AND THE QUALITY OF PRIMARY CARE IN AUSTRALIA

Prof. Anthony Scott
Dr. Lucio Naccarella
Dr. John Furler
Prof. Doris Young
Dr. Peter Sivey
Mr. Driss Ait Ouakrim
Ms. Lisa Willenberg
ACKNOWLEDGMENT

The research reported in this paper is a project of the Australian Primary Health Care Research Institute, which is supported by a grant from the Australian Government Department of Health and Ageing under the Primary Health Care Research, Evaluation and Development Strategy. The information and opinions contained in it do not necessarily reflect the views or policies of the Australian Government Department of Health and Ageing. Thanks go to the international reference group and policy advisory group for input, and members of the Cochrane Effective Practice and Organisation of Care (EPOC) group for comments and assistance with the Cochrane Review.
INTRODUCTION

The aim of this review is to examine the role of explicit financial incentives to improve the quality of primary care in Australia. Primary care provides integrated, easy to access, health care services by clinicians who are accountable for addressing a large majority of personal health care needs, developing a sustained and continuous relationship with patients, and practicing in the context of family and community (Vanselow et al. 1995).

The use of blended payment schemes in primary care, including the use of financial incentives to directly reward ‘performance’ and ‘quality’ is increasing in a number of countries. There are hundreds of examples in the US and the Quality and Outcomes Framework (QoF) for GPs in the UK is an example of a major system-wide reform (Christianson et al. 2008, Institute of Medicine 2007, Rosenthal and Frank 2006). In Australia, the Practice Incentive Program (PIP) was introduced in 1999 to improve the quality of care provided in asthma, diabetes, mental health and cervical screening. Despite the popularity of these schemes, there is currently little rigorous evidence of their success in improving quality, or of whether such an approach is cost-effective relative to other ways to improve the quality of care (Gosden et al. 2000, Scott 2007).

The role and impact of such incentives may depend on a number of factors, including the amount of the payment, how the payment is made (eg salary, fee-for-service, performance bonus or targeted at individuals or teams), and on the factors that motivate primary care teams and their members that include how important income is relative to other sources of intrinsic and extrinsic motivation they may have (e.g. professional autonomy). These sources of motivation can influence the success or failure of these schemes and may vary across doctors in different settings and doctors with different preferences and practice styles.

The success of financial incentives is also dependent on the costs of participating in these schemes, which are likely to vary across different primary care teams including differences in administrative costs borne by different primary care teams of different sizes, the additional effort required by the primary care team to increase quality of care to meet the incentive requirements, and support provided to help reduce these costs by regional primary care organisations or clinical networks (De Domenico M et al. 2005, Scott 2007). Both theory and empirical evidence suggest that, in some circumstances, poorly designed incentives may not work at all, can lead to lower levels of quality, or unintended and distorted behavioural effects (Casalino et al. 2007, Christianson et al. 2006, Conrad and Perry 2009, Fisher 2006, Institute of Medicine 2007, Mannion and Davies 2008, Prendergast 1999, Rosenthal et al. 2004, Nicholson et al. 2008). For example, financial incentives applied to one disease area may ‘work’, but at the cost of GPs spending less time in other disease areas or with other types of patient, such that the overall net impact on quality of care and costs is difficult to determine. Other unintended effects may include either the selection of healthy ‘eligible’ patients by physicians, or the selection of high performing physicians into the incentives schemes, such that it is these selection mechanisms that drive observed success, rather than the effect of the scheme itself.

The aim of this review is to summarise the evidence on the use of financial incentives to improve quality in primary care, and discuss the options for such an incentive scheme in Australia. After defining financial incentives and a broad conceptual framework, we conduct both a narrative review of the literature and a Cochrane systematic review. The narrative review examines why such incentives may or may not work by examining experience in a number of countries and contexts. The Cochrane review summarises the most rigorous evidence on the effectiveness of such incentives in changing the quality of care provided by primary care physicians.
CONCEPTUAL FRAMEWORK

The aim of the conceptual framework is to outline the main issues that influence the design, use, and impact of financial incentives targeted at health care providers. In turn, these issues will form the structure of data extraction for both the narrative and Cochrane review.

The review is concerned with examining the effect of financial incentives on the quality of primary health care. This is concerned with monetary transfers targeted at primary care health professionals or teams, where the monetary transfer influences the personal income or remuneration of the health professional. The 'incentive' is defined as the behavioural response to the monetary transfer, and is not necessarily fixed or embodied in the monetary transfer itself. The effect of the monetary transfer depends on the way the funding change is implemented and communicated, and on the behavioural response of those receiving the monetary transfer, which in turn depends on their preferences, costs of changing behaviour, and context (Giacomini et al. 1996).

There is a large, mainly theoretical, literature in economics about the use of incentives in organisations, and a large empirical literature in health economics (eg. Prendergast 1999, Conrad and Perry 2009). Other behavioural social sciences, mainly psychology, also have a role to play in examining the effects of incentives. These theories provide a rich set of predictions about how individuals respond to financial incentives.

There are number of factors that can influence behaviour, which can in turn influence the cost and quality of care provided.¹

CONTEXT

There are a number of issues related to the context in which the monetary reward is introduced. How health systems have evolved are important determinants of the nature and type of health care reform that is possible. The political climate for change, the professional climate for change, as well as the structure of the health system to support change (for example the existence of regional primary care organisations), are key factors. The system may already have a culture of performance management from previous reforms which may make a new pay for performance scheme less controversial and more acceptable. The system may already have experienced quality improvement initiatives, such as education, audit and feedback on performance. A scheme which links monetary transfers to the level, or changes in the level, of quality of care requires that quality of care can be measured and monitored. This requires substantial IT infrastructure and reporting requirements to third party funders. This is a necessary condition for any system of financial incentives linked to quality. Other types of intervention may therefore have already occurred or may occur at the same time of the monetary transfer, and it is important to distinguish their independent and combined effects.

How the policy is developed, implemented and communicated is a key factor in its effectiveness (Giacomini et al. 1996). The clarity of the policy influences how it is interpreted by the targets of the policy, and therefore the extent and direction of behaviour change. Vague or unclear

¹ Changes in payment may also influence quality and cost by influencing recruitment and retention, ie influencing the mix of providers, and their self-selection into the payment scheme. Changes in the mix of providers may then impact on quality, rather than the incentive having a direct effect on quality. This leads to selection bias when estimating the effectiveness of incentives, especially with voluntary incentive schemes (Prendergast 1999, Nicholson et al. 2008).
signals may lead to unintended consequences or interpretations not expected by the policy maker.

Christianson et al (2006) examines the role of organisational culture in the successful implementation of change. Klein and Sorra (1996) discuss implementation climate in terms of employees having shared perceptions of organisational goals and rewards, and the extent to which the innovation will meet these goals and values (value fit).

AMOUNT OF PAYMENT

The amount of the payment influences the size of the behavioural response. The relationship between the amount of the payment and the size of the behaviour change is the ‘effect size’ or ‘elasticity’. Economic theory suggests that monetary transfers change the ratio of the costs and benefits of the conditional behaviours. The monetary transfer would lead to the primary care physician improving the quality of care only if the additional (marginal) benefits are greater than the additional (marginal) costs of undertaking the behaviour, compared to current practice. The larger the cost-benefit ratio, the more likely the individual is to change their behaviour. Higher payments (with the same cost) lead to larger changes in behaviour.2 Lower costs (with a fixed payment) lead to larger changes in behaviour. The cost-benefit ratio, and therefore their behavioural response, will vary across primary care professionals and their teams because of variations in the cost of providing services, and variations in preferences and motivation.

The amount of payment may:

- be fixed in advance and subject to periodic negotiation (e.g. in a fee schedule or trade union bargaining agreement);
- be flexible such that providers have complete (in a market setting) or partial (e.g. fee/price controls) discretion as to the amount of money they can charge for their services;
- be reduced or withheld if behaviour does not comply with what is required (i.e. a financial penalty);
- vary depending on administrative rules based on the characteristics of the provider (e.g. whether accredited or qualified to a certain standard or seniority, or other eligibility requirements) or depending the characteristics of patients seen (e.g. ‘risk adjustment’ where more complex/costly services or patients receive higher payments).

The total amount of payments made by a third party payer (government or insurer) to a provider can be capped to influence the amount of risk sharing and therefore influence total expenditures. Risk sharing is an important part of any funding scheme to help encourage efficiency and cost containment. This influences the strength of the behavioural response. If payments are made in advance (prospective payment and capitation payment) this provides a fixed total payment pool for each provider. If total payments are greater than this amount, then the provider is at risk for this additional expenditure. If payments are less, then the provider can keep the savings. Where there is no overall cap on the amounts that can be paid,

2 Although this may not always be the case if the effect of the payment on income (income effect) is larger than the effect of the payment on the behaviour (substitution effect). In this case, the behaviour may be less likely to occur.
then the third party payer bears all of the financial risk, and the provider has no incentive to be cost-conscious. Once the cap is reached, either no further payments are made or the amount of the unit payment is reduced.

METHOD OF PAYMENT

Payments can be made in exchange for the following provider behaviours.

- Working for a specified time period (e.g. salary or sessional payment). The levels on a salary scale are usually based on experience/seniority. If accompanied by a subjective performance evaluation (e.g. annual appraisal), then achievement of levels of subjective performance can be built into the process of promotion, therefore providing incentives to increase performance over time. Salary payment usually involves direct employment of the primary care physician, whereas sessional payment or other fixed payments can be made to primary care physicians who are self-employed.

- Providing specific services/treatments/episodes/cases (fee-for-service).

- Providing care for a patient within a defined enrolled population (capitation).

- Providing a pre-specified level, or increases in the level, of performance (performance-based pay or target payments). Performance can be defined in a number of ways, but is usually linked to providing agreed ‘appropriate’ treatments in specific disease areas, that have been shown to improve health outcomes based on evidence-based guidelines. Payment for providing a level of performance that is high relative to the performance ranking of other providers is known as tournament-based pay, and is based on providers competing for a bonus or one-off annual payment.

Payments may be linear or non-linear:

- linear, so the same payment is made for each additional unit of service provided;

- non-linear, such that payment is conditional on reaching a threshold or target, or a series of thresholds, or that the amount of the payment changes with each additional service provided. This may be appropriate where the additional (marginal) costs increase with each successive service provided.

THE COSTS OF BEHAVIOUR CHANGE.

The additional costs of undertaking the behaviour change include the additional time and effort of the primary care professional/team, the financial costs related to improving quality (e.g. the purchase of equipment, IT infrastructure such as a recall and reminder systems, additional time spent with patients), and administrative costs of claiming the payment. These costs may depend on the size of the practice/team. High fixed infrastructure costs lead to high average and marginal costs per patient if the number of patients seen is small and/or the practice is small, due to diseconomies of small scale and scope. For example, a small practice in a rural area (with few patients) may find it more difficult to change behaviour compared to a larger practice in an urban area because it cannot spread its fixed costs over a high enough number of patients. This leads to high average and marginal costs per patient, which may be higher than the size of monetary transfer. The characteristics of primary care professionals and their teams
and practices are therefore important indicators of the costs of undertaking the behaviour change.

The additional (marginal) cost of treating each successive patient may be non-linear. A practice with 150 diabetes patients, may find it relatively easy (low cost per patient) to complete cycles of care for the first 75 patients, who may be regular attenders. The following 50 patients may require more extensive recall, thus increasing the costs to the practice. The final 25 patients may be very difficult to contact such that effort required may not be worth the additional reward. This may be an argument to have a nonlinear payment scheme, where the payment to complete a cycle of care is, for example, $40 for the first 75 cycles, $60 for the next 50 cycles, and $80 for the final 25 cycles of care completed.

THE BENEFITS OF BEHAVIOUR CHANGE

Health professionals are motivated by a range of factors, and it is these that need to be influenced in order to change behaviour. Knowledge of what motivates health professionals is therefore important in designing policies to change behaviour.

In addition to extrinsic monetary rewards which make up health professionals’ personal income, other sources of motivation include funding for other services that is unrelated to their income; their patients’ health and well-being; professional autonomy; intellectual satisfaction, and; social and professional norms (peer esteem/reputation/status). The relative importance of these factors varies across primary care professionals, and it is this variation that is likely to influence the effectiveness of monetary transfers. For professionals where monetary motivation is more important than other sources of motivation, then financial incentives may have a larger effect. For health professionals who are more intrinsically motivated, then the same size of financial incentive may have less of an effect on their behaviour.

Frey (1997a) draws on theory and evidence from cognitive social psychology and argues that there may be cases where extrinsic incentives ‘crowd-out’ intrinsic motivation (Motivation Crowding Theory). This is where the introduction of external monetary rewards, where previously there were none, can undermine intrinsic motivation,

“...and even reverse the most fundamental economic law, namely that raising monetary incentives increases supply” (Frey and Jegen 2001).

They argue that for tasks and jobs where intrinsic motivation is high, the introduction of extrinsic rewards may de-motivate workers. Extrinsic incentives can be viewed as controlling, and may reduce trust and clinical and professional autonomy (Frey 1997b). A body of empirical evidence exists to support motivation crowding theory (Frey and Jegen 2001) and although there is no empirical evidence in health care, some commentators have suggested that intrinsic motivation in the medical profession may be declining (Jones 2002).

For those primary care professionals who are motivated by improvements in quality of care and patients’ well-being, then it is important that they believe that the rewarded behaviour leads to such improvements. This depends on the mix of patients they see, and whether the disease area targeted by the monetary transfer represents a large enough proportion of their patients. Primary care professional’s beliefs about the effect of the newly remunerated activity on health and well-being is also influenced by the level of information on current ‘best practice’ as defined by evidence-based clinical guidelines.
If the rewarded behaviour is not congruent with the opinions of professional or social peers, then the behaviour change is less likely to occur. Professional and group norms will also play a role in the effectiveness of monetary transfers.

If primary care professionals think the behaviour change and monetary reward may damage the doctor-patient relationship or continuity of care, or make patients less likely to attend, they will be less likely to change their behaviour. Their reputation and status amongst patients, especially in more competitive markets, will influence their response to monetary transfers encouraging the provision of new services or standards of care.

These factors are likely to be traded-off and interact. For some primary care professionals a sufficiently high monetary reward may compensate them for a loss of autonomy, such that on balance, they will respond to the monetary reward. For other primary care professionals, no amount of monetary reward would compensate them for a loss of autonomy. Some may accept lower earnings from the new scheme if they believe that the net effect on patient health status is positive, or that the behaviour re-inforces professional norms.

DEFINING QUALITY OF CARE

The term ‘quality’ can be very broad such as that used by the Institute of Medicine: “the degree to which health care services for individuals and populations increase the likelihood of desired outcomes and are consistent with current professional knowledge” (Lohr 1990). Their report in 2001, Crossing the Quality Chasm, identified six aims for health care that should guide quality improvement efforts: safety, effectiveness, patient-centeredness, timeliness, efficiency, and equity. The popular definition used by Donabedian (1966) includes aspects of health care related to structures (e.g. IT systems or disease registers), processes (e.g. recording of blood pressure, or the existence of quality improvement programs), or outcomes (e.g. clinical indicators or the health status of patients) Campbell et al (2000) provides a review of several definitions of health care quality and suggests that Donabedian’s three dimensions may also be split into ‘access’ and ‘effectiveness’. Economists also use the term ‘utility’ to refer to the well-being of individuals which, although still a broad concept, excludes measures of structure and process of care provided, unless they influence patients’ utility (Torrance 1986).

There may also be different types of quality indicators used for different levels of the health care system. The quality of national and regional health systems are more likely to include population-based measures of access and uptake of preventive services, and overall health of the population including mortality rates. Monitoring the quality of hospitals is more likely to include re-admission rates, waiting times, and minimum standards of safety, such as infection rates. Individual physician measures of quality are more likely to include evidence-based clinical and physiological indicators of quality (e.g. HbA1c, blood pressure), clinical behaviours (e.g. prescribing and test ordering), or measures of patient’s health outcomes and quality of life, such as mortality or patient reported outcome measures.

In this review we focus on defining quality of the care provided by the physician and so include clinical and physiological measures, clinical behaviours, and patient reported outcomes and experiences.
METHODS

LITERATURE REVIEW

Both a Cochrane systematic review and a narrative review and synthesis were conducted. The objective of the Cochrane review was to examine whether the use of financial incentives led to increases in the quality of care, and includes empirical studies with rigorous study designs that can demonstrate a causal effect. The objective of the narrative review was to help address questions about the context of each incentive scheme, how schemes evolved, what needed to be in place in terms of infrastructure and previous reforms, why schemes did or did not work, and for which population groups or types of primary care physicians were incentives used. All searches were confined to the period 2000 - current.

COCHRANE REVIEW

The Cochrane review followed the methods of the Cochrane Effective Practice and Organisation of Care (EPOC) group. A detailed protocol was written and refereed by the Cochrane EPOC group editors, and is now published on the Cochrane Collaboration website (See Appendix 1 for the Cochrane protocol). This includes details of the search strategy and methods of review. Cochrane EPOC reviews only include empirical studies with strong study designs including randomised trials, quasi-randomised trials, controlled before and after studies, and interrupted time series studies. The review included the following types of interventions aimed at primary care physicians.

1) Where the intervention changes the amount of payment (dose-response)
2) Where the intervention changes the method of payment:
   a. payment per unit of time (salary/sessional payment)
   b. payment for each service/visit/treatment/episode provided (fee-for-service)
   c. payment for each patient enrolled or registered with the PCP (capitation)
   d. payment for improvements in ‘quality’ (performance pay)
3) Where the intervention changes who is paid (e.g. from an individual to a group or team)
4) Multifaceted interventions that include at least two of the above
5) Multifaceted interventions that include at least one of the above plus and at least one other type of non-financial intervention designed to change professional behaviour (e.g. education, feedback, audit).

Primary outcome measures included the quality of care provided by PCPs that are related to patients’ health and well-being. This includes:

i) Patient reported outcome measures:
   a. self-reported measures of health status and health-related quality of life (e.g. SF-36, EQ5D).
   b. self-reported measures of patient satisfaction and experience with the process of care
ii) Clinical behaviours, such as prescribing, test-ordering, referrals, treatments or advice provided.

---

3 Target payments may be included under (b), (c), or (d), depending on the behaviour that triggers the payment. Target payments are simply a non-linear version of these payment schemes.
iii) Intermediate clinical and physiological indicators (e.g. HbA1c, blood pressure, cholesterol)

For (ii) and (iii) the study authors’ should state that the change in behaviour or indicator is related to changes in the quality of care. This claim should be justified with reference to evidence of a positive effect of the behaviour or indicator on health and well-being, such as references to evidence-based clinical guidelines or systematic reviews. If a statement about improvements in quality of care is not made in the paper, or is made but not justified through references, then
- the authors will be contacted to establish whether any evidence existed at the time the study took place, and;
- the review authors will search for Cochrane reviews or national clinical guidelines that existed at the time the study was conducted.

In addition to (i) to (iv), if a study reports other primary outcomes that are not measures of quality of care (e.g. costs, health professional satisfaction or experiences, and clinical behaviours where there is no evidence of a link to quality of care, e.g. number of services provided or number of patients seen), then these were reported and summarised in separate tables, but not synthesised. This is important in helping to identify unintentional effects of the intervention.

NARRATIVE REVIEW AND SYNTHESIS

Our aim for the narrative review was to undertake a review and synthesis of the evidence base to identify the role of financial incentives to improve the quality of primary health care. This reviewed schemes that have attempted to link financial incentives to the provision of higher quality primary health care in Australia, United Kingdom, and United States4. Primary health care will be defined as services organized around primary care physicians or medical professionals (and their teams) practicing in primary care settings.

The conceptual framework identified a number of issues that helped to structure the narrative review (Table 1). The aim was, for each country, to address the questions in Table 1. The search strategy began through using pre-identified literature, which included studies already known by the review authors and references lists from existing systematic reviews of financial incentives in health care. An initial literature search in each country was conducted, and this provided the baseline literature to begin the narrative review for each country and incentive scheme. If necessary, more in depth searches were conducted as the review progressed. Common themes and issues from each incentive scheme were summarised.

DEVELOPMENT OF POLICY OPTIONS

As the literature reviews were being conducted in early 2009, face-to-face interviews with a group of key stakeholders were conducted to examine the Australian context for the potential introduction of financial incentives linked to the quality of primary health care (see Appendix 2 for the list of key informants). Was there potential? What needed to be in place? What were the potential problems? The information from these interviews was then used to inform the direction of narrative review. The results of both reviews were then used to identify a number of issues that formed the basis of the policy options for Australia. A draft set of policy options was developed for Australia which were then discussed with key stakeholders in a second face-

4 New Zealand and Canada were also examined, but no schemes exist currently that use monetary transfers to influence primary care physician incomes, conditional on the achievement of levels of quality of care.
to-face interview in spring 2009. These were then refined, based on their comments, to arrive at a final set of policy options.

Table 1. Key questions for the narrative review

<table>
<thead>
<tr>
<th>Research Objectives</th>
<th>Key Review Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>What contextual factors are relevant in each incentive scheme?</td>
<td>How did the schemes evolve and develop in each country?</td>
</tr>
<tr>
<td></td>
<td>What IT, infrastructure, professional support, role of regional primary care organisations, etc were in place or developed?</td>
</tr>
<tr>
<td></td>
<td>What other changes occurred alongside the incentive scheme?</td>
</tr>
<tr>
<td>What were the key characteristics of the incentive scheme?</td>
<td>What are the characteristics of the financial incentive scheme (eg how was financial reward linked to PCP performance to create an incentive)?</td>
</tr>
<tr>
<td></td>
<td>For what sub-groups of patients/population were incentives used?</td>
</tr>
<tr>
<td></td>
<td>What groups of primary care physicians/teams/practices were the targets of incentives?</td>
</tr>
<tr>
<td>What were the objectives and intentions of the scheme, and were they achieved?</td>
<td>What was the assumed mechanism by which these schemes were expected to work?</td>
</tr>
<tr>
<td></td>
<td>What were the original objectives of these schemes?</td>
</tr>
<tr>
<td></td>
<td>Were the original objectives of the scheme achieved?</td>
</tr>
<tr>
<td></td>
<td>Were there any unanticipated effects of the schemes?</td>
</tr>
<tr>
<td>What incentive approaches were (or were not) used for whom (specific patient</td>
<td>Was there evidence of effects on recruitment/retention (i.e. self-selection of physicians/teams/practices into/out of the scheme).</td>
</tr>
<tr>
<td>populations, GPs, teams), for which disease areas (chronic disease, prevention, acute) and why?</td>
<td>Were there reasons why some physicians/teams decided not to participate in the scheme?</td>
</tr>
<tr>
<td></td>
<td>For those that did participate, was there evidence of variation in effects of the scheme across different types of patient/population group, or different types of PCP or team</td>
</tr>
</tbody>
</table>


RESULTS

COCHRANE REVIEW

DESCRIPTION OF STUDIES
We identified 2933 potentially relevant studies, of which we excluded 2635 based on their title and year of publication. Of the remaining studies we retrieved 298 for more detail and from these we excluded a further 263 studies based on the identified exclusion criteria. (Appendix 3).

35 publications relating to financial incentives and quality in primary health care were selected for a more detailed evaluation. Of these, 29 studies were excluded from the review (insert Appendix of ref liist). Three cluster randomized controlled trials (An et al. 2008, Roski et al. 2003, Twardella and Brenner 2007); two controlled before and after studies (Gosden et al. 2003, Rosenthal et al. 2005), and; one interrupted time series study (Young et al. 2007a) met the EPOC criteria for inclusion in the review and are listed in the table of included studies.

TYPE OF OUTCOMES/TARGETED BEHAVIOUR FOR CHANGE
The three included C-RCTs examined how financial incentives for PCPs impact on their adherence to smoking cessation guidelines and patients’ smoking cessation behaviour (An et al. 2008, Roski et al. 2003, Twardella and Brenner 2007);. One study assessed the effect of change in payment method from capitation to salaried contracts and the consequences of such a change on patients’ assessment of quality of care (Gosden et al. 2003). The two remaining studies evaluated the impact of financial incentive programmes on clinical quality measures of diabetes (HbA1c, Urinalysis, LDL, and eye exam), cervical cancer screening and mammography (Rosenthal et al. 2005, Young et al. 2007a).

CHARACTERISTICS OF SETTINGS AND PROFESSIONALS
Among the included studies, four interventions took place in the USA, one in the UK and one in Germany. The four US studies used incentive schemes put in place by large private health plans to increase the quality provided by their contractors (individual or group practices).

Young (2007a) involved 334 primary care physicians from Rochester (New York) in Independent Practice Associations (IPA) contracting with Excellus health Plans, an HMO with approximately 300,000 patients. Roski et al. (2003) included forty clinics which provided primary health care service within a large Midwestern multispeciality group practice. They also surveyed 66,000 of their patients about their smoking cessation behaviour. Similarly, An et al. (2008) involved 49 primary care clinics which were members of the same physician association, and contractors with several major Minnesota health plans. Rosenthal (2005) examined the introduction of a pay-for-performance initiative in a large Californian group of health plans.

Gosden (2003) examined data from 10 general practices and 10 control practices in the UK. Data from around 200 patients from each practice was also used. Twardella (2007) study involved 82 medical practices, including 94 general practitioners in the Rhine-Neckar region (southwest Germany) and 577 of their patients, to evaluate new strategies to enhance the promotion of smoking cessation in general practice.
CHARACTERISTICS OF INTERVENTIONS

Four out of the six studies examined the effects of incentives on reaching pre-specified clinical targets (Roski, An, Twardella, Rosenthal). One study awarded bonuses based on relative performance of medical groups (Young), and one examined a change in the method of payment (Gosden).

Roski et al. initially randomised forty clinics providing primary health care services in two intervention groups and one control group. Three clinics closed during the trial for business reasons and were excluded from the analysis. The thirteen clinics in the first intervention group received printed versions of smoking cessation guidelines and bonuses paid to the medical group for reaching preset clinical performance targets set up by the medical groups' management. For the nine clinics in the second intervention group, the printed guidelines distribution and financial incentives system were combined with access to a telephonic smoking cessation counselling service. The last 15 clinics served as controls and received the same guidelines as the ones distributed in the intervention groups.

An's study is more restricted in terms of objectives as it mainly focused on clinicians' practice related outcomes. In this study, 49 clinics that provide adult primary care services were randomly assigned to either intervention or usual care group. Physicians in the 24 intervention clinics were encouraged to refer patients who currently smoked to a smoking cessation counseling service. Each patient referred received up to 3 telephone calls from the service over a 2 week period in order to be offered counseling. Clinics that referred 50 smokers would receive a $5000 performance bonus. Clinics would also receive $25 for each referral beyond the initial 50. Incentive payments were made to clinics as a lump sum at the end of the contract period. Payments were made to each clinic rather than individual physicians.

Twardella is a study on smoking cessation behavior and physician education strategies to enhance smoking cessation promotion in general practices. Eighty two medical practices, with a total of ninety four GPs, were randomly assigned to three intervention arms and one control group. In the first intervention arm 24 GPs from 21 medical practices were provided with two sessions of a two hour group tutorial on methods of promoting smoking cessation. GPs received €130 for each patient they recruited, who was a non-smoker twelve months after the end of the study. The second intervention arm included 23 GPs over 21 Practices, who received the same group tutorial and instead of personal financial reward, they had the possibility to offer cost-free prescriptions of drugs proved effective in supporting smoking cessation to their patients. The last arm was a combination of the interventions implemented in the two other arms. The 26 GPs allocated in this arm received the group tutorial on smoking cessation methods, a financial reward and the possibility of offering cost-free drug prescriptions to selected patients. The effect of the different interventions was measured by the number of self-reported abstinent smokers at 12 months follow-up and validated by serum cotinine analysis.

Rosenthal used a controlled before and after design and longitudinal administrative data on the performance of physician groups obtained from a large group of Californian health plans to assess quality improvement after the introduction of a pay-for-performance scheme. The incentive program was introduced in January 2003 for approximately 300 medical groups, members of the Californian network of the health plan. Overall the program targeted ten clinical and service quality measures, however the investigators focused on three of the clinical measures for which data was complete (rates of cervical cancer screening, mammography, and hemoglobin A1c (HbA1c) testing for diabetic patients). Performance targets were set as the 75th percentile of 2002 performance by the physician groups and were made known in advance to the participating physician organisations. Participants received a quarterly bonus of approximately $0.23 per member per month for each performance target met or exceeded. As a control site the investigators used data from 42 medical groups in the Pacific Northwest operating with the same health plan as the Californian network. The same data collection and feedback procedures had been put in place there without the incentive scheme.
Young et al. conducted an interrupted time series analysis to measure the association between a performance-based payment program and physicians’ adherence to diabetes control measures. The scheme was introduced in 2002 for GPs who were members of an independent practice association contracting with a private health plan. Physicians were placed at limited financial risk for their performance with respect to clinical quality, patient satisfaction and practice efficiency. Each physician had approximately 5% of his or her fees withheld to fund incentive pools. The money was then distributed to physicians based on their relative performance.

Gosden’s CBA investigated the effect of a change in payment method for GPs on the level of quality of care perceived by the patients. A policy change in the UK was an opportunity for the authors to compare salary and capitation payment systems in terms of their relative impacts on the quality of care. Ten practices in which GPs switched to salaried contracts were selected and matched with ten control practices with standard capitation contracts. Between 200 and 250 patients, drawn from the lists of the selected practices, were surveyed before and after the policy change occurred. Patients were asked to rate the quality of thirteen aspects of service provision on a 0-100 point scale: (Access, receptionists, continuity of care, Technical care, communication, interpersonal care, trust, knowledge of patient, nursing care, co-ordination of care, referral, recommendation, overall satisfaction). Information was obtained using the General Practice Assessment Survey (GPAS).

EFFECT OF INTERVENTIONS

Overall, the different interventions and financial incentives programmes had limited effects on quality.

An’s C-RCT showed the most significant effect despite a number of limitations and source of bias. Clinics who received a financial incentive had a rate of smoker’s referral 7.2 percentage points higher than the usual care clinics (11.4% compared to 4.2%). These results remained consistent after adjustment for several predictors of referral such as the clinic’s history of engagement with quality improvement activities, and the number of physicians in the clinic. The intervention was characterised by easy referral procedures (development of a single fax referral form and telephone number, system of referral triage to the Quit Helpline or appropriate health plan number) in which the intervention was implemented. Lack of blinding has also been reported as a potential source of bias in the study. Finally, the authors mention issues regarding the optimal magnitude, duration and timing of the incentive which could have limit the size of the effect.

Roski et al. showed a significant positive effect for only one outcome, as the percentage of patients who had their tobacco use status identified by the GPs at last visit increased by 14.1 percentage points in the incentive clinics, compared to an increase of 6.2 in the control clinics. In the second intervention group, the financial incentive was coupled with a smoking cessation counselling service but the increase in smokers’ identification rate was 8.1 percentage points. Results for the clinicians’ adherence to other smoking cessation guidelines recommendations (advice to quit and assistance to quit) were not statistically significant. Difference in results for patient smoking cessation behaviour between the three experimental groups were not statistically significant for six of the seven investigated outcomes. 3.3% of patients from clinics who received an incentive and had access to a registry, reported having used a counseling service for smoking cessation. This is compared with 1.3% of patients from incentive clinics and 1% from control clinics. The study is exposed to a number of sources of bias such as the lack of blinding. Also, an important legal decision against the tobacco industry that took place in Minnesota at the same time as the intervention may have influenced the behaviour of clinicians and patients and had an impact on the results of the study.

In Twardella’s intervention, the financial incentive had no effect on patient’s smoking behaviour. The investigators reported a smoking abstinence at twelve months follow-up for 3% of the patients in the training plus incentive group, 12% for those in the training and free
prescriptions group, and 15% in the group having the two interventions. Patients in the control clinics had an abstinence rate of 3%. The authors clearly state that “even a high payment is not an effective measure to increase general practitioners’ successful advice for smoking cessation”.

Rosenthal’s CBA was able to show a positive effect of the financial incentive scheme for only one of the three outcomes investigated. No difference was observed between the control and the intervention groups for HbA1c testing, and a non-significant difference of 1.7 percentage point was reported for mammography screening. For cervical cancer screening, the 3.6 percentage point difference in improvement between California and the Northwest was significant. Overall the quality increased in both intervention and control sites and again, the specific effect of the financial incentive on quality is hard to observe. Also, the design of the study exposes the results to a number of biases due to the lack of randomisation and blinding.

Gosden’s CBA on the change of method of payment of GPs has shown an increase in patient’s perceived quality for seven out of ten aspects of care investigated in salaried practices compared to practices under capitation contracts. However, none of the results was statistically significant, and there is no evidence to claim a difference in quality between the intervention and control practices. The CBA design exposes the study to a number of biases due to the lack of randomisation and blinding.

Young’ ITS reports a modest effect of the financial intervention on one of the four diabetes performance measures investigated. Physicians’ adherence to clinical guidelines for eye examination increased by seven percentage points in the year after implementation of the programme, and appeared attributable to the incentive scheme. It is not clear why only the eye exam showed a statistically significant improvement. The authors point to the fact that there was more opportunity for improvement as the procedure had the lowest baseline score. The overall limited effect of the schemes may be related to the specificities of the IPA network and the fact that individual practices may lack the resources and structures necessary to enhance clinical quality. The authors also argue that observable change in professional behaviour may require more than the three years of the study period.

NARRATIVE REVIEW AND SYNTHESIS

This section summarises the results from the synthesis of the narrative review for each country and incentive scheme. Details of the number of studies included in the narrative review are shown in Appendix 4. The review of the literature that addresses the questions in Table 1 for the US, UK and Australia is shown in Appendix 4.

The US has the most experience with performance pay schemes in health care. There are hundreds of schemes within Health Maintenance Organisations, large schemes involving coordination across a number of private health insurers, and national schemes are being currently developed and evaluated in Medicare. The UK National Health Service introduced the Quality and Outcomes Framework (QoF) for primary care physicians in 2004. This scheme is notable in that it covers over 98% of GPs in the UK, accounts for an average 25% of GPs’ income, and has XXX different indicators linked to financial incentives. Australia has the Practice Incentive Program which includes Service Incentive Payments that have been used since 2001. This scheme is operated by Medicare and has changed little since 2001.

KEY FINDINGS

Given the lack of rigorous evidence from the Cochrane review on the impact of these schemes on quality of primary care, a number of issues arise, that are common across the countries and schemes, that would enhance the likelihood of these schemes improving quality of care.
1. **Build on what exists already.**

The schemes evolve over time and usually involve a series of complex interventions that include accreditation, education, existence of quality improvement programs, investment in IT and data collection systems, and professional support. These are all necessary interventions that create the conditions for linking financial incentives to quality of care, and would normally precede or be introduced alongside the financial incentive. For example, in the US, the introduction of pay for performance in HMOs was preceded by a long history of managed care which introduced the necessary data systems for utilisation review, and physicians were used to external review of their activities and reporting requirements. In the UK NHS, there was a long history of structural reform, performance management, national clinical guideline development, target payments for immunizations, and development of IT infrastructure before the QOF was introduced. In Australia, the Practice Incentive Program was introduced in 1999, closely followed by accreditation (a requirement to join the PIP), and also additional investment in infrastructure via Divisions of General Practice, before Service Incentive Payments were introduced in 2001.

2. **Reward for improvement in quality as well as achievement of levels of quality.**

Rewarding for the achievement of a specified level of quality is most common, but may not be as effective at improving quality compared to rewarding for improvements in quality, for several reasons:

- for those who already are providing high quality of care, the reward represents a windfall financial gain with no change in behaviour, and so it rewards past achievements or already high levels of ability;
- it does not encourage improvements beyond the specified level of quality, and;
- for primary care physicians with very low baseline levels of quality, the costs of achieving the specified level seem very high and they may not respond or participate at all – yet these are the very groups that one would need to target and were substantial gains in quality could be made at little cost.

Rewarding for changes in quality encourages those with low baseline levels of quality to improve quality of care, and where the largest increases are likely to be made. The costs of achieving a given change from a low baseline are likely to be lower compared to the costs for those already providing high quality. For a given investment of resources, those with lowest baseline quality are likely to have the largest effects, compared to those with highest baseline quality, where the ‘easy’ gains have already been made. Rewards for those who already provide high quality care, but find it difficult to improve any further (i.e. rewards for maintenance of a given level of quality), should still be made but should not be the main element of the incentive scheme.

For example, the UK QoF policy documents did not mention rewarding for improving quality, but just for ‘providing’ quality of care.

3. **Financial incentives linked to quality should comprise at least 10% of physician’s total revenue.**

For many schemes in the US, rewards were less than 10%, and this is an often cited reason why such schemes may not have worked. The incentives were not large enough to cover the costs of changing behaviour. Incentives were funded from within existing cost or efficiency savings and so for some schemes no new resources were invested. However, in the UK there
was a context of increasing health care spending generally, and so rewards comprised around one-quarter of income, which included a substantial increase in income of 35% over two years when the scheme was introduced. Although there is almost 100% of doctors achieving the top targets in the QoF, there is little rigorous evidence of any changes in quality or whether the pre-existing trend would have continued, or whether any changes in quality are ‘worth’ the investment in resources. In Australia, Medicare data from 2003 suggested that the PIP scheme represented an average of around 9% of GP’s revenue from Medicare (Ref senate committee). It is likely that this is higher in rural and remote areas. At the time this represented additional resources invested in primary care.

4. Financial incentives are more likely to have an effect where there is one single funder of primary care services.

With multiple funders, the effects of incentives introduced by one funder will be less effective. Experience from the US suggests that where a physician contracts with multiple insurers or HMOs, any single HMO has less influence on changing their behaviour. The IHA scheme is a collaboration of a number of insurers in California that use pay for performance schemes that cover X% of the population. HMOs that employ salaried physicians or where physicians contract with only one insurer may have stronger effects on changing behaviour. In the UK, negotiations to implement the Quality and Outcomes Framework in the UK were simplified because a single funder (the UK Department of Health) could negotiate with doctors representatives (the BMA and RCGP) to agree on a new contract for General Practitioners.

5. A stable and enrolled population will strengthen the role of financial incentives.

It is easier to attribute changes in quality and performance to a specific provider/medical group who has been responsible for coordinating the health care of their patients. The incentives may be weaker if a patient visits multiple providers, since providers don’t feel ‘responsible’ or don’t feel they have control over their performance, and so may be less likely to change their behaviour. This is a particular issue in the US and Australia where patients can visit any primary care physician, and in the US where patients can visit any medical practitioner.

6. Incentives should be paid at the practice level, and payments to health professionals involved in improving quality of care should be equitable with respect to their skill and effort.

Incentives for quality should be targeted at those health professionals involved in the provision of health care to patients. There should be a mix of payments to the practice or organisation and payments to individual health professionals for quality improvements. There is little evidence to suggest that one or the other should be targeted. Paying only teams raises issues of how a team is defined, and how the financial rewards are used by the team. These issues may mitigate the intended effect of the financial incentives, though it does give discretion to the team to use the rewards in a number of ways which might include investing in further quality improvements, or rewarding team members, or a mix. Payments to individual professionals may not encourage team work or co-ordination of care (and thus may not improve quality of care where team-based coordinated care is needed). Any payments to team members should be equitable with respect to their skill and effort, or health professionals who feel unfairly treated may lose motivation to change behaviour.
7. Keep payments simple to administer for primary care practices.
This is a particular issue for Australia, where several ‘red tape’ reviews have taken place with respect to GP payments. The fee-for-service system of subsidies run by Medicare is costly to administer, and the temptation to ‘add’ items and programs to the Medicare Fee Schedule, each with its own regulations and eligibility, creates substantial complexity for GPs and patients.

8. Consider potential unintended consequences.

There is a possibility of undesirable consequences of these schemes, such as a focus on the remunerated areas at the cost of unremunerated areas, and these need to be thought through and managed carefully. Primary care physicians have limited time and need to allocate their time effectively – the relative prices and payments for different types of services will influence how time is allocated. Although these issues are often talked about, in the context of primary care there were few empirical studies examining potential adverse consequences of pay for performance. Evidence from the UK suggests that there were positive spillovers to unremunerated disease areas (Sutton et al. 2009). This reflects the issue that a practice that provides high quality of care in one disease area is also likely to provide high quality of care in other disease areas as well. Gravelle et al (2008) found evidence that UK GPs who failed to score maximum QOF points in the first year, ‘gamed’ the system by increasing their exception-reported patients (who would not count for their QOF scores) in the following year. However, the possibility of such undesirable consequences should not be a reason for inaction or used to argue that such schemes should not be introduced, and should be carefully considered when designing the scheme.


Only six empirical studies of reasonable quality were included in the Cochrane review. The need for rigorous and controlled evaluations of any changes, although presenting challenges, should be seriously considered before a scheme is fully implemented. Rigorous but feasible study designs should be developed that compare practices receiving new incentives with those who do not.
POLICY OPTIONS FOR AUSTRALIA

The aim of this section is to propose a number of options for the linkage of financial incentives to quality of care improvements in Australian primary care. These are based on the eight key issues identified from the narrative review and synthesis in the previous section, and the current policy context described below.

CURRENT POLICY CONTEXT

Funding of primary care services in Australia is largely through patient subsidies delivered through the Medicare Benefits Schedule, the Practice Incentive Program, funds delivered through Divisions of General Practice, and jurisdiction-run health centres. The current context provides a relatively strong basis for developing linkages between funding and the quality and performance of primary care services.

There are a number of contextual factors that could be built on in Australia:

- an existing blended payment model that has been in place since 1999 (the Practice Incentive Program), with elements of performance-based pay since 2001;
- a number of existing primary care funding models that involve alternative funding models, including Community Health Centres in Victoria, corporate and...(more);
- support provided to general practices through regional primary care organisations (Divisions);
- the majority of general practices are accredited and computerized;
- there is a growing culture of quality improvement in primary care based on accreditation and also for those practices participating in the Australian Primary Care Collaboratives Program (APCC);
- support for blended payment models from professional and primary care organisations such as the Royal Australian College of General Practitioners (RACGP), the Australia General Practice Network (AGPN), and the National Primary Health Care Partnership (NPHCP), and;
- there is an emerging consensus that such payments should be linked to chronic disease, complex conditions, and preventive activities, with FFS remaining for acute conditions.

The proposed reforms by the National Health and Hospitals Reform Commission (NHHRC) and the Draft National Primary Health Care Strategy included a number of recommendations that would further facilitate the linkage of financial incentives to quality in primary care (refs).

From the NHHRC report:

- the Commonwealth taking responsibility for funding all primary care services including elements currently funded by State and Territory governments;
- the ‘promotion of a culture of improvement through health performance reporting’ (p133) that includes clinician-level reporting of performance;
- the idea of a ‘health care home’ to facilitate voluntary enrollment of patients with chronic disease and complex conditions, and;
greater scope for blended payments whilst retaining fee for service, including: grant payments based on the size of the enrolled population; outcomes payments to reward the performance for enrolled patients; and episodic/bundled payments (p136.)

The Draft National Primary Health Care Strategy also moves in the same direction, and proposes:

- improvements in quality, safety, performance and accountability (including peer feedback and comparison of performance);
- new/changed funding models (but has not yet specified any specific proposals);
- voluntary enrollment for those with chronic disease;
- a greater role for regional primary care organisations, and;
- improved information systems to support improved quality and performance, prevention and access to chronic disease services.

The report of the Expert Advisory Group to the Primary Health Care Strategy recommends pay for performance as a key building block;

“...progressive introduction of pay-for-performance arrangements which are linked to improvement of patient health outcomes to support participation in the framework and continual quality improvement activities.” And, “...there is scope for these types of incentives to play a much larger role in achieving improved health outcomes.”

A WAY FORWARD FOR LINKING FINANCIAL INCENTIVES TO QUALITY IMPROVEMENT IN AUSTRALIA

The use of financial incentives to improve quality of care should be expanded in Australia. This should be done carefully and in stages, that recognise current data infrastructure limitations and that such a scheme should be driven by a quality improvement agenda.

Even in the absence of strong evidence from overseas of the effect of financial incentives on quality improvement, such schemes could be justified on the basis that such payments (and associated interventions) may shift the culture of primary care away from volume and towards health outcomes and quality. Relatively larger rewards are also likely to influence recruitment and retention in primary care, and so improve access even if quality does not improve. Arguments have been made that such schemes merely increase the recording of information to obtain payment (e.g., blood pressure and smoking status), but this in itself can be regarded as an improvement in quality of care and is certainly an important base for monitoring subsequent incentives to improve quality. Nevertheless, the investment of scarce health care resources in such schemes does require that they have some direct impact on quality.

PHASE 1. MODIFY CURRENT INCENTIVE SCHEMES (2011 TO 2013)

The PIP scheme and its Service Incentive Payments and Outcomes payments have changed little since it was introduced in 2001. GPs are used to it, and many GPs have become accredited to be able to claim these incentive payments. This scheme should therefore be built upon. Similarly, the General Practice Immunisation Incentives Scheme should be modified.
Two main changes should be made:

1. Reduce the value of SIP payments (or remove SIP payments), and re-invest this money to increase the value of outcome payments, which will be modified to reward quality improvements (see examples below).

2. To reward practices already providing high quality of care, but who may find it difficult to improve further, a practice which maintains a given level of coverage across all incentivised disease areas, for a minimum time period (e.g. one year), would receive an X% ‘quality supplement’ to their total PIP. This should be at least equal to the value of their SIP and outcome payments in the year before the changes are introduced, so they do not lose any income, but can earn additional income for further improvements.

For the changes to incentive payments in (1) above, the following should be considered.

**DIABETES.**
- Reduce the SIP payment to $10 for each cycle of care per patient and re-invest in outcomes payments.
- New outcomes payments should be introduced and paid for each 5% \(^{5}\) increase in the proportion of eligible patients receiving a completed cycle of care. They should be paid to practices. The payment should be a fixed amount per patient (say $50), so if a practice improves by 5% and this is equal to 5 patients, they will receive $250. If a practice improves by 5% and this is equal to 10 patients, they receive $500. This creates equity across practices with different patient loads.
- To reflect the increasing costs of each additional patient, the fixed amount per patient should increase as higher proportions of cycles of care are completed. For example, the payment may start at $30 for each 5 percentage point increase between zero and 50% of eligible patients. It could increase to $50 for each 5% increase between 50% and 75% of eligible patients, and to $70 for each 5% percentage point increase above 75% of eligible patients. The thresholds and size of payments would need to be based on evidence of how costs increase with increasing numbers of patients.

**ASTHMA**
- There is currently no outcomes payment for asthma since there are no centrally collected data on the number of patients with asthma treated by each practice, only cycles of care claimed. In the short term, prescribing of asthma medications could be used to construct a denominator for each practice, and the proportion receiving cycles of care can then be calculated.
- Using the same method as for diabetes, the SIP payment could be reduced to $10 and an outcomes payment introduced to reward improvements in quality as for diabetes.

**CERVICAL SCREENING**

\(^{5}\) 5% is arbitrary, and would need to examine Medicare SIP data and APCC data to decide actual proportion.
The current outcomes payment should be increased in value and modified. It should be paid for each 5% increase in the proportion of eligible patients screened. A fixed payment per patient should be paid, and this should be increased as a higher proportion are screened (see diabetes example). This should have the same effect as the current SIP payment for those who haven’t been screened in the last 4 years (the hard to reach), and so the current SIP payment can be removed.

CHILDHOOD IMMUNIZATION

The current outcomes payment should be changed and paid for each 5% increase in the proportion of eligible patients immunized, with a fixed payment per patient that increases as the proportion immunized increases, as in the diabetes example.

There should be a formal and rigorous evaluation of these modifications to inform further development in Phases 2 and 3. Initial research using Medicare data, data from APCC and BEACH should be conducted to establish the broad parameters of the scheme and model/simulate its likely impact on behaviour and expenditures. This modelling should be conducted during 2010 to inform Phase I. The study design for a national demonstration project and evaluation would be formulated in 2010, with the trial beginning in 2011. This trial would be continued throughout phases 2 and 3 below.

PHASE 2. DEVELOP INFRASTRUCTURE FOR A NEW QUALITY INCENTIVE SCHEME (QIS) FOR CHRONIC DISEASE AND COMPLEX CARE (2011-2013)

The modified incentive payments in Phase I require minimal reporting of data by GPs, as existing Medicare and other data sources are already used to calculate the SIP payments for each practice.

Further infrastructure for data collection is necessary if incentives are to be expanded to other priority disease areas. This should be based around the methods being used by the Australian Primary Care Collaboratives. They already used an established methodology and formal quality improvement cycle that support participating practices with IT infrastructure and data reporting. Divisions could be more heavily involved in supporting practices to undergo quality improvement cycles and with IT infrastructure. The collaboratives should be explored as the basis for a future expansion of quality improvement and pay for performance.

PHASE 3. INTRODUCE A NEW QUALITY INCENTIVE SCHEME (2014)

Abolish PIP. All SIP and outcome payments to be rolled into new QIS scheme with additional funding.

Roll all other PIP capitation payments (eg after hours, teaching, IT etc) into a single infrastructure grant based on the number of enrolled patients seen, with a loading for rurality and the maintenance of quality. This would help simplify the current PIP scheme. Minimum levels of infrastructure would need to be in place to receive these grants (e.g. accreditation, access (opening hours), disease registers etc) that could be specified within new accreditation to include the elements of infrastructure currently rewarded under the PIP scheme (e.g. after hours, IT). The size of the grant will be
determined by the size of the enrolled population per practice (across all disease areas that require enrollment), with % supplements for practices in rural areas. Grants could also have ‘opt-in’ and ‘opt out’ elements depending on which services are required to be provided by all practices, and which are at the discretion of practices to provide (i.e. dependent on local circumstances). For example, teaching, care for specific populations (e.g. homeless), procedural work, could be ‘opt in’ elements of the grant. Grants could also have a number of elements, including direct payments contributing towards the salaries of practice staff (e.g. practice nurses).

- Link outcome payments to patients who are enrolled. This would form the denominator for the incentive payments, so practices would be asked to report disease prevalence for their enrolled population. It would enable the scheme to be expanded to other disease areas.

- The APCC quality improvement scheme would be rolled out across Australia and managed by regional primary care organisations and APCC. This will require further investment in practice IT infrastructure to enable consistent reporting of performance indicators and disease prevalence.

- Consider extending to other disease areas, and beyond process measures such as the proportion of HbA1c tests or the proportion of completed cycles of care, to whether an HbA1c test was taken, to the proportion of patients who have blood sugar levels are within acceptable ranges. This will build upon more rigorous reporting of data from practices under the APCC framework. Other preventive areas such as risk factor reduction and other screening programs should be included, especially where a national screening program exists. Other chronic disease areas should only be considered if there are sufficient infrastructure to produce relevant data, and evidence-based clinical guidelines that clearly link performance indicators to higher quality of care and improved population health.

- Whether new disease areas should be included should be considered by an independent body which can summarise and synthesise evidence along similar lines to MSAC and PBAC. The evidence on existing disease areas in the scheme would be regularly reviewed and targets and incentives updated if necessary, and if behaviour change has been achieved/plateaued, the incentive scheme should be altered to further target the incentives or to focus on new disease areas or to reward maintenance of quality.

- The total payments offered would need to be higher than those made in the current PIP scheme to enhance the attractiveness of the scheme and encourage higher participation.

- The changes above should be carefully piloted and evaluated. Demonstration sites should be selected based on a range of factors, and include practices with low baseline performance.
REFERENCES


CENTRE FOR MEDICARE AND MEDICAID SERVICES (2009a). 'The Medicare Care Management Performance Demonstration Fact Sheet'.


COCHRANE EFFECTIVE PRACTICE AND ORGANISATION OF CARE GROUP (2009). 'Cochrane Resources for Review Authors'.


a six-year evaluation of a physician pay-for-performance program', Health Serv Res, Vol. 42, pp. 2140-2159; discussion 2294-2323.


APPENDIX 1. COCHRANE EPOC SYSTEMATIC REVIEW PROTOCOL

THE EFFECT OF FINANCIAL INCENTIVES ON THE QUALITY OF HEALTH CARE PROVIDED BY PRIMARY CARE PHYSICIANS

Review authors:
Prof. Anthony Scott, Dr. Peter Sivey, Mr. Driss Ait Ouakrim, Ms. Lisa Willenberg
Melbourne Institute of Applied Economic and Social Research, The University of Melbourne

Dr. Lucio Naccerella
Australian Health Workforce Institute, The University of Melbourne

Dr. John Furler, Prof. Doris Young.
Department of General Practice, The University of Melbourne

Acknowledgements
The review is funded by the Australian Primary Health Care Research Institute. Thanks for comments from Emma Tavender and the EPOC editors.
BACKGROUND

Primary health care provides integrated, easy to access, health care services by clinicians who are accountable for addressing a large majority of personal health care needs, developing a sustained and continuous relationship with patients, and practicing in the context of family and community (Vanselow et al. 1995).

The use of blended payment schemes in primary care, including the use of financial incentives to directly reward ‘performance’ and ‘quality’ is increasing in a number of countries. There are many examples in the US and the Quality and Outcomes Framework (QoF) for GPs in the UK is an example of a major system-wide reform (Christianson et al. 2007, Institute of Medicine 2007). In Australia, the Practice Incentive Program (PIP) was introduced in 1999 to improve the quality of care provided in asthma, diabetes, mental health and cervical screening. Despite the popularity of these schemes, there is currently little rigorous evidence of their success in improving quality, or of whether such an approach is cost-effective relative to other ways to improve the quality of care (Gosden et al. 2000, Scott 2007).

The role and impact of such incentives may depend on a number of factors, including the amount of the payment, how the payment is made (eg salary, fee-for-service, performance bonus or targeted at individuals or teams) and on the factors that motivate primary care teams and their members that include how important income is relative to other sources of intrinsic and extrinsic motivation they may have (e.g. professional autonomy).

These sources of motivation can influence the success or failure of these schemes and may vary across doctors in different settings and doctors with different preferences and practice styles. The success of financial incentives is also dependent on the costs of participating in these schemes, which are likely to vary across different primary care teams including differences in administrative costs borne by different primary care teams of different sizes, the additional effort required by the primary care team to increase quality of care to meet the incentive requirements, and support provided to help reduce these costs by regional primary care organisations or clinical networks (De Domenico M et al. 2005, Scott 2007).

Both theory and empirical evidence suggests that, in some circumstances, poorly designed incentives may not work at all, can lead to lower levels of quality, or unintended and distorted behavioural effects. For example, financial incentives applied to one disease area may ‘work’, but at the cost of GPs spending less time in other disease areas or with other types of patient, such that the overall net impact on quality of care and costs is difficult to determine.

DEFINING FINANCIAL INCENTIVES

In their most basic form, financial incentives are created through the transfer of money from a ‘buyer’ (in our case a patient or third-party funder/insurer) to a ‘seller’ (in health professional or team of health professionals) that is conditional on the seller behaving in a defined way, such as providing a good or service. In economic theory, these monetary transfers change the ratio of the costs and benefits of specific decisions or behaviours. The money may be used to reduce the costs of the behaviour change, thereby making it more likely to happen by increasing the cost-benefit ratio, or may be used as a reward that can be used to purchase other goods and services, again by increasing the cost-benefit ratio.

These principles apply in health care where the funders (the buyers) of health services (third party payers such as insurers or governments) wish to encourage providers (the sellers) to provide efficient and equitable health care for their population.

The monetary transfer is usually part of the personal income or remuneration of the health professional. In-kind transfers of specific resources (e.g. equipment, or salary for a practice nurse) are not included in this review. There are number of characteristics of the payments that can influence behaviour.
1. METHOD OF PAYMENT:
i) Payments can be made in exchange for the following provider behaviours:
   • working for a specified time period (e.g. salary, sessional payment),
   • providing specific services/treatments/episodes/cases (fee-for-service),
   • providing care for a patient or specific population (capitation),
   • providing a pre-specified level, or increases in the level, of quality of care (e.g. performance-based pay).

ii) Payments may be linear or non-linear:
   • linear, so the same payment is made for each additional unit of payment (e.g. service provided)
   • non-linear, such that payment is conditional on reaching a threshold or target, or a series of thresholds, or that the amount of the payment changes with each additional service provided.

iii) The timing of payments may be:
   • in advance (prospective payment that provides a fixed overall budget), or;
   • after the behaviour has taken place, including:
     o retrospective payment where there is no overall limit on payments, or
     o retrospective payments where there is a cap on the total payments that can be made. Once the cap is reached, either no further payments are made or the amount of the unit payment is reduced.

2. LEVEL OF PAYMENT
The level of payment may be fixed in advance and subject to negotiation (e.g. in a fee schedule or trade union bargaining of salary increases), or

providers may have complete (in a market setting) or partial (e.g. fee/price controls) discretion as to the amount of money they can charge for their services.

The amount of the payment may be reduced or withheld if behaviour does not comply with what is required (i.e. a financial penalty).

The amount may also vary depending on administrative rules based on the characteristics of the provider (e.g. whether accredited or qualified to a certain standard or seniority) or the patients seen (e.g. 'risk adjustment' or more complex/costly services or patients receive higher payments).

Changes in any of the above will influence the incentive effects of the payment, i.e. whether it encourages or discourages the desired behavior. Changes in behaviour in turn influence the cost and quality of care provided.6

---

6 Changes in payment may also influence quality and cost by influencing recruitment and retention, i.e influencing the mix of providers, and their self-selection into the payment scheme. Changes in the mix of providers may then impact on
DEFINING QUALITY OF CARE
The term ‘quality’ can be very broad such as that used by the Institute of Medicine: “the degree to which health care services for individuals and populations increase the likelihood of desired outcomes and are consistent with current professional knowledge” (Lohr 1990). Their report in 2001, ‘Crossing the Quality Chasm’, identified six aims for health care that should guide quality improvement efforts: safety, effectiveness, patient-centeredness, timeliness, efficiency, and equity (clinical quality, patient centred care, and efficiency). The popular definition used by Donabedian (1966) includes aspects of health care related to structures (e.g. IT systems or disease registers), processes (e.g. recording of blood pressure, or the existence of quality improvement programs), or outcomes (e.g. clinical indicators or the health status of patients) Campbell et al (2000) provides a review of several definitions of health care quality and suggests that Donabedian’s three dimensions may also be split into ‘access’ and ‘effectiveness’. Economists also use the term ‘utility’ to refer to the well-being of individuals which, although still a broad concept, excludes measures of structure and process of care provided, unless they influence patients’ utility (Torrance 1986). Utility therefore focuses on what is important to consumers/patients.

There may also be different types of quality indicators used for different levels of the health care system. The quality of national and regional health systems are more likely to include population-based measures of access and uptake of preventive services, and overall health of the population including mortality rates. However, monitoring the quality of hospitals is more likely to include re-admission rates, waiting times, and minimum standards of safety, such as infection rates. Individual physician measures of quality are more likely to include evidence-based clinical and physiological indicators of quality (e.g. HbA1c, blood pressure), clinical behaviours (e.g. prescribing and test ordering), or measures of patient’s health outcomes and quality of life, such as mortality or patient reported outcome measures.

In this review we focus on defining quality of the care provided by the physician and so include clinical and physiological measures, clinical behaviours, and patient reported outcomes and experiences. Studies must demonstrate there is a link between their primary outcome measures and patients’ well-being, either through direct measurement of patient well-being or through reference to current best practice, such as previous studies, existing clinical guidelines, or systematic reviews of the literature. The term patient ‘well-being’ is used to capture the wide range of patient reported outcomes that can potentially be used, including not only health outcomes, such as health-related quality of life, but also measures of patient experience with care provided. Our definition of quality of care is therefore, “the degree to which changes in physician behaviour improves the well-being of patients”

OBJECTIVES
The aim of this review is to examine the effect of changes in financial incentives on the quality of care provided by Primary Care Physicians (PCPs) and to identify:

• the different types of financial incentives that have improved quality.
• the characteristics of patient populations for whom quality of care has been improved by financial incentives.
• the characteristics of PCPs who have responded to financial incentives.

DIFFERENCES WITH PREVIOUS COCHRANE REVIEWS
Our review is broader than the two previous Cochrane reviews in this area with respect to the type of financial incentives, but also narrower with respect to the outcome measures being used (Gosden et al. 2000, Giuffrida et al. 1999). Both existing reviews examine primary care physicians, and we retain this focus.

Type of financial incentives. Giuffrida et al (1999) focuses only on target payments, which are when payments are made when a specific level of activity or quality is reached. This is a specific type of non-linear performance pay. Gosden et al (2000) consider interventions that change the type of payment between salary, capitation and fee-for-service (FFS). Our review includes all of these types of payments, in addition to changes in the level or amount of payment, pay for performance schemes, changes to who is paid (e.g. the physician or team), in addition to multifaceted interventions that include a mix of the above with a non-financial intervention (e.g. education).

Outcome measures used. Another difference with these reviews is the type of outcomes considered. In Giuffrida et al (1999) studies were included if they reported objective measurement of patient outcomes, health services utilization, health care costs, equity of care and PCP satisfaction with working environment. Gosden et al (2000) report similar outcomes as studies were included only if they reported objective measures of: health professional outcomes; health professional process, health service utilization; health care costs; and patient outcomes. Our review defines quality of care in terms of clinical behaviours, clinical and physiological measures, and patient reported outcome measures and experiences. It excludes health professional processes and outcomes, and health care costs.

CRITERIA FOR INCLUDING STUDIES IN THIS REVIEW
Four criteria will be used to select studies. Study design, types of participants, types of intervention, and types of outcome measures used.

STUDY DESIGN
We will include studies meeting the EPOC quality criteria (Grimshaw et al. 2003, Cochrane Effective Practice and Organisation of Care Group 2009) including all languages and unpublished studies. We will include:

1. Randomized, controlled trials (RCTs)
2. Quasi-randomized controlled trials
3. Controlled before and after (CBA) studies with at least two intervention sites and two control sites (EPOC 2009)
4. Interrupted time series (ITS) with a clearly defined point in time where the intervention occurred and at least three data points before and after the intervention.

TYPES OF PARTICIPANTS
Primary Care Physicians (PCPs). PCPs are defined as doctors holding a medical degree. Primary care physicians include general practitioners, family doctors, family physicians, family practitioners and other generalist physicians working in primary health care settings who fulfill
primary health care tasks (Gosden et al. 2000). PCPs working in all sectors (public, private, and private not-for-profit) and from all countries will be included in the review.

Primary Care Teams. These are defined as teams or groups of health professionals that include a PCP.

TYPES OF INTERVENTIONS
We will include studies that examine the following financial interventions.

   i) Where the intervention changes the amount of payment (dose-response)

   ii) Where the intervention changes the method of payment:
         o payment per unit of time (salary/sessional payment)
         o payment for each service/visit/treatment/episode provided (fee-for-service)
         o payment for each patient enrolled or registered with the PCP (capitation)
         o payment for improvements in ‘quality’ (performance pay)

   iii) Where the intervention changes who is paid (e.g. from an individual to a group or team)

   iv) Multifaceted interventions that include at least two of the above

   v) Multifaceted interventions that include at least one of the above plus and at least one other type of non-financial intervention designed to change professional behaviour (e.g. education, feedback, audit).

Note: Target payments may be included under (ii) to (iv) depending on the behaviour that triggers the payment. Target payments are simply a non-linear version of these payment schemes.

TYPES OF OUTCOME MEASURES
Primary outcome measures include the quality of care provided by PCPs that are related to patients’ health and well-being. This includes:

Patient reported outcome measures:
   i) self-reported measures of health status and health-related quality of life (e.g. SF-36, EQ5D).
   ii) self-reported measures of patient satisfaction and experience with the process of care
   iii) Clinical behaviours, such as prescribing, test-ordering, referrals, treatments or advice provided.
   iv) Intermediate clinical and physiological indicators (e.g. HbA1c, blood pressure, cholesterol)

For (ii) and (iii) the study authors’ should state that the change in behaviour or indicator is related to changes in the quality of care. This claim should be justified with reference to
evidence of a positive effect of the behaviour or indicator on health and well-being, such as references to evidence-based clinical guidelines or systematic reviews. If a statement about improvements in quality of care is not made in the paper, or is made but not justified through references, then

- the authors will be contacted to establish whether any evidence existed at the time the study took place, and;
- the review authors will search for Cochrane reviews or national clinical guidelines that existed at the time the study was conducted.

In addition to (i) to (iv), if a study reports other primary outcomes that are not measures of quality of care (e.g. costs, health professional satisfaction or experiences, and clinical behaviours where there is no evidence of a link to quality of care, e.g. number of services provided or number of patients seen), then these will reported and summarised in separate tables, but not synthesised. This is important in helping to identify unintentional effects of the intervention.

SEARCH STRATEGY FOR IDENTIFICATION OF STUDIES

The searches will be conducted with the advice and assistance of the Effective Practice and Organisation of Care (EPOC) Group.

1) The EPOC Register will be used in the first instance and contains monthly searches for studies in EPOC’s broad area of interest. These retrospective searches include MEDLINE, HealthSTAR, EMBASE and CINAHL.

2) Specific searches of MEDLINE, HealthSTAR, EMBASE, CINAHL, PsychLIT, ECONLIT, the Cochrane Controlled Trials Register (CCTR), and Cochrane Data base of Systematic Reviews (CDSR). Both free text and formal search terms will be used.

3) Searches of internet-based Economics and Health Economics Working Paper collections, including RePEc (Research Papers in Economics), and the Social Science Research Network (ERN).

4) Studies identified from previous research of the review authors and personal contacts in the area.

5) The reference lists of retrieved articles.

6) Direct contact with key authors in the field.

7) Websites of key organisations. UK- National Primary Care Research & Development Centre, NHS Service Delivery and Organisation R & D Programme, NHS Centre for Reviews and Dissemination; USA- Commonwealth Fund, Robert Graham Centre; Europe - European
Observatory on Health Systems and Policy; Canada - Canadian Health Services Research Foundation; Australia - Primary Health Care Research Information Service (PHCRIS).

8) Grey literature identified through key informants and policy contacts. International and Australian Government policy documents, commissioned reports; position papers / policy statements of professional bodies or associations.

9) Handsearching may be required, and will be focused on any journals identified from the above search strategies which appear to have a high yield of relevant studies.

The searches will be confined to the period 2000 current. We will search MEDLINE using the search strategy (outlined below), which has been modified from an overview of reviews of financial incentives being conducted by Prof. Martin Eccles, in addition to the search terms used by Gosden et al (2000) and Giuffrida et al (1999). This will be adapted to other databases with the assistance of EPOC.

**MEDLINE strategy:**

exp Primary health care
(primary adj2 care).ti,ab.
exp Physicians/
Physicians, Family
((community or family or general) adj2 (doctor? or physician? or practitioner?)).ti,ab.
Family Practice/
professional practice/ or exp group practice/ or house calls/ or institutional practice/ or office management/ or office visits/ or exp partnership practice/ or exp practice management/ or exp private practice/exp Private Practice
(practice? adj2 (community or family or general or group or institutional or partner? or partnership? or physician? or private or solo)).ti,ab.
("house call?" or (office adj2 (visit? or visiting))).ti,ab.
physicians/ or foreign medical graduates/ or physicians, women/
Physician's Practice Patterns/ or ((physician? or practitioner?) adj2 "practice pattern?").ti,ab.
or/1-11
exp Reimbursement mechanisms
Insurance, Health, Reimbursement
exp Fees and Charges
Income
Physician Incentive Plans
Salaries and Fringe Benefits
exp Physician's Practice Patterns
((reimburse* or payment or remunerat* or incentive* or (financ* adj3 penalt*) or financial or salar* or fee or fees or capitait* or (pay adj3 perform*) or (payment adj3 perform*) or (pay
adj3 reduc* or (payment adj3 reduc*) or (pay adj3 penalt*) or (payment adj3 penalt*) or payor or payer or payee or prepaid or “pre-paid” or prepay$ or “pre-pay$“).ti,ab.
or/13-19
12 and 21

METHODS OF THE REVIEW

SELECTION OF STUDIES
Two reviewers (LW and DA) will read all the abstracts identified from the electronic search to identify publications that meet the initial inclusion criteria of the types of interventions, studies, participants and outcome measures defined above. Where it is unclear from the abstract whether a study meets the inclusion criteria, the reviewers will retrieve and read the full paper. Disagreements will be resolved by discussion between the two reviewers, with further discussion with PS and AS if necessary. The other review authors are experts in the field of primary care organisation (LN) and general medical practice (JF and DY) and will be used to adjudicate inclusion, especially in relation to the studies claims to have measured quality of care.

DATA EXTRACTION
Two reviewers (LW and DA) will independently extract appropriate information from each included study using a modified version of the EPOC data collection checklist. This will include information on the study design, intervention, controls, participants, setting, methods, outcomes studied and results. We will contact study authors if possible to obtain missing information about relevant studies.

ASSESSMENT OF RISK OF BIAS
Two reviewers will assess the validity of each included study according to the EPOC risk of bias guideline. (EPOC 2009) and the methodology developed in Chapter 8.5 of the Cochrane Handbook for Systematic Reviews of Interventions (Higgins and Green 2008). We will present our findings in a ‘Risk of bias table’ and will use graphs and figures to summarise our assessments across studies.

ANALYSIS
On the basis of the two previous Cochrane reviews in this area, which uncovered only a handful of studies set in different contexts and with different outcome measures, we anticipate that combining the data from different studies is likely to be infeasible. We will summarise the results of included studies with the estimated effect of an intervention and confidence intervals in a table where possible. Results will be presented grouped by intervention type, disease area, and country.

MEASURES OF INTERVENTION EFFECT AND UNIT OF ANALYSIS
ERROR
For RCTs, Cluster-RCTs and CBA studies we will report the baseline and the intervention differences between study and control groups, in natural units. Where possible we will calculate absolute differences and relative percentage differences. For dichotomous and continuous outcomes appropriate summary statistics will be calculated. Statistical meta-analysis will be
conducted if possible and results displayed graphically to assess for heterogeneity. For ITS studies we will report changes in level and in slope as well as relative percentage changes at 1 month, 6 months, 12 months and yearly thereafter.

For Cluster-RCTs we will identify that the clustering has been taken into account and that the unit of allocation and the unit of analysis are the same. For studies presenting a unit of analysis error we will attempt to reanalyze the data appropriately either by extracting the necessary data from the articles or by obtaining these data from the study investigator. If this is not possible we will report only point estimates and no confidence intervals or p-values.

If there appears to be a sufficient number of studies amenable to meta-analysis we will follow the approach developed in the Cochrane Handbook (chapter 16.3) to combine studies with unit of analysis error. We will use external estimates of the intracluster correlation coefficient (ICC), obtained from similar studies, to reduce the concerned trials to their effective sample size.

For studies using instrumental variables within the four study design types, we will summarise the authors’ theoretical justification for the choice of instrumental variables, and summarise the statistical tests used to examine the strength and validity of instruments (Bound et al. 1995, D'Agostino Jr and D'Agostino Sr 2007).

**INVESTIGATION OF HETEROGENEITY**

If there are studies that are similar enough that it would be meaningful to combine their results we will investigate how the following factors might explain the differences in the impacts of financial incentives:

- Characteristics of the financial incentive (e.g. amount of payment, method of payment)
- Characteristics of the patient population (e.g. type of disease, socio-economic status, age, gender)
- Characteristics of the PCP (e.g. age, gender, solo/group practice)

We will assess the statistical heterogeneity and inconsistency across studies both by inspecting graphically the confidence intervals for the results of individual studies and by calculating appropriate statistics (Chi-squared test, I-squared test). We will report the results of these tests and explore the causes of heterogeneity by conducting subgroup analysis or meta-regression according to the methodology recommended in the Cochrane Handbook (sections 9.6.3 - 9.6.4, 2008).
APPENDIX 2. LIST OF KEY INFORMANTS

1. Dr. Tony Hobbs, Chair, Expert Reference Group for the National Primary Health Care Strategy

2. Department of Health and Ageing
   - Ms. Megan Morris, First Assistant Secretary, Primary and Ambulatory Care Division, DoHA (Megan.Morris@Health.gov.au)
   - David Braggett (Director, Divisions Funding and Performance section)
   - Jacinta Holdway (A/g Director, Practice Incentives section)


4. Prof. Justin Beilby, Commissioner, National Health and Hospitals Reform Commission (justin.beilby@adelaide.edu.au)

5. Prof. Mark Harris, UNSW, Member of Expert Reference Group of National Primary Health Care Strategy

6. Prof. Hal Sverrisen, LaTrobe University, Member of Expert Reference Group of National Primary Health Care Strategy (H.Swerissen@latrobe.edu.au)

7. Australian General Practice Network
   - Mr. David Butt, CEO AGPN
   - Ms. Leanne Wells, Executive Director, Policy and Business Development
   - Ms. Rachel Yates, Director Policy.

8. Russell McGowan, Consumer Health Forum

9. Janet Laverick, Director, Primary Health Care Branch, Department of Human Services, Victoria.

10. Royal Australian College of General Practitioners
    - Lauren Cordwell, Senior Adviser (Health Systems Reform & Safety)
    - Dr. Greg Wilson, National Policy Advisor
APPENDIX 3. FLOWCHART OF STUDY SELECTION FOR NARRATIVE & EPOC REVIEW

Initial database search

2616 articles
Databases searched:
- CINAHL (342)
- EPOCRegister (26)
- Cochrane (41)
- Medline (834)
- EconLit (75)
- PAIS (14)
- EMBASE (1175)
- PsychInfo (109)

2134
(482 articles published prior to 2000 deleted)

162
(1972 articles culled at title stage)

Pre-identified literature + reference list search

317 articles
Identified through:
- Professional & academic experts
- Reference lists of systematic reviews & corresponding articles
- Advisory committee
  (selection based on title)

186
(131 articles published prior to 2000 deleted)

160 empirical
126 non-empirical

348
(database articles + pre-identified literature)

298
(50 duplicates detected & deleted)

34
(263 articles culled at abstract stage)

6 articles included
28 articles excluded
APPENDIX 4. REVIEW OF COUNTRY-SPECIFIC INCENTIVE SCHEMES

Our narrative review and synthesis of the literature is based on a conceptual framework that identified a number of issues that structured the review for each country. Table 1 shows the questions that were used. For each country, the literature was used to answer these questions.

UNITED STATES

BACKGROUND

Primary care physicians in the US include family physicians, internists, paediatricians, obstetricians and other specialists, who may also refer to themselves as primary care physicians. There is no national requirement for gatekeeping although this occurs in many HMOs. Family physicians have traditionally played a minor role in the health system relative to specialists.

There is considerable variation and diversity in the role of primary care physicians. Some are largely fee-for-service, and others may contract with a single or a range of managed care or Health Maintenance Organisations or Federal and State programs such as Medicare and Medicaid. The payment arrangements of these contracts can vary. Some PCPs may be salaried and employed by HMOs, and others may be paid capitation or bonus payments. The diversity of different payment schemes, including the use of pay for performance, means there are hundreds of examples of such schemes and a large literature discussing and evaluating their effects.

Given this diversity, we answer the key questions of the narrative review from three perspectives.

- The findings and views of a number of previous reviews of P4P in the US are summarised. This helps to document the general trends in the US and how schemes have evolved, including the more recent proposals centred around the medical home.

- The Quality Improvement Program (QIP) of the Integrated Healthcare Association (IHA) in California, one of the largest private schemes in the US.

- Medicare’s Practice Group Demonstration Program.

OVERVIEW FROM PREVIOUS REVIEWS OF THE US LITERATURE

What contextual factors are relevant in each incentive scheme?

“Three recent trends—the collapse of managed care, the evidence of substantial quality problems, and the return of costs rising well in excess of economic growth—have led to a new round of innovations by payers.” (Galvin 2006)

Since 2000 the emphasis has shifted from the role of managed care with capitation and consumer driven approaches such as the public reporting of performance, to a more direct emphasis on quality. Advances in the measurement of quality have also played a role in the growth of P4P schemes, and rising health costs have increased demands for accountability in the health system (Fisher 2006, Institute of Medicine 2007, Thomas 2007).
The report by the Institute of Medicine, Crossing the Quality Chasm (Institute of Medicine 2007) was an important part of this change in emphasis. The report, and subsequent reports by the IOM including one on pay for performance (Institute of Medicine 2007) was mandated by the US Congress. A key issue is whether quality improvement programs (e.g. feedback and continuous quality improvement) and public reporting of performance information by themselves are sufficient to stimulate behaviour change amongst enough physicians, and whether they need to be supplemented by financial incentives.

There are a large number of pay for performance schemes in the US that have emerged from 2000 onwards, and around 160 schemes in 2007 (Baker and Delbanco 2007). Initially most common in private settings (HMOs, PPOs and IPAs), Medicare and Medicaid have introduced pilot demonstration projects of their own (Gilmore et al. 2007, Mullen et al. 2009, Young et al. 2007b, Kautter et al. 2007).

Rosenthal (2004) surveyed 31 private HMO plans with P4P and followed them up in (Rosenthal et al. 2007) to examine their evolution. Most were still in existence three years later, and had expanded to include specialist groups as well as PCPs. The coverage of performance measures had also increased, with cost and efficiency and IT measures more likely to be included, as well as an increase in intermediate outcome measures such as HBA1c, cholesterol, or blood pressure control. Some quality measures were dropped because their levels were consistently high or there was little variation across providers (e.g. screening and vaccinations). Incentives remained a relatively small proportion (<5%) of PCPs total income, although some plans had intentions to increase this. A higher proportion are now rewarding quality improvement, and also using risk-adjustment to adjust payments depending on the characteristics of the population served. Many of the quality measures used are from the the Healthcare Effectiveness Data and Information Set (HEDIS), developed by the National Committee for Quality Assurance (NCQA) and are the accepted standard in quality measurement.

Rosenthal et al (2007) found that respondents indicated three main challenges to making P4P more successful: improved engagement of physicians; determining the size of the incentive pool to capture PCPs attention, and; finding the resources to continue funding the programs. Starting small and building trust and capacity over time was also one lesson learned. Using nationally accepted/recommended performance measures based on clinical rather than administrative data was also deemed important in scheme’s success. Other respondents noted the absence of a clear return on investment in terms of net savings as an issue for future sustainability of the programs.

**What were the key characteristics of the incentive scheme?**


The first survey in 2004 emphasised the role of the insurer having a large market share, thus giving it ‘leverage’ to influence PCP behaviour. Physicians may contract with many health plans, and each one by itself may have little effect on behaviour. This leads to complexity for each provider who may find it difficult to monitor specific quality targets (Bokhour et al. 2006). This is a general issue about the effectiveness of payment schemes in the US, and that they will less effective, the lower the market penetration of the purchaser (Rosenthal et al. 2004, Christianson et al. 2006). Incentives introduced by one funder will be less effective if there are multiple funders of PCPs. Multiple funders may operate together, such as the IHA program, to strengthen their leverage and overall market share. In 2007, those programs which self-reported a positive impact also had higher market shares and higher rewards as a percentage of PCP pay (Rosenthal et al. 2007).

A more representative survey of 252 HMOs in 2006 found that ninety percent of schemes targeted physicians rather than hospitals (Rosenthal et al. 2006). The plans mainly involved bonuses or penalties, and some only rewarded the top-ranked physicians or physician groups, and so relative performance was used as the basis of rewards/penalties. Those with low
baseline performance may find it less costly to improve quality, than those with a high baseline performance, especially if there are diminishing returns to effort. Two-thirds offered rewards for the maintenance of a performance threshold, and 20% rewarded quality improvements. Many HMOs offered rewards which were around 5% of total payments.

How incentive money was distributed within PCP groups was also important (Bokhour et al. 2006). Equal sharing, regardless of individual performance, relied on good data systems, and alternatives to equal sharing could be divisive. Equal sharing encouraged working together and so could enhance quality in other ways. However, it was more common for money to be distributed according to relative performance, which is a more high powered incentive. However, this is more problematic when a patient may have more than one physician or where physicians have small groups of patients. Physician groups have discretion on how to use the payments. This may reduce the power of the specific incentive, but enable its use for more locally-relevant quality improvement (e.g. hiring additional staff). This avoids divisiveness and attribution problems – it creates cohesiveness rather than competition. It may also lead to spillover effects on quality of care in other areas not related to the incentive and at the discretion of the physician group (Forrest et al. 2006). Medical group payment may also lead to increased shared accountability for a population and cooperation between members of the group (Forrest et al. 2006). A mix of approaches to the distribution and use of rewards within medical groups, such as retention of a percentage by the group and distribution of reminder according to performance, may dilute the more extreme incentives of sharing of rewards and risk.

Characteristics of patients/population

Rosenthal et al (2004) uses the broad structure, process and outcome framework to classify quality. In 2006, over 80% of plans focused on clinical quality, in addition to patient satisfaction and IT. Diabetes, mammography, asthma, were the most common targets of these schemes, followed by blood pressure, antidepressant medication and cholesterol management (Rosenthal et al. 2006). HMOs were most likely to use pay for performance where enrollees were required to select a primary care physician as a gatekeeper, and where capitation payment was already used. This suggests that having a stable population where performance can be attributed to a single provider or medical group may be a key factor influencing the success of these schemes.

The accuracy, validity and source of data were also important in the success of such schemes. Claims data was more likely to be mistrusted or inaccurate (Bokhour et al. 2006). National measures, although standardized, did not always feel relevant to a physician’s clinical practice.

Characteristics of PCPs in the schemes

For plans which targeted physicians, over half targeted medical groups, and only 13% targeted individual physicians (Rosenthal et al. 2006). Two-thirds involved a primary care provider as a gatekeeper.

What were the objectives and intentions of the scheme and were they achieved?

In the review by Rosenthal et al (2004), most incentive programs rewarded achievement of a certain level of quality, and hardly any rewarded quality improvement, though this had changed by 2007. They argued that physicians who are already good performers have an advantage and are being reward for historical improvements:

“This rewarding of historical investments in quality, however, strikes us as not altogether consistent with the stated goals of most of the programs: to improve quality for all enrollees/beneficiaries.”
To the extent that physicians are already motivated to provide high quality of care, then the role of incentives in improving quality is limited (Bokhour et al. 2006, Mooney and Ryan 1993). Rosenthal et al (2005) found that physician groups at baseline who were already meeting the target threshold improved their quality the least, but received the majority of the incentive funding. Monetary transfers were therefore seen by these physicians as reward for good work done, therefore re-inforcing intrinsic motivation, rather than being regarded as a reason to further change their behaviour. Physicians seemed more responsive to feedback on quality, than to the monetary rewards themselves (Bokhour et al. 2006). May feel they were being unfairly punished for lack of compliance of patients, if quality targets depend on patient behaviours as well as their own (Bokhour et al. 2006).

Even where quality was not improving, Rosenthal et al (2007) found that three motivations were responsible for keeping the programs in existence. The first was that the program had not yet found the ‘right’ type of incentive scheme, and so believed that the the ‘right’ program could still work. The second was that paying more for higher quality is fairer than paying for quantity provided, even if quality didn’t improve (Bokhour et al. 2006). Finally, P4P was regarded as an agent for change and a step toward other organisational goals, such as transparency to consumers (Bokhour et al. 2006).

Physicians’ mistrust of schemes is based on a different interpretation of the goals of P4P schemes than those implementing them. Physicians fear that they will be used to encourage efficiency rather than only higher quality (Fisher 2006). More recent developments have begun to include efficiency measures in the schemes, in addition to quality.

Self-reported unanticipated effects were patient dumping and collapse of the scheme due to disagreements about how performance bonuses were distributed (Rosenthal et al. 2007).

What incentive approaches were used for whom, for which disease areas, and why?

A number of reasons for physician non-participation, include: i) mistrust of the objectives of quality measurement; ii) quantitative measures being inappropriate to medical practice which requires a careful balance of risks and benefits, patient preference and medical knowledge (Fisher 2006); iii) careful attribution, and iv) small sample sizes for some physicians/physician groups. There are also high costs of collecting the data, doubts about whether the size of rewards are high enough to cover costs and motivate behaviour change, especially amongst those with a low baseline performance. Physician acceptance and early ‘buy in’ of these schemes is particularly important (Casalino et al. 2007, Forrest et al. 2006). There is also evidence that the size of reward is associated with participation of physicians (Mullen et al. 2009, de Brantes and D'Andrea 2009).

What contextual factors are relevant in each incentive scheme?

A general consumer backlash against managed care and financial incentives for utilisation, lead to a number of health plans introducing report cards and P4P schemes during the 1990s in California. With considerable diversity amongst these schemes and multiple payers, there was confusion amongst payers, physicians and consumers (Integrated Healthcare Association 2006). The Californian Association of Physician Groups, the Californian Association of Health Plans, and the Pacific Business Group on Health (employers), were the main players, and physician groups representatives approached IHA (established in 1996) to establish a uniform program for quality improvement.
incentives, although some mistrust developed in the physician groups as the scheme developed and this almost stopped the scheme. Initially the funding source from purchasers was an issue, but incentives were funded through some premium increases and other efficiencies within the plans, without lowering usual reimbursement. A clear set of guiding principles helped steer the program. A steering and technical committee was established in 2001 to develop the indicators (Integrated Healthcare Association 2006).

What were the key characteristics of the incentive scheme?

In July 2003 PacifiCare began paying quarterly bonuses of $0.23 per PacifiCare member per month for each performance target that was met or exceeded (at or above the 75th percentile of 2002 performance) for ten common measures agreed upon by IHA as well as six measures of quality and patient safety for the hospital to which the group admits the majority of its patients. From 2004, PacifiCare became one of seven participating health plans in the non-profit Integrated Healthcare Association (IHA) initiative in California, comprising around 60% of the total revenue of physician group members. This is the largest private grouping in the US, and in 2009 covered 228 medical groups, 40,000 physicians, and 12 million members. As other health plans joined the IHA scheme, they offered a different structure of incentives, such as having threshold payments at 20th and 30th percentiles (Mullen et al. 2009). By 2006 only PacificCare used absolute thresholds, whilst the rest used relative rankings of providers, with payments on a sliding scale between lower (e.g. 50th) and upper (e.g. 100th) percentiles.

The total payment for the average physician group was around 5% of capitation payments to the group, or about 0.8% of total revenue. Performance was assessed using a rolling year of data with a 6 month lag, so changes made early in the year may not be rewarded for up to 18 months. When the other plans joined the IHA scheme, this increased the potential size of the average bonus by a factor of 10 (Mullen et al. 2009).

The ten IHA domains (for 2003) are cervical cancer screening, mammography, childhood immunizations, HbA1c, LDL cholesterol, patient satisfaction, satisfaction with primary care physician (PCP), satisfaction with referral process, satisfaction with specialist, and effective PCP communication. In 2004, these measures changed slightly, and a third tier threshold of 85% was introduced, doubling payments for the highest performers, and care co-ordination was also added. There were further changes in 2005. In 2007 risk-adjusted indicators of cost and resource use were also added to the scheme (www.iha.org). The three main domains: clinical, patient experience and IT, were weighted by 50%, 30% and 20% respectively in 2006, with these weightings constantly reviewed. Two plans don’t pay for IT and only three plans use the recommended weights. These are based on electronic data from administrative records, rather than rely on reviewing patient charts.

Each physician group was required to have a minimum number of enrollees, of which 163 out of 300 medical groups met these criteria. In addition to the incentives, physicians were also receiving feedback (a report card) on their performance, and this continued in the control group (Rosenthal 2005). All physician groups had been receiving and collecting performance data since 1993, and this was made public in 1998, so IT infrastructure was already in place for HMO utilisation review and report cards showing relative rankings of physician groups for some quality indicators.

What were the objectives and intentions of the scheme and were they achieved?

A set of guiding principles were developed to help shape the IHA program (Integrated Healthcare Association 2006). Their overriding goal is:

“to create a business case for quality improvement through a compelling set of incentives that would drive breakthrough improvements in the quality and experience of health care”.

48
There is a clear emphasis on quality improvement, and this was further emphasized in changes to the original mission and priorities from 2006 onwards. An early evaluation of the PacificCare Scheme found that only cervical screening indicator improved by 10% (Rosenthal et al. 2005). Mullen et al (2009) extended this evaluation to include the larger scale scheme operated by IHA and including PacifiCare and found that the positive effect for cervical screening continued, and that a previous decline in chlamydia screening was reversed. Appropriate asthma medication and preferred antibiotic usage fell, and there no effects on other measures such as childhood immunization, breast cancer screening, cholesterol testing, or HbA1c testing. They found no evidence of spillover effects into unremunerated clinical areas.

**What incentive approaches were used for whom, for which disease areas, and why?**

Physician groups which were at the target levels of performance at baseline were the least likely to improve, but received the highest financial rewards per group. Those which had the lowest level of performance at baseline improved the most, but received the lowest average rewards (Rosenthal et al. 2005). Mullen et al (2009) found that some clinical areas increased and some decreased, suggesting important interactions between the relative rewards for different disease areas. They highlighted economies of scope in production for those disease areas which required recall registers and frequent visits (i.e., diabetes, asthma, screening, immunisation), although found little evidence of similar quality improvements across all disease area with common production processes.

**PAY FOR PERFORMANCE UNDER MEDICARE**

There are currently two Medicare demonstration projects, one is for large group practices, the Medicare Physician Group Practice (PGP) Demonstration Program, and the other is designed for small and solo practices, the Medicare Care Management Program (MCMP).

The former has been operating since 2005 with the demonstration due to end in 2010. There have been three evaluation reports to US Congress that include a detailed description of the program and evaluation of changes in quality scores and expenditures compared to non PGP Medicare beneficiaries. The MCMP has been operating since 2007 and is scheduled to end in 2010. The first year included payments for providing data on quality, whilst the second year (2008-9) introduced the first incentive payments. An evaluation is underway but is not yet completed. The discussion below will focus largely on the PGP demonstration where more data are available, but the key features of the MCMP will also be summarised.

A third development in very early stages is the concept of the medical home. Recommended to Congress in 2008 by the Medicare Payment Review Commission, this is for patients with chronic disease and where practices would receive case management fees for caring for that patient (Medicare Payment Advisory Commission 2008). The recommendations did not go as far as recommending pay for performance, although earlier commentaries have suggested that this is desirable (Iglehart 2005). A Medicare Demonstration Program has been proposed and designed (Maxfield et al. 2008) that includes the care management fee only, adjusted by severity. However, this is now to be integrated into an 'Advanced Primary Care' demonstration in collaboration with private insurers and State-based pilots who have already set up pilots based around the concept of the medical home, with the Medicare demonstration beginning in 2010. [http://healthreform.gov/newsroom/factsheet/medicalhomes.html](http://healthreform.gov/newsroom/factsheet/medicalhomes.html)
What contextual factors are relevant to each incentive scheme?

The growth in private HMO-based pay for performance schemes, coupled with the failure’s of managed care schemes to control costs, and a growing concern about quality, led to Medicare introducing a pay for performance scheme in 2005. This was aimed at physician groups and started as a demonstration program in 2005. “The PGP Demonstration seeks to align incentives for physician groups to manage the overall care of their patients, especially beneficiaries with chronic illnesses and high-risk patients who account for a significant portion of Medicare expenditures. The demonstration provides a financial incentive similar to those used by managed care organizations and other commercial payers to reward quality improvement and encourage efficiency” (Leavitt 2006).

What were the key characteristics of the incentive scheme?

The design of the PGP demonstration is outlined in Thomas (2007), (Kautter et al. 2007), and (Cromwell et al. 2007). Leavitt (2006) outlines the details of the scheme. Ten physician groups took part in the 3-year demonstration which has now been extended to 5 years until 2010. This includes 5,000 physicians and 220,000 Medicare fee-for-service patients. The physician groups are large (at least 232 physicians per group) and diverse. Physician groups were paid fee-for-service and so are rewarded for improving care for a non-enrolled population. Annual bonus payments cover both lower than average growth in costs, and meeting thresholds for up to 32 ambulatory care quality measures covering five conditions; diabetes, heart failure, coronary artery disease, hypertension and preventive care. The participating physician groups have used the payments to re-invest in practice services, rather than reward individual physicians.

The incentive payment is made up of a ‘cost performance payment’ and a ‘quality performance payment’, with an increasing proportion of the total payment determined by the quality performance payment over the 3 years of the demonstration project. The cost performance payment is based on the PGP keeping expenditures for its beneficiaries below 2% of its target. These ‘savings’, then form the incentive payment. 80% of these savings are returnable to the PGP. Any losses can be carried forward to the next year and PPGs are not liable for these losses.

The quality performance payment is made if all quality targets are met in the previous year. If only some are met, then a proportion is retained by Medicare. The quality measures were derived from the Doctors’ Office Quality Project that provided validated measures. The number of indicators increased over the period, and by year 2 27 indicators were being used (out of a possible 32). Each of indicator was assigned a number of points. Points earned each year are divided by total points possible to arrive at the percentage of quality targets met. For each quality measure, a PGP is required to meet two threshold targets and one improvement target. The first threshold target is 75% compliance, and the second is the achievement of the 70th percentile of Medicare HEDIS measures. The improvement target is a 10% or greater reduction in the gap between the baseline and 100% compliance. It was not apparent from Leavitt (2006) how the size of the quality performance payment pool is determined. This pool provides an effective expenditure cap on the scheme, though its size will determine the strength of the incentives. The final performance payment cannot be more than 5% of the PGP’s target expenditures. Finally, Medicare withholds 25% of the payment until the end of the 3 year demonstration.

The MCMP demonstration is occurring across four states and includes almost 700 small to medium size (up to 10 Physicians) practices, with over 500 practices currently participating, with an average practice size of 3.3 physicians (Wilkin et al. 2007). In the first year practices were rewarded for reporting data on 26 clinical quality measures to Medicare for diabetes, heart failure, and coronary heart disease, immunizations, blood pressure screening, and cancer screening. The second year, payments were made for attaining thresholds similar to the PGP demonstration, ie 75% compliance or the top quartile for Medicare HEDIS measures. Based on the percentage of total points earned, practices can earn $70 for each patient in the disease categories, and $25 per patient with chronic disease for the preventive measures. Practices scoring 90% or more will eligible for the full payment per beneficiary. Practices scoring less
than 30% will not earn any incentive payments. The payment will be prorated for those scoring between 30% and 90% in a specific category. Unlike the PGP demonstration, there is no payment for quality improvement, just attaining the thresholds. Up to an additional 25% bonus can be made if an electronic health record is used and reporting the quality data electronically.

What were the objectives and intentions of the scheme and were they achieved?

The PGP demonstration original objectives were to: i) test the use of incentives for health care groups; ii) improve coordination of health care furnished under Medicare Parts A and B; iii) encourage investment in care management infrastructure and processes for efficient service delivery; iv) reward physicians for improving health care processes and outcomes.

In the latest evaluation report to Congress, (Sebelius 2009) found that in year two of the demonstration, all PGPs achieved benchmarks for 25 out of the 27 measures in use. Compared to the base year, in year 2 PGPs improved their quality scores by 9 percentage points for diabetes, 11 percentage points for heart failure, and 5 percentage points for coronary artery disease. In year 3, quality scores had increased by 10 percentage points for diabetes, 11 percentage points for heart failure, 6 percentage points for coronary artery disease, 10 percentage points for for cancer screening and 1 percentage point for hypertension (Centre for Medicare and Medicaid Services 2009b). Seven measures were able to be compared to a comparison group of Medicare beneficiaries in the local area who did not visit a PGP. Four out of seven of these indicators showed greater improvements amongst PGPs than their comparison groups, after adjusting for pre-demonstration trends. However, the 10 PGPs were self-selected and so these results cannot be generalized. Four out of the 10 PGPs achieved the 2% savings target, and so were eligible for the incentive payment. Net savings to Medicare were $2.26m, though this does not take account of the costs to PGPs of implementing the measures, and is relatively small compared to target expenditures.

The objectives of the MCMP demonstration are to: “to promote the adoption and use of health information technology to improve the quality of patient care for chronically ill Medicare patients.” In the first year, 88% of practices were receiving the maximum payments for reporting clinical information to Medicare, with each practice earning an average of $2,500. In year 2, the average payment per practice was $14,000. 25% of practices used the electronic health record to submit data (Centre for Medicare and Medicaid Services 2009a). No formal evaluation or comparison against control groups has yet been reported.

What incentive approaches were successful for whom, for which disease areas, and why?

For the PGP, there were no upfront payments, so any investments required to meet the targets had to be made by PGPs upfront for an uncertain future payoff. This may reduce participation and the effect of the incentives as these costs will vary across practices. 25 of the 32 quality measures require medical record review, and this was more work than anticipated by many PGPs (Leavitt 2006). The four PGPs earning performance payments were different in a number of respects to those which did not earn performance payments, reflecting selection bias. They were either free-standing groups or were affiliated with an academic medical centre; were not affiliated or integrated with a hospital; had lower inpatient and outpatient expenditures than the other six PGPs, and; were more likely to have been making savings before the demonstration.
UNITED KINGDOM

The National Health Service in the UK introduced the Quality and Outcomes Framework (QoF) in 2004, as part of a change to the overall remuneration of GPs.

What contextual factors were important for the QOF?

Prior to April 2004, General Practitioners (GPs) in the UK National Health Service (NHS) were paid for their services almost entirely from a mix of capitation, fees, fixed payments and limited target payments for preventive care. The previous major revision of reimbursement for GPs in 1990 had introduced very limited performance-related pay in the form of target payments for immunizations and cervical screening.

The election of the Labour Government in 1997 lead to large funding increases in the NHS starting 3 years later in 2000 (Maynard and Street 2006). Real NHS spending increased by 7.4% annually between 2002/3 and 2007/8, well over the long-term average of 3% real annual growth (Wanless et al. 2007). The increase in funding was spread across primary and secondary care and was accompanied by extensive reform, including reforms of remuneration systems. A strong performance management framework had evolved over the preceding 10 years in secondary care (Propper et al. 2008), and this was strengthened to attempt to ensure that sufficient improvement accompanied the funding increase. The regime included ‘star ratings for NHS hospital trusts and waiting time targets that were rigorously enforced through incentives providing increased autonomy and funding. This was later followed by the introduction of ‘payment by results’ (Farrar et al. 2009), a prospective activity-related payment system to replace block-contracts for secondary care. In primary care the QOF was seen as a way to increase funding and improve performance: “a something for something deal” (Hutton 2003).

“UK expenditure on primary care will rise from £6.1bn in 2002/03 to £8.0bn in 2005/06, an increase of 33 per cent. About two-thirds of the increased investment will be spent on rewards for higher quality.” (Royal College of General Practitioners 2003)

We expect the large increases in funding which accompanied the QOF were favorable to increases in quality in a capacity-constrained public healthcare system. GPs would be able to spend more time with their patients, by employing more practice nurses and salaried GPs, and could afford better facilities and equipment to deliver quality care. The increased funding in secondary care would also presumably help improve primary care outcomes as patients needing elective surgery would have been seen quicker.

A policy that was evolved in the 10-15 years before the QOF was the introduction of regional primary care organizations to commission (purchase) secondary care from NHS hospitals. Some GP Fundholders evolved into regional ‘Multifunds’, and after GP fundholding was abolished in 1997, ‘Primary Care Groups’ were created in 1999 and progressed to become ‘Primary Care Trusts’ (PCTs) in 2002 as they began commissioning both primary and secondary care services. Some researchers have suggested PCTs have improved services in primary care, developed clinical governance and strengthened the place of primary care in the overall health system (Smith and Mays 2007). It is likely that the introduction of PCTs helped GPs to adapt to the change in funding mechanisms in the NHS in a similar way to how Divisions of General Practice were involved in the implementation of the Practice Incentives Program (PIP) in Australia (Scott and Coote 2007).

It is important to acknowledge that quality of primary care in the UK was improving prior to the introduction of the new policy. Campbell et al (2005) present data on the trends in performance in three chronic disease areas later included in the QOF, from a random sample of English GPs between 1998 and 2003. They show that GP performance, in terms of indicators of quality in coronary heart disease, asthma, and diabetes improved markedly in the five-year period. The authors cite in particular the introduction of new clinical governance initiatives.
including national service frameworks (Department of Health 2000, Department of Health 2001) as policies which may have stimulated this improvement in quality.

The finding that quality indicators similar to those being used in the QOF, were improving markedly before its introduction is key to any attempt to evaluate the QOF. If quality was already improving, was the QOF necessary, and more specifically was it cost effective expenditure by the government? Would the QOF improve rates of quality improvement above those shown by Campbell et al, or could the financial incentives and bureaucracy of the QOF be counterproductive? At the very least, the QOF was criticized for paying doctors for improvements in quality they were making independently of the financial incentive.

An important contextual factor for the QOF is that the reform was not implemented by itself but as part of a new contract for GPs negotiated in 2003 (Department of Health 2003, Royal College of General Practitioners 2003).

The contract emphasized ‘more flexible delivery of services’, ‘developing human resources’ and ‘modernizing infrastructure’ alongside a large increase in funding through the QOF (Royal College of General Practitioners 2003). The new agreement also included a new weighting formula for allocating capitation funds across GP practices.

A part of the new contract which had a significant impact on how GPs worked was the new arrangements for optional ‘Enhanced Services’. These are services which are essentially optional for GP practices and which the local primary care trusts must commission from GPs separate from the overall GMS contract. Crucially, out-of-hours care (evenings and weekends) became an enhanced service and so GPs were controversially allowed to opt out of their previous 24 hour responsibility to provide out-of-hours care (Eaton 2007, Heath 2007). This was interpreted as a significant ‘perk’ for UK GPs. Another perk was inserted into the contract when it emerged that some GPs could ‘lose out’ when the new weighting allocation formula was applied. The government included a ‘minimum practice income guarantee’ clause to prevent capitation funding decreasing from previous levels. It is likely that the QOF being packaged with other perks helped ‘sell’ the reform to GPs.

What were the key characteristics of the incentive scheme?

In April 2004 a new funding scheme for GPs started in which a substantial portion of GP income, around 25%, was to be earned in a pay-for-performance scheme called the Quality and Outcomes Framework (QOF) (Roland 2004). In this scheme, GPs earn points according to their achievement in a number of indicators. The indicators are grouped into three sections: practice organization, clinical care and patient experience. These sections broadly correspond with the structure-process-outcome concept of quality (Donabedian 1966) with the exception that the ‘clinical care’ section includes both process measures (eg measuring blood pressure) and outcome measures (blood pressure managed within certain limits).

The clinical care indicators were initially limited to 10 chronic conditions treated in Primary Care, with the most points available for Coronary Heart Disease, Hypertension, Diabetes and Asthma. The other diseases included were stroke, hypothyroidism, mental disorders, COPD, Epilepsy and Cancer. Points are generally awarded within each clinical area according to the proportion of eligible patients achieving a desired health outcome (eg proportion of patients with heart disease achieving blood pressure below 150/90). Eligible patients are those that are not ‘exception reported’ by the GPs. GPs can exception report patients who refuse care, or for whom additional care is inappropriate (eg because of already receiving the maximum doses of appropriate medication, or having a terminal illness).

Practices are rewarded according to each point earned in the scheme but the amount (in pounds per point) is adjusted for prevalence. This was initially done using an ‘adjusted prevalence factor’ (Guthrie et al. 2006b) which was based on the square root of true prevalence. This was revised in 2009 to be based simply on true prevalence (British Medical Association 2005).
The QOF is not limited to clinical quality indicators and a range of organizational indicators were also included. These were grouped into five categories: Records and information about patients, communicating with patients, education and training, management of medications, and management of the practice. The first category mentioned had at least twice as many points assigned as any of the other categories, indicating the importance of data collection in the QOF.

The total number of points available in 2004 to practices was 1050 of which 550 were for the clinical indicators, 184 for the organizational indicators, 100 points for patient experience, 86 for access and other minor categories (pap tests, maternity, immunizations) and 130 points for performing well across all areas.

In 2005 each point was worth approximately £120, putting the potential value of the scheme at £126,000 per practice per year or £42,000 per GP per year (Roland 2004). The QOF represented a very substantial proportion, 30-50% of earnings (Kmietowicz 2006), for the average GP as well as a substantial expenditure by the government.

The QOF has been reviewed every two years with the first review in 2006 (British Medical Association 2005) when the number of chronic conditions included had risen to 19. The extra conditions included were heart failure, atrial fibrillation, palliative care, dementia, depression, kidney disease, obesity, learning disabilities and smoking. In the 2006 review some of the original indicators were removed or had the number of points reduced, these included many of the organizational indicators which had compensated practices for the ‘set-up’ costs of the QOF.

Another key part of the review process has been the changes in the thresholds (targets) to be achieved for each indicator. All indicators whose previous lowest threshold was 25% had the threshold raised to 40% and all upper thresholds previously under 90% were raised to 90%. These represented relatively modest changes in thresholds compared with the achievement levels of practices (Doran et al. 2006).

A key feature of the development of the QOF was that it was approved by the target population of physicians rather than being opposed by them. Nearly 80 percent of physicians who voted on the reform approved the introduction of the QOF. In order to achieve this approval, the Government had to make sure the QOF represented ‘extra’ money on top of GP’s standard pay packet. This is a contrast to the U.S. policy environment where there is some debate about how to achieve ‘budget neutral’ P4P schemes (Galvin 2006).

What were the objectives and intentions of the QoF, and were they achieved?

In order to evaluate the QOF as a policy intervention we must carefully define its objectives. The Royal College of General Practitioners (2003) outlined the ‘new GP contract’ in a summary paper to its members before the QOF was implemented. The QOF (the ‘pay-for-performance’ part of the contract) was only a part of a more general package agreed between the government and GPs signing on to the GMS contract. The objectives of the scheme as ‘sold’ to GPs by the RCGPs (Royal College of General Practitioners 2003) were to ‘reward’ quality care provided by GPs. There was not an emphasis on ‘improving’ quality in these documents but on ‘addressing’ the emphasis on high volume rather than quality of care in previous contracts.

“...existing GMS contract places far greater emphasis on high volume than quality of care. The new contract will address this through introducing a quality and outcomes framework based on the best available research evidence.” (BMA, 2004)

“The framework... is in line with professional opinion and reflects the ethos that higher quality care is most likely to be achieved through the use of incentives.” (Royal College of General Practitioners 2003)
A guidance document for PCTs released around the same time (Department of Health 2003) provides a more straightforward objective of the QOF:

“The new contract, above all else, is a strategic tool to improve the quality and range of services for patients;…”

The QOF was introduced across the whole of the UK at the same time without being piloted, making evaluation difficult. However, because data collection is an integral part of the policy, evidence was soon available on the levels of performance of GPs in the incentivised quality indicators. Performance in the first year was high (Doran et al. 2006), and was higher than expected or budgeted for (National Audit Office 2008). The median reported achievement across all indicators and all GP practices was 83.4%, this is the proportion of eligible patients who met the prescribed performance targets. In terms of points earned, 95.5% of available points were achieved. This level of achievement led to the government spending 10% (£1.76bn) over the planned budget for GP services.

A key study in evaluating the QOF is a follow up of the study evaluating the evolution of quality prior to the QOF (Campbell et al. 2007, Campbell et al. 2005). Whilst the first study shows an increasing trend in quality indicators for chronic disease areas (asthma, diabetes and heart disease), the follow up shows the trends accelerated following the introduction of the QOF. The authors find no statistical difference between the trends of incentivised and non-incentivised indicators.

Another study which presents data on quality indicators before and after the introduction of the QOF is Sutton et al (2009). The data in this study includes only process measures of quality – recording of risk factors such as blood pressure and cholesterol. The recording of risk factors was incentivised in the QOF for some patients (those with chronic diseases) but not for others. The results indicate the QOF increased recording of risk factors for both incentivised patients (19.9 % - points) and non-incentivised patients (5.3 % - points). The authors interpret these results as evidence of a positive effect of the QOF and ‘positive spillover’ effects in non-incentivised areas. A similar study looking not at the recording of indicators, but at achievement in indicators of asthma and hypertension care, was conducted on a sample of English GPs by Steel et al (2007). The authors found in this case that performance improved substantially (75% to 91%) for incentivised indicators, but not significantly for non-incentivised indicators. Taking both of these results at face value suggests that the QOF had positive spillovers on the recording of quality indicators but not on achievement in those indicators.

The ability of GPs to ‘exception report’ patients coupled with the targets for performance in the QOF being less than 100 percent, may incentivise GPs to neglect the provision of care to the ‘most difficult’ patients. Some studies have sought to examine this issue, for example Doran et al study how the performance of GPs in the QOF varies by the deprivation level of patients. The study finds that there was some inequity in median achievement between the most and least deprived areas in the first year of the QOF, with the least deprived areas scoring four percentage-points higher than the most deprived areas. However, this gap narrowed by the third year of the QOF to 0.8 percentage points. The authors of this study draw the conclusion that this finding is evidence for financial incentives reducing health inequity with respect to deprivation. The attributed causality in this study may be questionable because data before the introduction of QOF is not examined.

A group of Scottish-based authors have also analysed the equity implications of the QOF (Guthrie et al. 2006a, McLean et al. 2007, McLean et al. 2006, Sutton and McLean 2006). A study which might be seen to support the view that the QOF reduces health inequalities is Sutton and McLean (2006). Analysing cross-sectional variation in QOF scores they find GP practices in deprived areas score higher than those in less deprived areas, controlling for confounders such as practice size, location and ex-fundholding status. The implied reduction in health inequalities in this study would be very small as there was little variation in QOF scores with most practices achieving high scores for most indicators (Doran et al. 2006). A similar
study explores this issue in greater detail for an expanded dataset and different definitions of quality performance (McLean et al 2006). The results seem to contrast with their BMJ paper: when looking at delivered quality (including exception-reported patients in the ‘denominator’), the QOF outcomes are lower in deprived areas.

A different perspective on the effect of the QOF on equity is to consider equity of financing. Taking this approach, Guthrie et al (Guthrie et al 2006) give an alternative view of the effect of the QOF on inequalities: the authors find that the payments made by the QOF are markedly lower (per patient treated) in practices with a high burden of disease. This feature of the QOF remuneration system comes about because the ‘adjusted prevalence factor’ used to allow for differences in prevalence, which increases payment with the square root of true prevalence. The inequality in payment violates the inverse care law and has implications for GP workload. Practices in more deprived and more highly diseased areas have to treat more patients than other practices to achieve the same financial reward. McLean et al (2007) show how this financial inequality extends to comparisons between UK countries, with GPs with higher prevalence of disease in Scotland and Wales being penalized by the QOF formula. Partly in response to these studies, the adjusted prevalence factor was replaced by true prevalence in 2009 (British Medical Association 2009).

A number of UK studies have looked for relationships between quality of care as evidenced by QOF scores and emergency admissions. No association was found for emergency admissions for a range of conditions included in the QOF (Downing et al. 2007). However, Shohet et al found that admissions for epilepsy were less common in practices with high scores for epilepsy (Shohet et al. 2007) The inconsistency of these relationships is perhaps not surprising, partly because not all indicators would be expected to have a direct or rapid impact on admissions (the epilepsy ones would), but also because of the strong influence of other factors such as socio-economic deprivation on emergency admissions (Giuffrida et al. 1999, Aylin et al. 2007).

Undesirable consequences

The exception reporting behaviour of GPs is of interest because it relies on GPs making subjective decisions about the ‘eligibility’ of patients. High levels of exception reporting may reflect (1) that some indicators are unsuitable for a large number of patients, or (2) that GPs are ‘gaming’ and falsely exception reporting large numbers of patients in order to improve their performance scores. Doran et al (2008) have analysed the associations between the rate of exception reporting in a practice for a variety of QOF indicators and characteristics of the practice and the patients in the practice. The median level of exception reporting across practices was 5.3% with an interquartile range of 4.0 to 6.9%. Little of the variation in the exception reporting rates is explained by the type of indicator. Indicators associated with measuring disease and providing advice (process measures of quality) were generally found to have much lower rates of exception reporting than indicators associated with providing treatment and achieving intermediate outcomes. This could be explained by GPs ‘gaming’ as treatment and outcome indicators require more effort than the simple process measures. Equally, the authors note there are legitimate reasons why fewer patients would be ‘eligible’ for the treatment and outcome indicators, for example, patients may fail to comply with treatment.

The hypothesis that GPs have been ‘gaming’ the QOF through exception reporting is more formally tested in Gravelle et al (2008). The authors of this study present an economic model of the QOF in which a GP practice aims to maximize its utility (in monetary and non-monetary terms) from the incentives offered. They explicitly consider how the incentives to GPs change over the range of performance (from 0% to 100%) and use this to derive some tests of GP altruism and gaming. The authors find that many GPs show evidence of altruism, by exceeding the maximum remunerated performance levels. However, there is also evidence of gaming: GPs performing below maximum remunerated performance levels in 2004/5 have a higher level of exception reporting the following year.
Two recent studies have also been published which start to address the concern that financial incentives may need to neglect of non-incentivised conditions (McGlynn 2007). This concern does not appear to have been realised in two recently published studies, one from the US (Ganz et al. 2007) and one from the UK (Steel et al 2007). However this type of study inevitably compares quality of care for aspects of care that can readily be measured. Much of the criticism of the Quality and Outcomes Framework in the UK relates to the potential loss of the caring aspects of the general practitioner’s work (Mangin and Toop 2007). The literature continues to provide a wide range of largely anecdotal views on the QOF, ranging from those who believe that good care is being appropriately rewarded and that patients will benefit, to those who believe that the ethos and values of general practice are being fundamentally undermined to the detriment of patient care.

**Attitudes of PCPs towards the QOF**

Taking the perspective of the ‘communication model’ (Giacomini et al. 1996) the perceived intentions of policymakers, communication between policymakers and doctors is clearly important in the response of some GPs to the policy. Mangin and Toop (2007) provide one of the most critical appraisals and raise many common concerns of doctors subject to P4P reforms. Their view is summarized by this quote:

“..UK general practice has moved from having an internal framework of professionalism that supports it, to an external framework that holds it up and embraces a market model of healthcare...”

The authors are also concerned that the public health emphasis of the QOF:

“..conflicts with the individual; patient-centred ethos of general practice.”

These critics of the QOF have also questioned previous quality improvement programs including the National Service Frameworks (Toop and Richards 2001). The following issue of the journal published four letters from UK GPs, two of which (Osborne 2007, Davies 2007) agree with Mangin and Toop’s criticisms of the QOF and two others (Thomas 2007, Van den Heuvel 2007) who cautiously disagree, pointing out some benefits of the QOF. From this debate we can only conclude the response of GPs to the QOF is varied, while some view the intentions of the policy makers behind the QOF as a misguided attempt at “Applied Public Health” (Osborne 2007), others see some positive benefits.

Some more formal research into GPs attitudes and responses to the QOF (McDonald et al. 2007) reinforces the heterogeneous opinion of GPs towards the QOF. In a series of qualitative interviews from two GP practices the authors find that GPs were generally supportive of the effect of the QOF on their clinical practice. A common comment was that GPs acted as ‘the chasers or the chased’ with regard to checking the performance of each other in the practice’s QOF quality scores. GPs generally had a positive view of either taking responsibility for ‘nagging’ fellow GPs or being nagged themselves. However, the authors report that nurses had a less positive experience and were often left to do the more tedious administrative work behind meeting QOF targets.

One way in which GPs and policymakers can communicate is through academic GPs who can straddle the divide between policy and practice. An example in the UK is Martin Roland, who is a practicing GP, academic and was involved in the initial negotiations on the QOF in 2002. He has published prolifically on the topic of the QOF and often strikes a conciliatory tone, for example acknowledging problems in the scheme but also reminding readers that the exception reporting rates and health inequalities observed in the QOF have been better than expected (Roland 2007). He also emphasises a forward-looking view in which the QOF is ‘here to stay’ and problems are there to be solved not simply criticized. Other similar authors are Tim Doran (Doran 2008) and Martin Marshall (Marshall and Roland 2002, Marshall and Smith 2003). These authors potentially provide an important intermediary link, helping to communicate the intentions of an incentive scheme to practitioners, and feeding back insights and criticisms from practitioners to policymakers.
AUSTRALIA

The Practice Incentive Program was introduced in Australia in 1999. This includes capitation payments and incentives payments for a limited number of disease areas.

What contextual factors were important for the PIP?

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. Medicare funding of GP services is based on the Medicare Benefits 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 ranged 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 poor quality of care (Scott and Shiell 1997, Beilby 2003). As there is no financial incentive to register patients in a practice there is evidence of a lack of continuity-of-care in Australia: Schoen et al (2005) show that Australian patients see significantly more (different) doctors on average than patients in the UK or US.

Before the Practice Incentives Program (PIP), there was very little pay-for-performance for Australian GPs. The General Practitioners Immunisation Incentive Scheme (GPII) was introduced in 1998 to offer incentives for each immunization completed (a Service Incentive Payment or SIP) and for overall immunisation coverage (an outcomes payment) (Brand and Wright 2005). The structure of the payments was the same as those that were later implemented for diabetes, asthma, mental health and cervical screening in the PIP in 2001.

An important policy context for the implementation of the PIP has been the Divisions of General Practice, which were first introduced in 1992 and by 1998 covered all of Australia. Divisions are networks of GPs defined by geographical areas: there were 119 divisions in 2007 representing 22,868 GPs (nearly 98% of all GPs) in Australia (Hordacre et al. 2008). Divisions are ‘bottom-up’ organizations characterized by voluntary membership and governance from GPs within the division. They are, however, funded by the Commonwealth and have an important role in implementing government healthcare policy through educating and supporting members. Considerable heterogeneity exists between Divisions in the size and scope of their activities which includes the degree to which government policies are implemented. Divisions had a particularly important role in the implementation of the PIP, especially in the administrative processes such as installing practice IT infrastructure. Scott and Coote (2007, 2009) have shown that Divisions affect GPs PIP participation through improvements in practice infrastructure but do not directly influence measures of clinical quality.

A policy development which has run complementarily, albeit on a much smaller scale, to the PIP is the Australian Primary Care Collaboratives Program (APCC 2009). This is a quality improvement program encouraging GP practices to join a scheme in which general practices undergo training in using a ‘model of improvement’ (Langley et al. 1996) to focus on improving patient outcomes in a few important areas of care. In addition to assisting practice to implement the quality improvement cycle, the collaboratives also assist with IT infrastructure and data reporting. The program started in 2004 and initially covered two chronic diseases (diabetes and CHD) and patient access to care. This was extended to include COPD and chronic disease prevention in 2009. The program initially funded only 500 practices taking part
in the program with the anticipation that this would be extended to around 1000 practices in 2008 (Department of Health and Ageing 2009).

The collaboratives program is important in the context for the PIP because it focuses on chronic disease, and specifically on one disease (diabetes) that is also subject to incentives in the PIP. This extra focus on chronic disease might be expected to increase the potential effect of the PIP incentives on quality of care provided by GPs. Conversely, GPs working through the collaborative ‘model of improvement’ may be motivated to improve quality independent of financial incentives. The collaboratives also encourage regular and detailed collection of data on quality indicators.

What were the key characteristics of the PIP?

The Practice Incentive Program (PIP) has its roots in the Better Practice Program (BPP), introduced in 1995. The BPP became the PIP in 1998, and in 2001 incentives were introduced relating to quality of care in diabetes, asthma, mental health and cervical screening (Brand and Wright 2005). We concentrate on the post-2001 aspects of this policy which relates specifically to incentives for quality care. In 2003 this program contributed to almost 10% of GPs’ income (Senate Select Committee on Medicare 2003). 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 mental health incentive payments in the PIP were phased out in 2007 and replaced with new GP mental health care items available to all GPs (Medicare Australia 2006 p. 74). The asthma Service Incentive Payments were also reformed in 2005, and a number of minor changes have taken place to other PIP payments over time. However, it has largely remained unchanged in terms of the amount of financial incentives.

As an example, the details of the PIP scheme for diabetes are as follows: The sign-on capitation payment is paid to practices who have a register of diabetes patients and a recall system. In 2008, this was 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 for diabetes.
To claim a SIP for a diabetes cycle of care, a GP needs to provide 12 quality of care items to a diabetes patient within 12 months that includes among others three types of blood test (HbA1c test, lipid profile test, and microalbumin test), feet and eye examination, and lifestyle and self-management advice (Diabetes Australia, 2007). For asthma, the cycle of care involves a ‘structured approach’ to care over three consultations within four months (Beilby and Holton 2005), this was later reduced to require only two consultations (Medicare Australia 2009) and was $100 per patient. The SIP for mental health came under the banner of the ‘Better Outcomes in Mental Health Care’ initiative (Hickie and Groom 2002) and offered $150 for completion of a three-step mental health plan (assessment, planning and review), again occurring over three consultation (Hickie et al. 2004).

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.

When they were introduced, the PIP incentives in the four disease areas were also ‘labeled’ and promoted separately in each disease area, often as part of a wider national initiative or program for quality improvement. The diabetes incentives are part of the ‘National Integrated Diabetes Program (Georgiou et al. 2006, De Domenico M et al. 2005), the mental health incentives were part of the Better Outcomes in Mental Health (Hickie and Groom 2002, Hickie et al. 2004) and the asthma incentive part of the Australian Asthma Management programme (Beilby and Holton 2005). There is even seems to be a reluctance to use the overall name ‘Practice Incentives Program’, with some studies only discussing the service incentive payments (SIPs) as part of the disease-specific initiative. An exception is Beilby and Holton (2005) which categorises all the PIP incentives across disease areas while looking only at data for asthma.

Also adding to some confusion about the PIP is their relationship to a different set of Medicare items introduced in 1999: Enhanced Primary Care (EPC) items. EPC items include a number of areas from funding for allied health workers in primary care to childrens health literacy, but also prominently include ‘Chronic Disease Management’ items. These items include the preparation of ‘care plans’ for patients with chronic disease which seem inherently similar to the cycles of care incentivised in the PIP. There is concern in the profession over the administrative burden posed by numerous additional Medicare items (Chew 2004, Aloizos 2005).

What were the objectives of the PIP and were they achieved?

The Medicare Australia website has the following statement on the objectives of the PIP:

“Under fee-for-service arrangements, practices that provide numerous quick consultations receive higher payment than those that take the time to look after the ongoing health care needs of their patients. High throughput can result in unnecessary prescribing, tests and referrals.

To improve this situation, the Practice Incentives Program (PIP) was developed to provide incentives that encourage general practices to improve the quality of care provided to patients.” (Medicare Australia 2009)

There have been few evaluations of the PIP, and the policy has been low-profile in terms of media coverage and the relative size of government expenditure in comparison to high-profile schemes such as the QOF in the UK. In general, evidence has shown variation in participation in the scheme across groups of GPs (Saunders et al. 2008) and low uptake in some disease areas (Zwar et al. 2005). Studies have agreed about the importance of divisions of general practice in promoting the PIP incentives (Saunders et al. 2008, Scott and Coote 2009, Scott et al. 2009).

Two studies concentrate on take-up of the Service Incentive Payment (SIP) for asthma (Zwar et al. 2005) and diabetes (Georgiou et al. 2004) using data from Medicare Australia (previously the Health Insurance Commission). The Zwar et al study shows that asthma SIPs were poorly taken-up by GPs and concludes that this was due to administrative burden of in this particular
disease area. Another study has commented that poor take-up of SIPs for asthma may be due to problems diagnosing eligible asthmatic patients in the absence of disease registers (Beilby and Holton 2005).

For diabetes, studies have been more positive about the PIP. Georgiou et al (2004) study shows that in 2002/3, 95,486 diabetes SIPs were claimed by Australian GPs, and estimate this to be 10% of the diabetic population (or 20% of the diagnosed population). The authors conclude that this represents an improvement in quality when comparing to previous estimates (published in 2000 and 2002). Other studies using data on diabetes have attributed improvements to diabetes disease registers run by divisions of general practice (Harris et al. 2002, Georgiou et al. 2006, Wan et al. 2006).

Hickie et al (2004) report on data for the take-up of the mental health incentives. They are generally positive, citing the better than expected take up of the most basic training required for the initiative (3,046 GPs versus an expected 1,500 GPs) and the figure of 11,377 three-step cycles of care being completed in 15 months. However, this figure is not related to the overall prevalence of mental health conditions (as they are in the Georgiou et al 2004 study for diabetes) so it is hard to assess if this is a meaningful proportion.

Scott et al (2009) provide the main attempt to evaluate the PIP program. They use data from a representative patient encounter-level survey of GPs, and define a treatment group of those GPs in the PIP and likely to be claiming SIP payments. They use an instrumental variables approach, exploiting the correlation between PIP participation and activity of divisions of general practice. The model estimates the effect of PIP on the probability of an HbA1c test for diabetes patients. The authors find a relatively large effect of the financial incentive: they estimate the PIP program increases the probability of an HbA1c test occurring by around 20 percentage points (for example, from approximately 20% to approximately 40% of patient encounters).