QQUIP Quest for Quality and Improved Performance

Financial incentives, healthcare providers and quality improvements

A review of the evidence

Jon Christianson, University of Minnesota

Sheila Leatherman, University of North Carolina and London School of Economics

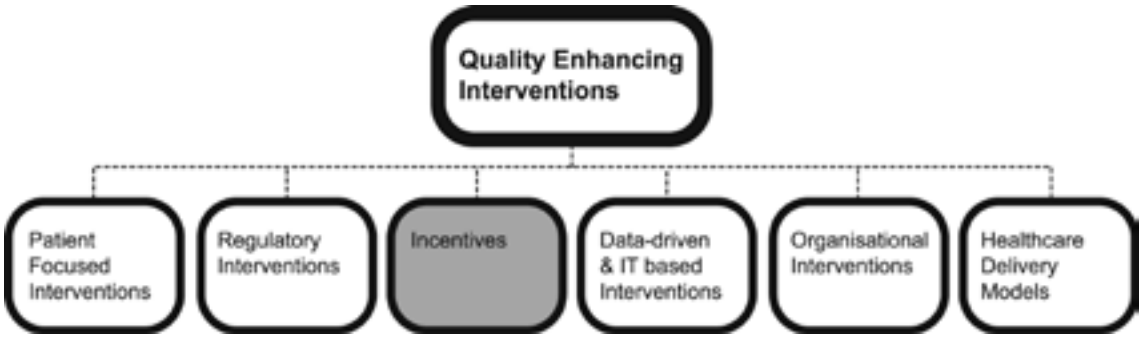
Kim Sutherland, University of Cambridge

The Health Foundation

QQUIP and the Quality Enhancing Interventions project

QQUIP (Quest for Quality and Improved Performance) is a five-year research initiative of The Health Foundation. QQUIP provides independent reports on a wide range of data about the quality of healthcare in the UK. It draws on the international evidence base to produce information on where healthcare resources are currently being spent, whether they provide value for money and how interventions in the UK and around the world have been used to improve healthcare quality.

The Quality Enhancing Interventions component of the QQUIP initiative provides a series of structured evidence-based reviews of the effectiveness of a wide range of interventions designed to improve the quality of healthcare. The six main categories of Quality Enhancing Interventions for which evidence will be reviewed are shown below.



For more information visit www.health.org.uk/qquip

Published by:

The Health Foundation
90 Long Acre
London WC2E 9RA
Telephone: 020 7257 8000
Facsimile: 020 7257 8001

www.health.org.uk

Registered charity number 286967
Registered company number 1714937

First published 2007

ISBN 978-1-906461-02-7

Contents

Executive summary

Within healthcare there has been a longstanding interest in how the type and amount of payment to healthcare organisations and practitioners affects the type and amount of services received by consumers and, ultimately, the costs of healthcare to individuals, employers, insurers and governments. There has been an equally longstanding interest in how the cost of health services to consumers affects the services they seek out and use, along with the implications of this for overall costs at various levels of aggregation. But there has been much less attention devoted by researchers to the impact of financial incentives on the quality of care. The Institute of Medicine in the USA has defined quality of care as 'the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge'(p. 21).

This study reviews the healthcare literature that examines the effect of financial incentives on the behaviour of healthcare organisations and individuals with respect to the quality of care they deliver to consumers. Its purpose is to provide guidance to policy-makers in government and decision-makers in the private sector in their efforts to improve quality of care through payment reforms. In this review and synthesis we assess the quality of the evidence relating to the relationship between financial incentives for providers and quality improvement. Specifically, we address:

1. How effective are efforts to reward providers for improving the quality of care they provide or achieving benchmark levels of quality?
2. Whether, and to what degree, financial incentives intended to restrain costs and utilisation have had secondary effects on quality of care?

We also review studies where the main focus was on analysis of utilisation and costs, and find clear evidence that financial incentives can influence these two outcomes. Although this is outside the scope of our 'quality-enhancing' brief, we include these studies in an appendix in order to provide a sense of how the literature on the impact of financial incentives on quality has evolved over time, and because many countries have a growing interest in modifying use of health services as costs continue to escalate.

Financial incentives directed at improving quality

We reviewed 36 published articles that examined the impacts of financial incentives directed at improving the quality of care delivered by institutional providers and by healthcare practitioners, particularly physicians. The findings from this review can be summarised as follows:

- The findings from studies on the effect of payer initiatives that reward providers for quality improvements or the attainment of quality benchmarks are mixed. Relatively few significant impacts are reported, and it is often the case that payer programmes include quality improvement components in addition to incentive payments, making it difficult to assess the independent effect of the financial incentives.
- Very little research has been done on the impact of direct payments to hospitals to improve quality. The published research to date in this area is too limited to draw conclusions with confidence.
- Though relatively more attention has been paid to preventive services, there is limited evidence that targeted interventions employing financial incentives to improve the delivery of preventive services are effective. The few studies in this area with strong research designs find small, if any, effects of payments to providers that are intended to improve quality.

- The accumulated body of research described in this chapter is not yet sufficient to assess the relative significance of identified barriers to the effective design and implementation of pay-for-performance initiatives.

There are large pay-for-performance programmes underway in the US and the UK with more evaluations likely to appear in the peer-reviewed literature in the near future. Because of the variation in the way these programmes have been designed and implemented, synthesising their findings to provide useful guidance for decision-makers will be challenging. It will be especially important to have comprehensive reporting of results in future studies (not limiting results to a subset of quality measures rewarded by payers), accompanied by complete descriptions of study context and possible confounding factors. In the meantime, policy-makers can support, and learn from, process evaluations of ongoing pay-for-performance efforts with particular attention to accurate documentation of costs as well as continued tracking of outcomes.

Secondary impacts on quality of financial incentives directed at reducing utilisation and costs

We reviewed 45 published articles that addressed, in some manner, financial incentives and their secondary impact on quality of care. We drew the following conclusions from that review:

- The evidence regarding the secondary impacts of financial incentives on quality of care is not compelling. There are several possible explanations. First, the incentives studied were designed, for the most part, to reduce utilisation of services. Generally, the hypothetical link between service reduction and quality in the studies is not clear, especially where utilisation may have been excessive prior to the introduction of different payment arrangements. Second, the literature reports results for a wide range of quality and outcome measures, making it difficult to detect patterns in the findings. The most commonly used outcome measure – mortality – may not be sensitive to the relatively modest changes in financial incentives found in many studies. Also, mortality can be influenced by a host of factors, many unrelated to medical care, making it difficult to isolate the marginal effects of financial incentives.
- How incentives are transmitted to the level at which decisions about treatment are actually made is not clear in most studies. Typically, information is lacking concerning other efforts to address quality via the health plan, hospital, physician practice or government agencies. It seems likely that these efforts would interact with financial incentives for providers to influence quality of care. Most studies do not control for these quality management efforts when drawing conclusions about the impact of financial incentives.
- The use of multiple quality of care and patient outcome indicators in a single research study enables a richer interpretation of findings. However, when results are conflicting in these situations, no clear overall picture of the impact of incentives emerges. Also, it is not clear in most of these studies if the authors adjusted their statistical tests to account for the multiple comparisons undertaken in their analyses.
- The exact nature of provider payment arrangements often is not clearly described in the studies. This is true in particular for comparisons of quality of care under different insurance arrangements. Because the relationships between payment arrangements and quality are likely to be more subtle than the links between payment and service utilisation, the absence of a description of provider payment incentives makes interpretation of findings even more difficult.

- Many of the studies were cross-sectional in design. There may have been considerable variability in provider quality of care, irrespective of financial incentives, that made it difficult for researchers to detect the influence of financial incentives on quality without access to adequate control variables.

Observations

The literature on the influence of financial incentives for healthcare providers on quality of care is under-developed, but that situation seems to be changing at a relatively rapid pace. The science of measuring quality in the healthcare arena is advancing at the same time that purchasers and funders are intensifying their efforts to measure and reward quality improvement. This is likely to generate a significant amount of new research on the topic in the near future. In addition to documenting the relationships between financial incentives and adherence to best practices or changes in patient outcomes, this research should also:

- contribute to a better understanding of the linkages between financial rewards for quality and practitioner behaviour
- assess the cost-effectiveness of pay-for-performance initiatives of different types
- thoroughly document unintended and unexpected effects of pay-for-performance on the healthcare system as a whole.

Definitions and methods

The use of financial incentives to influence behaviour is common in all areas of commerce. The design and impact of incentives has been examined in the research literature at many different levels:

- in theoretical economics, the principal–agent relationship has been studied with the goal of specifying optimal financial incentives in contracts under different assumptions (see Pratt and Zeckhauser, 1985; Eisenhardt, 1989; Sappington, 1991; Milgrom and Roberts, 1992; Prendergast, 1999)
- literature on employee compensation examines use of different payment approaches to encourage desired behaviours on the part of workers (for example, Gerhart and Rynes, 2003; Rynes, Gerhart and Parks, 2005; Sliwka, 2007)
- marketing literature addresses consumer responses to targeted incentive programmes.

Further, an empirical, applied literature on the impact of financial incentives on behaviour can be found pertaining to each sector of the economy, including healthcare, and related conceptual issues are addressed in several academic disciplines, such as psychology and decision-making, in addition to economics (see, for instance, Rynes, Gerhart and Parks, 2005).

Interest in the impact of financial incentives on provider behaviour has, until the last decade, focused on an imperative in publicly funded systems to improve efficiency and, in market-based systems, a desire on the part of purchasers, private and public, to moderate the growth in healthcare costs. It is only recently that attention has broadened to include the specific relationship between financial incentives aimed at providers and quality of care. This new focus has come about, to a large degree, because of a growing body of research and suggests that:

- there is wide variation in the services delivered to patients, unexplained by clinical variables, in response to specific diagnoses (see Wennberg and Gittelsohn, 1973)
- the evidence basis for determining which services to provide in treating many clinical conditions remains weak (see Eddy, 2005)
- for decisions where there is clear evidence that a particular treatment approach embodies superior quality of care, patients frequently do not receive this course of treatment, whether for preventive, acute or chronic cases (see Schoen et al, 2005; Schoen et al, 2004; Hussey et al, 2004; McGlynn et al, 2003).

In reviewing the literature relating to our first question (see introduction above) a significant challenge is establishing an up-to-date picture of findings. Study results are appearing in a variety of academic journals at an accelerating pace, reflecting the increasing number of interventions and pilot programs being launched by purchasers that reward providers for quality improvement or the achievement of quality benchmarks. To address these challenges, we have taken a relatively inclusive approach when searching the literature.

A challenge in reviewing the literature relating to our second question (see introduction above) is identifying the relevant research. This is because the impact of financial incentives frequently is addressed as a sub-question – sometimes an afterthought – in the context of studies designed to address the impact of incentives on other outcomes, especially service utilisation and costs.

Electronic searches were performed by the Centre for Reviews and Dissemination (CRD) at the University of York, with supplementary searches undertaken by the research team. We conducted electronic searches of MEDLINE®, EMBASE, Cochrane Database of Systematic Reviews (CDSR), Database of Reviews of Effects (DARE) and EconLit. We also drew on work from the Agency for Healthcare Research and Quality (AHRQ), Organisation for Economic Co-operation and Development (OECD) and World Health Organization (WHO). Search terms are shown in Appendices 2 and 3.

We aimed to bring together the best available evidence on incentives as used in healthcare contexts. In line with the search strategy adopted across the Health Foundation's quality enhancing inititiatives (QEI) series, we used a 'best evidence' approach to conduct our review, initially searching for high quality systematic reviews. If reviews were sparse or out of date, we then searched for individual studies within the following hierarchy of evidence:

- randomised controlled trial
- quasi-experimental study
- controlled observational study (for example, cohort or case-control study)
- observational study without control group (for example, cross-sectional study, before-and after study, or case series).

Broad inclusion criteria were adopted because of the methodological challenges inherent in assessing the impact of incentives on outcomes and processes of healthcare. The search strategy comprised two phases: the first focused on retrieving systematic reviews, using a wide range of search terms (see Appendix 2). The second used a more limited number of search terms but included a wider range of research designs (see Appendix 3).

Article titles and abstracts for 1564 articles were reviewed by two report authors during an initial search, which covered the period from January 1988 to August 2006. We updated our search results in June 2007, adding newly published articles as well as those that came to our attention through further scans

of reference lists in related articles. There is a substantial research literature prior to our starting date, particularly with respect to the impact of financial incentives on utilisation and costs. We rely on the results of published research syntheses to characterise the findings from this early literature (see Appendix 1).

We selected articles that were empirical and that focused on the effect of financial incentives on some aspect of provider behaviour. At this stage, we did not require that the studies address impacts on quality specifically. However, we did eliminate studies from consideration that focused on providers that delivered only mental health and substance abuse services. There are a number of strong studies in this area, but we chose to focus specifically on providers of medical services. We also eliminated studies of the impact of financial incentives on provider location decisions, choosing to focus only on how incentives affect treatment decisions more directly. Again, there is a substantial literature in this area, largely consisting of evaluations of government programmes designed to reward physicians for locating in under-served areas.

We retrieved and reviewed the full articles in every case, writing an article summary using a standard abstract format. The majority of the articles that we retrieved through our search process, and therefore included in this review, analysed payment reforms and initiatives set in the US. This no doubt reflects the multiplicity of payers in the US, creating more opportunities to evaluate different payment initiatives, as well as the relatively large number of academics conducting research on health services there. With respect to studies involving statistical analyses, there was wide variation in the sophistication of the analytic approaches used. Not surprisingly, studies where large datasets were employed, and where findings were published in academic research journals, used the most sophisticated statistical techniques. In contrast, those published in journals aimed more at practitioners or policy-makers and that sought to offer ‘early findings’ regarding specific interventions, were more likely to present simple comparisons of mean values, sometimes without statistical tests for significance of observed differences.

1. Financial incentives and quality improvement: background and overview

Definition and background

Financial incentives, as the term is used in this review, can be described as the influence that payments to organisations and individuals have on the health services they deliver to consumers. Economic theory suggests that different types of payment approaches can lead to different types of behaviours, with each payment approach having its own strengths and drawbacks. In this review, we focus on explicit financial rewards or sanctions that are directed towards improving quality of care delivered by healthcare organisations or practitioners to patients (Chapter 2). However, these rewards or sanctions often are layered on top of, or blended with, payment approaches designed to affect service delivery in other ways (Robinson, 1999), as described in Box 1.

When investigating the literature on the impact of financial incentives on quality, the first challenge is to understand the specific nature of the financial incentives in question. In experimental studies, clear, simple financial incentives can be employed. The drawback is that incentives in practice are likely to be much more complicated and nuanced. The provider payment incentives in observational studies reflect the more complicated reimbursement arrangements that exist in the 'real world', which is a strength of these studies. However, in observational studies it often is difficult to sort out the effect of financial incentives from other factors that influence provider behaviour, and researchers often do not take the time to understand this, opting instead to characterise incentives using the capitation/fee-for-service/salary shorthand categorisation (see Box 1).

It is presumed, based on theoretical models of behaviour and a considerable body of empirical evidence from healthcare and other fields, that financial incentives can be manipulated to affect provider treatment decisions. The challenge to funders or payers is to assemble the mix, or blend, of financial incentives, rules and monitoring effort that results in the desired provider behaviour. There are many factors that merit consideration. For instance, how high must the level of payment be to achieve the desired outcome (Conrad et al, 2006; Young and Conrad, 2007)? It may be that the required payment level to achieve a desired behaviour exceeds the value of achieving the desired outcome. If a research study finds that, in a particular situation, providers did not respond to a financial incentive by changing their behaviours, it might seem reasonable to recommend that policy-makers increase the payment in order to achieve their goal. However, this would not be warranted if the value of the gains from the behaviour change was not sufficient to justify the higher payment.

The general literature on payment arrangements suggests that, even under the best designed payment scheme, desired outcomes are not likely to be achieved through financial incentives alone; resources also need to be devoted to measurement and monitoring activities. The greater the cost of these activities, the less likely that use of financial incentives, all else being equal, will be a cost-effective approach to changing provider behaviours. Equally important, the effectiveness of any financial incentive scheme in eliciting changes in provider behaviours depends not only on the amount and type of payment, but also on the characteristics of recipients themselves (for example, their preferences for monetary versus other rewards, such as autonomy, security and so on) and of the context in which they practise (for example, whether the risks associated with a particular incentive scheme are borne by the individual provider or shared within a provider group).

Structure of report

This chapter has provided an overview of the use of financial incentives to influence performance and the nature of incentive arrangements commonly employed.

Chapter 2 focuses exclusively on evaluations of programmes or initiatives where payments to either healthcare organisations or practitioners depended in part on improvement from baseline measures of quality or meeting predetermined quality benchmarks. We provide the greatest detail regarding the studies reviewed in this section, because of the current interest on the part of both governments and private payers in the notion of 'pay-for-performance', and because there are fewer existing syntheses of this literature. However, there are significant challenges in generalising the findings from these studies, which we also discuss in our review.

Chapter 3 summarises findings from studies where quality measures were introduced explicitly into the analysis, and where impacts of financial incentives on quality were assessed empirically. However, in contrast to Chapter 2, the articles in this chapter evaluated the influence of incentives on, with a secondary focus on quality impacts.

Chapters 2 and 3 are organised using the same format. After a general introduction to the subject matter, we summarise the findings from the literature. Then we describe the studies we reviewed, by type of study, followed by more detailed summaries of individual studies.

Appendix 1 reviews studies where the main focus is on analysis of utilisation and costs, and where the impact of financial incentives on quality is discussed only in a speculative way, if at all. We provide this Appendix for two reasons: to offer a sense of how the literature on the impact of financial incentives on quality has evolved over time, and because many countries have a growing interest in modifying use of health services as costs continue to escalate.

Box 1: Payment approaches

Capitation: Under capitation payment, the provider agrees to deliver a specified list of health services to a predetermined group of individuals for a fixed amount per person per time period. The provider bears financial risk in situations in which the actual cost of these services exceeds that fixed amount. Conversely, the provider retains at least a portion of monies that accrue because the cost is less than the predetermined reimbursement. The most frequently voiced concern about how capitation might affect quality of care is that the provider entity receiving the payment might act too aggressively in constraining service use, eliminating some necessary as well as some 'unnecessary' services. The result could be lower quality of care for patients. This is especially true, it is thought, if there is no sharing of risks or surpluses, if the capitated contract is short term, and if contract renewal does not depend on measures other than costs. In this situation there are few, if any, financial incentives for providers to improve quality, unless that improvement reduces costs as well. However, others have argued that if provider organisations, reimbursed by capitation payments, care for an enrolled population over a period of time, they have an incentive to provide services that maintain or improve the health of that population as this will, over the longer term, be financially beneficial. In this more stable environment, capitation payment could contain incentives that are quality-enhancing.

Continued over

Box 1 continued

Fee-for-service: Fee-for-service payment in medical care is essentially the counterpart to 'piece-rate' payment in other sectors of the economy. Providers are rewarded for providing more services to patients. Robinson (2001) has noted that this form of payment for services is not common outside of medical care because it contains such a powerful incentive for over-provision of services on the part of the 'agent' and therefore necessitates a substantial amount of costly monitoring on the part of the payer, especially when output cannot be easily measured. In some respects, it seems intuitive that more medical care would benefit patients in most cases, and therefore that fee-for-service payments would not have a detrimental impact on quality of care or patient outcomes. However, there is research suggesting that more care does not necessarily imply higher quality or better outcomes. Provider incomes increase whether or not the services they provide are needed, or of dubious worth, and a growing literature suggests that there is a risk to patient health associated with 'over-treatment', just as there is with 'under-treatment' (see Institute of Medicine, 2001). To modify the incentives under fee-for-service, arrangements based on payment per episode or, for institutions, payment per admission have been introduced. These approaches bundle services for payment purposes, creating incentives for providers to limit the services they provide in response to a specific event. However, unlike capitation, providers receive more revenue the greater number of events they treat. They are not assigned the responsibility to provide all needed services to a defined population.

Salary: Under salary arrangements, the provider is paid a fixed amount per time period. There is no incentive to deliver unnecessary services, nor is there an incentive for 'under-provision', except to the degree that providers may attempt to provide fewer services – that is, work less hard – than expected in the contractual arrangement. There is no particular incentive under a pure salary method of payment for providers to deliver high quality care, so typically there is a heavy reliance on enforcement of rules and procedures thought to enhance quality. The result could be quality-enhancing or, to the degree that rule enforcement limits provider ability to bring professional judgement to bear in treatment decisions, result in lower quality of care.

Budgets: At the organisational or institutional level, providers may be reimbursed through a negotiated budget process. The nature of the incentives in this payment arrangement can resemble capitation, when the number of individuals served in a given period is fixed, and the organisation carries the risk for budget over-runs as well as retaining savings. Or, the incentives can resemble those of salaried providers when the organisation serves patients who seek care, but does not assume responsibility to provide care to a fixed number, or enrolled group, of individuals for a specified time period.

Together, these observations from the literature suggest that drawing the appropriate insights from provider responses to financial incentives will depend greatly on an understanding of the context in which the incentives were employed (see Town et al, 2004). The same incentives, employed in different contexts, could yield quite different results relating to quality of care. This implies that there is not likely to be one single answer that can be drawn from the research regarding how to structure financial incentives for providers in order to improve quality of care. This observation is independent of the strength of the research design, or fidelity of implementation of the payment arrangement. The following are three examples of context:

1. Entity receiving payment: The literature typically speaks of financial incentives for providers, or paying providers to improve various aspects of performance, without considering exactly what this means in practice. For instance, when the provider is the hospital, and the reward is for improvements in hospital procedures, it is expected that the payment will motivate hospital administrators to restructure processes or take other steps to encourage change in the desired direction. However, it is not clear that the financial incentive actually will be passed through to physicians, nurses or others delivering care to patients. Hospitals

could respond in any number of ways to the financial incentive. The results could reflect the effectiveness of the hospital as an organisation as much as the type or level of the incentive payment. Similarly, when physicians are rewarded for improvement in chronic care processes, the impact of the reward could depend on whether the physician practises in a solo setting or as part of a group. If the reward is paid to the group, then how the group decides to distribute the monies to individuals in the group (if at all) could have a major impact on the type or amount of behavioural change that occurs (Christianson et al, 2006; Young and Conrad, 2007).

2. Scale and scope of effort: At present there are a large number of programmes attempting to blend financial incentives that reward quality into existing payment arrangements. These efforts range from small-scale experiments, with limited time frames and targets for improvement, to national initiatives, such as the pay-for-performance scheme, that has been implemented for general practitioners (GPs) in the UK (Roland, 2004). In the former, relatively little of the physician's practice revenue is affected by the financial incentives. Because of the pilot nature of these payment arrangements, the physician practice may determine that it makes little financial sense to invest in practice reforms necessary to achieve the designated improvements or benchmarks. In contrast, where the incentive scheme is integrated into physician payment, includes a broad range of targets, has a potentially significant effect on provider incomes and is perceived to be a systemic change, rather than a pilot effort, physician practices may respond in fundamentally different ways. Financial incentives that are structured the same with respect to achieving goals in a specific clinical area could have quite different effects depending on the scale and scope of the overall payment reform effort.

3. Concurrent incentive programmes: In countries with pluralistic healthcare financing systems featuring many different payers, it is common for providers to be paid in multiple ways, each with different incentives to improve quality and with different quality goals. For example, one scheme might reward a physician group for achieving benchmarks on diabetes treatment, while another might emphasise medication management for heart failure patients. Some of the financial incentives implemented by different payers may reinforce each other, but others may not. This situation becomes even more complicated when individual purchasers have incentive schemes that reward cost and utilisation control along with quality enhancement, or when medical care organisations reward provider productivity along with achievement of quality benchmarks (Reschovsky and Hadley, 2007). The physician practice or individual physician then must decide where to allocate resources and attention in response to a complicated set of financial signals from purchasers. In this context, one might expect a different response than if the same incentive was implemented in a system where there was one payer, or a dominant payer whose actions others mimicked and where physicians were rewarded only based on quality measures.

In summary, research on the impact of provider financial incentives on the quality of healthcare must overcome several challenges for findings to be useful to decision-makers. As with all research, a rigorous study design enhances the power and credibility of the findings. But, in addition, researchers will need to be aware of, and shape their research approach to address, the complex blended payment arrangements likely to be present in observational studies. They will need to address the context in which financial incentives are implemented when interpreting their findings, especially regarding how they can be applied to other settings.

2. Financial incentives directed at improving quality

In this chapter we review the research relating to financial incentives employed by purchasers/payers to improve quality of care and patient outcomes. Recently, incentives with this objective have been labelled 'pay-for-performance' (in the USA this is sometimes abbreviated to P4P). Providing financial rewards to organisations or workers within organisations, based on measures of performance, is commonplace. Corporate entities often base part of their payment to their suppliers on the quality of the product they are purchasing. However, the rapid expansion in the use of this approach by healthcare purchasers over the last five years is noteworthy.

Historical context

The impetus behind the current era in the research on provider payment incentives and its impact on quality began to emerge in the mid-1980s, reflecting attempts at that time by funders and payers to alter the way in which they reimbursed providers. In the UK, a Good Practice Allowance was proposed to provide financial incentives for high quality care, but was rejected in a 1986 meeting of the British Medical Association (BMA) (Roland, 2004). Subsequently, in the revised 1990 contract between the NHS and GPs, there were financial incentives for performance on immunisations and the monitoring of Pap smears, along with the establishment of health promotion clinics. The scale of this effort greatly increased in 2000 when the UK allocated more funds to the NHS. Subsequent negotiations between the BMA and the NHS resulted in the allocation of £1 billion ($1.8 billion) towards additional payments for high quality care (Roland, 2004). GPs could accumulate points that would lead to a potential increase of £42,000 per doctor, a scheme labelled the Quality and Outcomes Framework (see Box 2). More recently, the concept of practice-based commissioning (PBC) was introduced in the UK. Under PBC, GP practices may retain financial savings from innovation in the provision of care outside of the inpatient setting to reinvest in patient care (Department of Health, 2006). A new hospital consultant contract implemented in 2003 focused on organising consultant time, but did not contain direct financial incentives for achieving quality indicators (Williams and Buchan, 2006).

Box 2: UK context – contracting with general practitioners (GPs)

In April 2004 new primary medical care contracting arrangements were introduced in England. The General Medical Services (GMS) contract covered those GPs who work for the NHS as independent contractors. Personal Medical Services (PMS) contracts were offered to those GPs who, rather than maintaining independent contractor status, opt for salaried appointments. Around 40 per cent of GPs work to a PMS contract. The new arrangements sought to:

- reward practices that offer higher quality care
- allow GPs greater flexibility in their working lives
- ensure patients benefit from a wide range of services in the community
- create flexibility for primary care trusts (PCTs) to commission from traditional and new providers and to innovate, in terms of workforce, infrastructure and IT.[1]

A key feature of the GMS contract is the Quality and Outcomes Framework (QOF) – a system of financial incentives which rewards GPs for delivering quality care. QOF links up to 25 per cent of practice income to performance. Currently, a maximum score of 1000 points is possible, and each point awarded has a monetary value. Achievement is measured across a scorecard of 146 indicators in the following domains:

Continued over

Box 2 continued

- **clinical** standards linked to the care of patients (655 points)
- **organisational** standards relating to records and information, communicating with patients, education and training, medicines management and clinical and practice management (181 points)
- **patient experience**, based on patient surveys and length of consultations (108 points)
- **additional services**, covering cervical screening, child health surveillance, maternity services and contraceptive services (36 points)
- **holistic care,** based on achievement across the clinical domain (20 points).[2]

In the USA, by the early 1990s some managed care plans had contracts with incentives that rewarded achievement of quality benchmarks. However, research evidence was accumulating that quality of care – whether for members of health maintenance organisations (HMOs) and similar arrangements, citizens who received care through government-managed healthcare systems or enrollees in relatively unmanaged, private fee-for-service plans – was questionable in many respects (see, for instance, McGlynn et al, 2003; Institute of Medicine, 2001). There appeared to be considerable opportunities for improvement in quality, but there was little consensus among managers, purchasers or providers about how to accomplish quality improvement. Influential bodies such as the US Institute of Medicine suggested that, among other initiatives, financial incentives in payment systems needed to be realigned to encourage providers to deliver care that met existing standards for quality. Health plans in the USA took up this challenge, seeing it as a way to 'create value' for their corporate and government customers, and some large purchasers were convinced that rewarding providers for better quality services was a win–win proposition as it could lower costs while at the same time increase the satisfaction of employers or citizens.

Rosenthal et al (2004) summarised the largest pay-for-performance programmes underway in the USA in 2003. More recently, Rosenthal et al (2006) have documented programmes implemented by HMOs, while Epstein (2006) has described public sector efforts and Nichols and O'Malley (2006) have discussed hospital programs. McElduff et al (2004), Roland (2004), and Smith and York (2004) have described pay-for-performance initiatives in the UK, and Pink et al (2006) have contrasted pay-for-performance programmes in the USA, the UK and Australia. These programmes have been applied to both hospitals and physicians, but their application to physician payment seems to have raised the most complex design and implementation issues.

Conceptual issues

Some authors have pointed out that paying for performance can diminish the intrinsic satisfaction present in carrying out 'work' (for example, Berwick, 1995; Gagné and Deci, 2005), although all do not agree on the potential significance of this effect (for example, Eisenberger and Cameron, 1996). The argument is thought to apply especially well to medical care providers, who are believed to be motivated to a substantial degree by a desire to help their patients and a duty of professionalism. The argument is made that paying providers for achieving standardised metrics relating to care processes trivialises the non-financial motivations of providers and could lead them to resist pay-for-performance efforts. However, most of the concerns expressed in the literature about applying pay-for-performance

1 Caron Hartley (2005), *General Medical Services (GMS) Contract (England),* NHS National Library for Health website

2 BMA (2006), *Revisions to the GMS contract, 2006/07,* BMA website www.dh.gov.uk/assetRoot/04/12/85/97/04128597.pdf

to physicians relate to possible problems of implementation. Sceptics argue that there are a number of implementation barriers which, if not carefully addressed, could result in pay-for-performance initiatives that are ineffective or that actually reduce quality of care. The most common barriers to effective physician pay-for-performance systems cited in the literature are set out below:

- The problem of small numbers: Many physicians may not have sufficient numbers of patients in their practices with specific medical problems (for example, diabetes) to measure their performance in a reliable way. When there is a relatively small number of patients of a particular type in a physician's practice, performance on quality metrics is likely to reflect, to a significant degree, random variation (Hofer et al, 1999). For physicians practising in medical groups, this problem could be addressed by aggregating the performance of individual physicians to the group level. However, the 'incentive effect' of the payment at the level of individual physicians then becomes a matter of group policy (Christianson et al, 2006), and it will not always be the case that group physicians who perform the best on pay-for-performance metrics receive higher payments (Bokhour et al, 2006). Where multiple payers are involved, the small numbers problem can be addressed by pooling patient data across payers. This may or may not yield numbers of patients 'large enough' for a valid assessment of physician performance. This approach assumes that payers are both willing to share data and able to agree on a common set of performance measures to reward – these are both likely to be daunting tasks. Where the small numbers problem cannot be addressed effectively, the resulting performance measures may lack credibility and weight with physicians, who therefore may not be willing to expend effort and practice resources to improve their performance on these measures.
- Managing to the metric: This involves linking a portion of physician payment to performance on a limited number of predetermined measures. This could diminish the quality of care experienced by patients whose care is not 'targeted' for improvement, especially if practice resources are reconfigured to improve pay-for-performance scores to the neglect of other aspects of care. Managing to the metric also can, in theory, stimulate physicians to seek out patients who are likely to do well, all else being equal, in the areas being measured, while referring less attractive patients to other providers when possible. A concern often expressed is that less educated, or possibly less motivated, patients will find it more difficult to access care because physicians may view them as less dependable in managing their illnesses by following physician instructions. As a result, health outcomes for these patients could decline even as the outcomes of other patients improve. Finally, as Berwick (1995) has noted, pay-for-performance can discourage innovation in areas not addressed by pay-for-performance metrics. The use of targets in the UK has demonstrated both the power of concentrating attention and resources on certain conditions and populations as well as the unintended consequences that can occur (Galvin, 2006).
- The luck of the draw: Patients are not likely to be randomly distributed across physicians with respect to the severity of their health conditions. Some providers become expert at caring for difficult cases and over time accumulate a disproportionate number of these patients in their caseloads. Where this occurs, these more expert physicians could grade poorly on performance for metrics based, for instance, on the percentage of practice patients in compliance with pay-for-performance standards. The solution is to 'risk-adjust' performance measures using a methodology that (statistically) evens the playing field across physician practices with respect to severity of patient mix. However, application of risk-adjustment techniques is often controversial. They can be difficult to explain and require sophisticated statistical methods to implement, which can cause physicians to view them as arbitrary 'black boxes' and to be suspicious of their validity.
- Pay-for-performance costs: An obvious cost of pay-for-performance is the money paid out to providers who achieve improvements in quality of care or perform at targeted levels with respect to pay-for-performance metrics. If these payments are not 'new money', then

hospitals, physicians and other healthcare service providers are likely to view pay-for-performance as an attempt to reduce reimbursement (Ferman, 2004). However, pay-for-performance programmes that funnel substantial additional monies into the healthcare system may entail costs that are not acceptable or sustainable by payers and purchasers who believe expenditure is already too much for variable or even substandard care. Even if payments are not new money, payers incur other costs associated with pay-for-performance initiatives, including costs of analysing data and administering the reward process. It remains to be seen whether payers and purchasers will sustain pay-for-performance initiatives over time if their net effect is to add to medical care costs, rather than to reduce them, even where quality gains are evident.

- How much is enough?: The general premise that individuals will work harder, or will provide more of a particular behaviour, if rewarded to do so, is intuitively appealing. It also seems reasonable to assume that higher rewards will result in greater effort, up to a point. However, a more complete theory of physician and institutional decision-making with respect to pay-for-performance would incorporate transactions costs related to collecting and reporting data, as well as reconfiguring processes to achieve higher scores on pay-for-performance metrics. The size of the payment necessary to elicit the desired behavioural change clearly will be related to the transaction costs incurred by providers in raising their performance (Fisher, 2006; Young and Conrad, 2007). These transaction costs will vary by provider characteristics and the nature of pay-for-performance performance metrics. Therefore, there is no single ideal payment level needed to bring about the results desired by payers and purchasers. In some instances, pay-for-performance will be ineffective because the performance reward is too small while in other cases the size of the reward will be more than is necessary to bring about change. It is likely that some pay-for-performance programmes will not be effective simply because of the difficulty in specifying appropriate reward levels.

Survey and interview data provide a basis for concerns about whether financial incentives will motivate institutional and individual providers to invest in quality of care improvements, or whether they will be viewed as negative and unproductive. For example, a physician survey conducted in Canada (Anderson et al, 2006) found that physicians considered participation in a specific pay-for-performance programme to be ‘burdensome and time-consuming’ (p 470). In-depth interviews with physician practice executives suggest that a desire to be a ‘good doctor’ may be a more powerful incentive for performance improvement than financial rewards (Bokhour et al, 2006). When Coleman et al (2001) interviewed physicians about their response to financial incentives rewarding them for providing anti-smoking advice to patients, he found that they reported no change in their practice as they did not want to ‘confront’ their patients. A survey of physicians in Minnesota (Keating et al, 2004) found that those who received payments based in part on performance reports and quality profiles were more likely to feel pressure to limit referrals. Young et al (2007), based on a survey of physicians in Massachusetts and California, reported generally positive physician attitudes towards pay-for-performance, but negative attitudes towards the amount of incentive money in these programme, their understanding of programme details and their expectations about whether the programme would improve quality. In a national survey of general internists in the USA, Casalino et al (2007) found that most physicians believed that quality measures in pay-for-performance were not accurate and that their use could divert attention from other aspects of care delivery or cause physicians to avoid some types of patients. Two studies based on interviews conducted in the UK reported that physicians appeared to use their rewards from pay-for-performance programmes in ways that could improve quality. Spooner, Chapple and Roland (2001) concluded that financial incentives were important in motivating physicians to improve patient care, with reward money spent mostly on augmenting staff and equipment in the GP’s practice. Benson (2002) found that GPs are more likely than specialists to incorporate computers in their practices. The reason for this, the authors suggest, is because computers assist GPs in achieving financial rewards for quality improvement, with no such rewards available for specialists.

Overall, the literature based on physician surveys and interviews suggests that physicians have very mixed views concerning pay-for-performance (for example, Bodenheimer et al, 2005), and that this could be a significant challenge to payers in implementing successful pay-for-performance initiatives. Complementing these studies, the literature describing potential barriers to the successful deployment of physician pay-for-performance initiatives, while speculative in nature, certainly raises questions concerning how these programmes should be designed and implemented to enhance their likelihood of success (McMahon et al, 2007; Rosenthal, 2007). Programme success can be assessed through empirical research. However, as Dudley has noted (2005), drawing lessons from evaluations of disparate pay-for-performance programmes may be as challenging, for a variety of reasons, as successfully implementing these programmes. Implementation challenges notwithstanding, support for introduction of pay-for-performance incentives in physician and hospital payment schemes has come from both government and the private sector (Roland, 2004; Milgate and Chen 2006; Kahn et al, 2006).

In the remainder of this chapter, we review the existing literature, which is at an early stage of development. We begin with an overall summary of the findings. We then discuss findings from published reviews of studies that assess the use of financial incentives for providers to improve quality. Following this, we discuss the results of specific studies. The literature comprises two main types of study:

- The first set of studies consists of highly structured, but relatively limited, experiments in physician payment. To some extent, this literature provides the scientific underpinnings for the recent proliferation of purchaser pay-for-performance efforts.
- The second set consists of evaluations of more broadly based initiatives in the USA and the UK that have restructured provider reimbursement to reward improvement relative to quality metrics or the achievement of quality goals. We separate research pertaining to institutional versus physician payment on the assumption that these two settings pose different challenges for the design and implementation of pay-for-performance initiatives.

In conclusion, we discuss where future research efforts relating to pay-for-performance could prove to be the most useful for decision-makers.

Evidence summary

We reviewed 36 published articles that examined the impacts of financial incentives directed at improving the quality of care delivered by healthcare organisations and by healthcare practitioners, particularly physicians. The findings from this review can be summarised as follows:

- Studies on the effect of payer initiatives that reward providers for quality improvements or the attainment of quality benchmarks have mixed results. Relatively few significant impacts are reported, and it is often the case that payer programmes include components in addition to incentive payments, making it difficult to assess the independent effect of the financial incentives.
- Very little research has been done on the impact of direct purchaser/payer payments to institutional providers (hospitals and clinics) to improve quality. The published research to date in this area is too limited to draw conclusions with confidence.
- Though relatively more attention has been paid to preventive services, there is limited evidence that targeted interventions employing financial incentives to improve the delivery of preventive services are effective. The few studies in this area with strong research designs find small, if any, effects of payments to providers that are intended to improve quality.

- The accumulated body of research described in this chapter is not yet sufficient to shed light on the validity of the concerns that have been raised about potential barriers to the effective design and implementation of pay-for-performance initiatives.

Description of findings

Literature reviews

We identified seven articles that reviewed literature on the relationship between financial incentives and quality of care, where the incentives were intended to improve quality. Five of these review articles had a broad focus in their search strategies, including all types of care (Armour et al, 2001; Dudley et al, 2004; Petersen et al, 2006; Rosenthal and Frank, 2006; Scott and Hall, 1995), while two limited their scope to specific types of care: immunisations (Achat et al, 1999) and preventive care (Town et al, 2005). In practice, there was little difference across the seven articles in the actual studies reviewed, because most of the incentive schemes uncovered in the authors' literature searches involved paying physicians or physician groups for improvements in preventive care of some type. The reviews did not include recently published evaluations of purchaser pay-for-performance initiatives.

In the earliest review, Scott and Hall (1995) identified studies that examined changes in GP reimbursement, observing that this literature lacked linkages between reimbursement and actual patient outcomes. They cite one study done in the UK in 1991 that reported that GPs were more likely to reach some quality targets if paid specifically to do so, but noted that the research design was weak. Achat, McIntyre and Burgress (1999) followed with a systematic review of the literature assessing the impact of financial incentives for patients and providers on immunisation rates. The review focused on two provider studies published in the early 1990s – Kouides et al (1998) and Ritchie et al (1992) – that evaluated experiments in which primary care physicians were rewarded for improving levels of immunisations or achieving targets. Both studies indicated some degree of improvement. Armour et al (2001) conducted a literature review which also found few studies relating to the impact of financial incentives on quality of care. In the studies where quality metrics were used as outcome measures, there were mixed results. The authors noted the very limited amount of research that had been done to assess the impact of financial incentives on quality of care.

Four reviews published more recently employed very specific screening criteria regarding research design. In a non-systematic review of the literature, Conrad and Christianson (2004) concluded that studies of explicit financial incentives and their impact on quality are directed almost entirely at preventive care, with inconclusive results. They noted that these studies provide little insight into the role of market and organisational factors as mediators of the effects of financial incentives on quality. In a systematic literature review, Dudley et al (2004) required that articles use random assignment to be included in the review. Of the eight studies they identified that used randomised trials, four had incentives directed at individual practitioners. There were significant improvements in quality for five of the seven quality indicators used in these studies. Across all studies in this review, it appeared that significant quality improvements were more likely to be observed when the physician incentives modified a fee-for-service payment, as compared to taking the form of a bonus payment. The authors note that seven of the eight randomised trials involved preventive care. Town et al (2005) addressed the impact of financial incentives exclusively on preventive care. Their search process, which focused on literature published between 1966 and 2002, identified six studies involving eight different uses of financial incentives and concluded that only one incentive was associated with a greater provision of preventive care on the part of physicians. However, the authors cautioned that the incentives generally were weak, suggesting only that small rewards are unlikely to motivate significant improvements in use of preventive services. In a review published in 2006, Petersen et al searched the literature from 1980 to 2005, identifying 17 studies for review. Thirteen of the studies focused on process of care measures, mostly relating to preventive

care. The authors found that five of the six studies with physician-level incentives, and seven of the nine with provider group incentives, found some positive effect on quality. In four cases there was some evidence of unintended side effects. As in the other reviews, the authors noted the relatively few studies assessing the impact of financial incentives on quality of care. In another review published in 2006, Rosenthal and Frank found seven empirical studies of the effects of paying for quality of care. They reported two positive findings in these studies, with studies that had stronger research designs reporting no impacts associated with the use of financial incentives. They also noted the focus of these studies on preventive care.

All of the reviews noted the small amount of rigorous research on the impact of paying providers for improvements on quality of care. They also noted that the few published studies meeting their inclusion criteria mostly addressed preventive care. The most reasonable conclusion, based on these reviews, is that the evidence base for both justifying and designing pay-for-performance schemes to improve quality of care is thin. Nevertheless, as noted above, there are a large number of pay-for-performance initiatives now underway. Evidence relating to the effectiveness of these initiatives is just beginning to emerge in the peer-reviewed literature and was not included in the review articles.

Achat, McIntyre and Burgress (1999) reviewed use of incentives to influence immunisation uptake, identified issues in developing incentive programmes and examined findings in the context of a new immunisation incentive scheme in Australia. They conducted a MEDLINE search, in English, under immunisation and financial incentives, from 1966 to 1998. They discussed a US study in New York by Kouides *et al* (1998) which found that, when primary care physicians were rewarded for reaching a 70 per cent target with a fee increase of 10 per cent, the average rate was 73.1 per cent compared with 55.7 per cent in a comparison group. Incentives were less influential in practices with fewer than 100 patients. Ritchie *et al* (1992) looked at changes in rates after the implementation of a new contract for GPs in Scotland in 1990. GPs received additional payments of £1800 (high target – 90 per cent) and £600 (low target). The number of physicians achieving 95 per cent or more rose from 31 to 81 per cent for primary immunisation and from 23 to 64 per cent for preschool boosters. The reasons for the increase were not clear, and there were other factors at work in addition to the financial incentives. Based on the discussion provided, it is not possible to determine the strength of the study designs used by these authors.

Type of study: Systematic review

Armour *et al* (2001) reviewed the impact of explicit financial incentives at the physician-level on resource use (hospital and visits) and quality measures. The literature review was conducted following the Cochrane Collaborative handbook. The authors did not state how many articles were identified through their review, but they discussed two articles related to resource use and four related to quality of care. One article related to resource use was based on data from a survey of medical directors (Hillman, 1989). The second examined the impact of bonus payments at the physician versus the physician group level. Incentives directed at the individual physician-level were found to be the most effective. The authors reported mixed results regarding the four studies where quality measures were used as outcome variables. One study found no impact while another reported that quality of care, measured by children's immunisation rates, improved. The authors noted the very limited amount of research related to the impact of imposing direct financial incentives on physicians.

Type of study: Systematic review

Dudley *et al* (2004) conducted a literature review of the evidence on strategies to support quality-based purchasing, which includes a review of the literature on use of financial incentives for providers to improve quality. The authors concluded that a performance-based provider payment could 'plausibly be introduced by a purchaser'. A variety of different outcomes were measured across the studies that were reviewed. The authors interrogated MEDLINE and Cochrane databases, as well as databases documenting ongoing work. Eight randomised trials in which the trial used a performance-based payment as the intervention were identified and included in the review. In four of the articles the recipient of the incentive payment was the individual provider, while in the other four the recipient was either a provider or provider group. In the four studies where the incentive targeted the individual provider, there were five positive and two negative results. In the remaining studies there were one positive and two negative results, where 'positive' indicates a result in the desired direction, and 'negative' means there was no significant effect. In seven studies the target for the incentive was the physician. In these studies there were five significant positive effects and four cases of no significant effect. Positive effects were more likely to be observed when the incentive took the form of an addition to fee-for-service payment than when the incentive was paid as a bonus. Seven studies (and nine dependent variables) addressed preventive care.

Type of study: Systematic review

Petersen *et al* (2006) reviewed the literature on studies where there was an explicit financial incentive to improve quality. They conducted a PubMed search of the English language literature from 1 January 1980 to 14 November 2005. The 17 empirical studies identified were classified according to the level of incentive (for example, physician, group, payment system) and the type of quality measure rewarded. Thirteen of the 17 studies examined process of care measures, most related to preventive care. Five of the six studies of physician-level incentives, and seven of the nine studies of provider group incentives, found partial or positive effects on quality. One study found a negative effect on care for the sickest patients. Results in four studies suggested unintended side effects. No studies examined optimal duration of incentives or their sustained impact after termination. Overall, the authors observed that there were few empirical studies of the effects of explicit financial incentives on quality.

Type of study: Systematic review

Rosenthal and Frank (2006) reviewed the literature on paying for quality in healthcare, including brief reviews of the pay-for-performance literature in other fields as well. In 2003, the authors examined the peer-reviewed empirical literature using five databases: MEDLINE, EconLit, ABI Inform, PsychInfo and the Social Science Citation Index. Additional citations were found by examining the reference lists of articles. The review focused on studies that assessed quality-based payment schemes. Studies were excluded that assessed the impact of payment systems on quality of care. The authors located seven published, peer-reviewed empirical studies of the effects of paying for quality in healthcare. Another study located by the review related to contracting for substance abuse treatment, but the rewards were not spelled out so it was excluded from the review. The authors concluded that the empirical foundations for pay-for-performance in healthcare are 'rather weak'. There were only two positive findings and studies with the strongest research designs were more likely to find no impact related to financial incentives. However, the studies were narrowly focused and tended to relate to preventive care. Their implications for more recent pay-for-performance initiatives are not clear.

Type of study: Systematic review

Scott and Hall (1995) reviewed the effects of different payment methods on GPs using a variety of measures of costs and outcomes of care. Four sources were used in searching the literature: MEDLINE, Social Sciences Citations Index, citations in articles received and citations known to authors. Studies were identified that examined actual changes in GP reimbursements or differences

in GP reimbursed in different ways. The authors did not summarise their findings across these studies. Their main conclusion was that, based on the literature, it was not possible to make recommendations about optimal payment systems. Much more research is needed. According to the authors the most 'fundamental criticism' of the literature was that it didn't say whether patients were better or worse as a result of reimbursement changes. Only one study attempted this, comparing actual practice with clinical guidelines. This study, by the Department of Health in the UK (1991), found that GPs were more likely to hit some payment targets when paid specifically to do so. However, this was a before–after study with no controls and unclear data sources.

Type of study: Systematic review

Town *et al* (2005) reviewed the impact of financial incentives on preventive care delivery. A unique aspect of the review is that it is limited to randomised trials. There were eight different financial interventions identified in the review. The incentives included direct payments or bonuses to providers, as well as more diffuse incentives. The authors searched EconLit, Business Source Premier, PsychInfo and MEDLINE. Reference lists were reviewed to identify other articles. The search focused on English language articles published from 1966 to 2002 that addressed primary or secondary prevention or health promotion. Studies using interventions with multiple components, where it was not possible to identify the effect of financial incentives, were also excluded, as were studies that compared outcomes under different payment systems. Two independent reviewers abstracted each article. Only six studies met the inclusion criteria and they generated eight different findings. Of the eight different financial incentives reviewed, only one led to a significantly greater provision of services. The authors noted that this doesn't necessarily imply that financial incentives won't motivate physicians to provide more preventive care. The incentives in the study were weak as the rewards were small. They concluded that small rewards probably won't motivate doctors to change their practices with respect to preventive care.

Type of study: Systematic review

Targeted studies of the impact of payment incentives on quality

The literature reviews described above, as might be expected, drew heavily from a core group of studies: Grady et al (1997); Kouides et al (1998); Hillman et al (1998, 1999) and Fairbrother et al (1999, 2001). The sample sizes in these studies were small to moderate, and the extent to which their findings can be generalised to other settings is questionable. Also of concern is whether the interventions in the studies could be effectively increased in scale if purchasers chose to do so. Grady et al (1997) examined the impact of financial incentives on referrals by primary care physicians of patients 50 years and older for mammograms. In what Grady et al characterised as a 'token' reward, physicians received a bonus payment on top of education and reminders, with the payment related to the percentage of patients referred. Sixty-one practices in Ohio and Massachusetts participated in the randomised study over a three-year period. The authors reported that use of chart stickers increased referrals but that feedback in combination with the bonus payment did not add to improvement obtained through the use of chart stickers alone. They speculated that the financial incentive may have been too small to be effective. Kouides et al (1998) examined the impact of paying physicians for influenza immunisations for elderly patients, with physicians receiving a payment of $0.80 per immunisation if a rate of 70 per cent was achieved and $1.60 if a rate of 85 per cent was reached. The strengths of the study, which was conducted in 1990/91, were the random assignment of physicians to a payment and control group and the fact that the financial incentive was simple, easy to understand and not combined with any other major efforts to increase immunisation rates. The authors found that immunisation rates increased by 10.3 per cent in the payment group compared with 3.5 per cent in the control group, with the average rates being 68.6 and 62.7 per cent respectively. They concluded that even modest financial incentives could be effective at improving the rate at which physicians provided influenza immunisations to elderly patients.

Two studies by Hillman et al addressed attempts to improve preventive care in the Medicaid population in the USA using physician bonus payments. In the first (Hillman et al, 1998), half of 52 primary care practices were randomly selected to receive the intervention, which combined bonus payments with written feedback relating to performance. The three intervention sites that had the highest scores in cancer screening for women received a bonus equal to 20 per cent of their capitation payment. The three sites with the next highest scores received a 10 per cent bonus, as did the three sites that improved the most. Eight other practices also received bonus payments; overall, bonuses ranged from $570 to $1260 per practice. The combined intervention of bonuses and feedback did not improve screening significantly. The authors speculated that the financial incentives may have been too small or poorly communicated to physicians. While there was no differential improvement that could be attributed to the intervention, both intervention and control sites experienced substantial improvement during the study period, presumably due to exposure to national efforts to improve preventive care. In their 1999 study, Hillman et al used a similar design to test the impact of financial incentives on paediatric preventive care in a Medicaid population. Fifty-three practices were assigned to one of three groups: feedback plus incentive, feedback only and a control group. No sites received only the financial incentive. This time, bonus payments were larger, ranging from $772 to $4682. Again, the authors found that neither intervention improved care, with only about half the practices reporting that they were aware of the interventions. These two analyses clearly underscore the difficulties in implementing an effective financial rewards programme in physician practices, including communicating with physicians about the existence and nature of the programme.

Fairbrother et al (1999, 2001) also carried out two analyses of the impact of financial incentives directed at improving preventive care, in this case relating to immunisation rates for children. In the first study, physicians were randomly assigned to one of three groups: bonus payment plus feedback on performance, enhanced fee-for-service payment plus feedback, and feedback alone. The study was set in nine deprived neighbourhoods in New York and 61 physicians participated. The results appeared encouraging, with a 25 per cent improvement in up-to-date immunisations in five immunisation categories for the bonus group compared with six improvements in the other groups. However, closer examination led the authors to conclude that the improvement was primarily due to better documentation of immunisations. The second study used a similar approach to explore whether improvements persisted over time in the bonus group and whether actual immunisation practices, as opposed to documentation, improved over time. In this study, both bonus payments and enhanced fee-for-service payments improved immunisation coverage but, again, the conclusion was that the improvement was due primarily to better documentation. The Fairbrother studies are instructive in that they underscore the importance of not taking reported changes in quality indicators at ‘face value’ when assessing the impact of financial incentives. In some cases, improvement in metrics on which payment is based may not represent actual improvements in care.

All six studies described above have several common features. They all employed a research design featuring random assignment, they all focused on preventive care and they all examined the response of relatively small numbers of physicians to financial incentives. None of these studies provided unequivocal support for the premise that the use of financial incentives for physicians is an effective way to improve performance on quality of care indicators. Possible explanations for a lack of effect focused on implementation difficulties and an incentive that may have been too small. There were several studies that were excluded from some literature reviews because they did not employ random assignment. In other studies, the measure of quality was relatively indirect, also resulting in their exclusion from review articles.

Ritchie et al (1992) used a longitudinal research design to track changes in childhood immunisation rates in a region of Scotland after the introduction in 1990 of a new GP contract that provided direct financial rewards for higher practice immunisation rates. Rates increased dramatically during the study period, but other changes occurring concurrently made it impossible to attribute the increases to the new financial

incentives. Amundson et al (2003) reported results of an analysis of a programme that was implemented within a single US HMO to pay physicians to advise patients who were smokers to quit. Results were based on an audit of medical records from 1996 to 1999. The authors reported that documentation of providing advice to quit increased from 32 to 53 per cent. The strength of the financial incentive related to this specific goal, as experienced by physicians, was not clear, and there were multiple other measures for which physicians were rewarded as well, making it difficult to draw definitive conclusions from this study. St. Jacques, Patel and Higgins (2004) analysed a multi-faceted programme directed at changing the behaviour of 31 anaesthesiologists at a university hospital. The hospital implemented a programme of profiling, reporting of performance and financial incentives, with performance tracked on five measures, mostly relating to treatment delays. Performance improved on most measures, but there was no control group, measures were indirect indicators of quality at best, the number of physicians involved was small and, because the intervention combined different approaches to physician behaviour modification, it was not possible to determine the marginal effect of the financial incentives. Ashworth et al (2005) investigated the association between physician income and the achievement of quality performance indicators for a group of 151 inner city practices in the UK. Income was assumed to be determined by number of patients, staff expenses and payments for performance received from the NHS. They concluded that achieving performance targets had little impact on incomes, but that higher staff budgets did. Based on these results, they hypothesised that the rewards for performance were offset by the higher costs of achieving higher quality. In general, the specific impact of financial incentives on quality of care in these three studies was somewhat difficult to determine due to the designs of the interventions being examined.

Amundson *et al* (2003) analysed a programme to reward physicians in a single HMO to advise smokers to quit. Physician groups received bonus payments for achieving target scores on various quality indicators, including providing advice to patients to quit smoking. The authors did not indicate the amounts received by groups specifically for improving in this area. Audits of 14,489 ambulatory patient records were undertaken in 19–20 medical groups from 1996 to 1999. Statistical tests of before–after group means were conducted. Identification of tobacco use in patient records increased from 49 to 73 per cent, and advice to quit increased from 32 to 53 per cent. The number of medical groups in which 80 per cent of patient targets were met increased from zero to eight. The impact of financial incentives by themselves on provider behaviour was difficult to determine because the strength of the incentive was not clear, the incentive at the individual physician-level (as opposed to group) was unclear and the intervention was multi-faceted.

Type of study: Controlled observational study

Ashworth *et al* (2005) conducted a multivariate analysis of the relationship among factors related to physician incomes and achievement of quality performance indicators in an inner city health authority in the UK. The income of GPs depended on the number of patients in the practice, staff expenses and payments for performance. The time period was two years before new, higher payments for performance were instituted in the NHS. Data were collected on 151 practices in an inner city health authority for 2001 and 2002. Regression analysis, including path analysis, was used to explore relationships. The authors concluded that GPs were able to maximise their incomes by taking on more patients. Achievement of performance targets had little impact. Higher staff budgets were associated with better performance on quality indicators, suggesting that the rewards for performance, which were not large, were offset by the higher costs of achieving higher quality.

Type of study: Observational study without control group

Fairbrother *et al* (1999) examined the effect of different financial incentives on immunisation coverage, specifically the percentage of children up-to-date on a variety of immunisations. Physicians were assigned to one of three groups: bonus and feedback, enhanced fee-for-service and feedback, and feedback only. Physicians were randomly assigned to the three groups and immunisations were

measured at three points in time, approximately four months apart. Nine neighbourhoods in New York City with the highest poverty rates were selected as study sites. Eighty-three physicians were invited to join the study and 61 accepted. Data were collected through chart review. Logistic and linear regression models were used to evaluate outcomes. There was a 25 per cent improvement in up-to-date immunisations in five categories for the bonus group, with no significant changes in the other groups. Much of the improvement appeared to be the result of better documentation.

Type of study: Randomised controlled trial

Fairbrother *et al* (2001) conducted a follow-up to a previous study to analyse whether bonus payments and enhanced fee-for-service improved immunisation rates for children, specifically the percentage of children with up-to-date coverage on immunisations. Bonus payments were $1000 and $2500 for 30 point and 45 point improvements, $5000 for reaching 80 per cent up-to-date coverage and $7500 for reaching 90 per cent up-to-date coverage. In the enhanced fee-for-service group, physicians received $5 for each vaccine administered within 30 days of its due date and $15 for each visit at which all due vaccines were administered. A control group received feedback. A previous study by the authors left questions unanswered including: would the improvements in a bonus group continue, would actual practices (as opposed to documentation) improve over time and would the enhanced fee-for-service group begin to have an impact? Also, the previous study focused on inner city children served under Medicaid. This study included all payers. Incentives were given to 57 randomly selected physicians in New York City four times at four-month intervals based on performance for 50 randomly selected children in their practices. Logistic regression models and linear regression models were used to analyse the data. The lower response rate in this study (compared with the authors' previous study) was a limitation. Both types of financial incentives increased documented immunisations. The authors concluded that the incentives were not sufficient to overcome entrenched physician behaviour patterns and that true immunisation coverage was higher than documented in charts.

Type of study: Randomised controlled trial

Grady *et al* (1997) evaluated the success of three different approaches designed to increase referrals by primary care physicians of patients 50 years and older for mammograms. The experiment added what the authors call a token reward for referrals to a strategy of 'cue enhancement' and education. The reward was a cheque based on the percentage referred in a given audit period (for example, $50 for a 50 per cent referral rate). In order to have a comparison group, the rewards for some were not initiated until the second year of the experiment. The study was based on a randomised trial involving 61 practices in Dayton, Ohio and Springfield, Massachusetts over a three-year period resulting in a sample of 11,426 patients. The actual years covered in the analysis are not mentioned. Various statistical techniques were employed, primarily repeated measures ANOVA (analysis of variance), to test for statistically significant differences among the groups. Chart stickers were effective in increasing referrals. Peer-performance and feedback, combined with a reward, did not increase referrals over cueing alone. The authors speculated that the reward offered may have been too small and isolated to have had an impact.

Type of study: Randomised controlled trial

Hillman *et al* (1998) evaluated a randomised controlled trial of an intervention intended to improve compliance with four preventive care screening exams for women 50 years and older, with financial incentives for physicians being part of the intervention. The three intervention sites with the highest compliance scores received a full bonus of 20 per cent of capitation. The three with the next highest

scores and the three that improved the most got a 10 per cent bonus. Bonuses ranged from $570 to $1260 a site with an average of $775 per audit. Seventeen of the 26 sites received a bonus. Half the 52 primary care sites received the intervention, which included written feedback along with the financial bonus. The study was conducted from 1993 to 1995 in Philadelphia. Tests for the significance of differences in group means were carried out. Financial incentives and feedback did not improve physician compliance. The magnitude of the incentive may have been too small or the physicians may not have been aware of the change in incentives. Both groups saw dramatic increases in preventive care during the study period reflecting national initiatives.

Type of study: Randomised controlled trial

Hillman *et al* (1999) conducted a randomised trial of two different interventions, one of which involved a financial incentive, to improve paediatric preventive care in a Medicaid population. The three practice sites with the highest total compliance scores with recommended practices received the full bonus of 20 per cent of capitation. The next three received a 10 per cent bonus, as did the three sites showing the greatest improvement, provided scores increased by at least 10 per cent. Bonuses ranged from $772 to $4682, with an average of $2000. Thirteen of 19 sites received at least one bonus and six sites received two bonuses. The purpose of the study was to determine if a system of semi-annual assessment and feedback, coupled with financial incentives, could improve paediatric preventive care guidelines as evaluated by semi-annual chart audits from 1993 to 1995. Fifty-three primary care sites in Philadelphia were assigned to three groups: feedback plus incentive, feedback only and a control group. Chart audits were performed at six-month intervals for 18 months. A statistical comparison of means was carried out. Neither intervention resulted in improved care. The authors noted that only 56 per cent of sites reported awareness of the programme despite repeated mailings.

Type of study: Randomised controlled trial

Kouides *et al* (1998) conducted an empirical evaluation of the impact of a 1990 Medicare influenza project set in Rochester (New York State), with randomisation of physicians to a control group and an incentive group. Physicians could receive an additional $0.80 per shot or $1.60 per shot if practice immunisation rates of 70 per cent and 85 per cent were achieved respectively. The study took place in 1990 and 1991. Multiple regression techniques were employed in the analysis of physician reports of immunisations. A survey of physician offices was conducted to gather data on practice characteristics. The mean immunisation rate for practices in the incentive group was 68.6 per cent compared with 62.7 per cent for the control group. The median improvement was 10.3 per cent in the incentive group and 3.5 per cent in the control group. The authors conclude that, although the financial incentive was modest, it improved immunisation rates by about 7 per cent.

Type of study: Randomised controlled trial

Ritchie (1992) tracked immunisations in a single region in Scotland before and after introduction of a new contract for primary care physicians in 1990. In this contract, 'item of service' payments were replaced by target payments to encourage GPs to increase rates of childhood immunisations. The details of the payment change were not described. In the study region, this change was accompanied by a records system that provided feedback to GPs regarding their immunisation performance. Immunisation rates for 95 practices encompassing 313 GPs were calculated for children aged two and five on the first day of each quarter for the seven quarters ending in March 1990 and subsequent three-month periods to September 1991. The analysis was retrospective and descriptive and used

data drawn from the computer records maintained by the Grampian region in Scotland. A variety of statistical analyses were conducted using linear, non-linear and logistic regression methods. The practices achieving immunisation rates of at least 95 per cent increased from 31 to 81 per cent for primary immunisations. Achievement of 95 per cent rates for pre-school booster immunisation increased from 23 to 64 per cent. The authors noted evidence of 'sustained improvement' but no change in overall trends. They suggested that the reasons for the change were complex and should not necessarily be attributed to the new contract and the change in financial incentives introduced by it.

Type of study: Observational study without control group (longitudinal)

St. Jacques, Patel, and Higgins (2004) assessed the impact of implementing a programme of physician profiling, reporting and incentives on the behaviour of anaesthesiologists. For each study month physicians were eligible to receive a variable financial payment of $0–500 depending on individual scores relative to one another. The payment was credited to the physician's personal expense account. Performance was tracked in five areas: percentage of first cases of the day at the room before or at start time, percentage of cases where preparation time was less than a target, percentage of cases delayed while waiting for anaesthesiology evaluation, percentage of cases delayed during anaesthesiologist controlled time and percentage of cases delayed while waiting for anaesthesiology attending. Thirty-one anaesthesiologists in a university hospital were tracked for six months. A statistical comparison of means was carried out. Compared to the first month, performance improved on most measures. Because the programme combined profiling with incentives it was not possible to determine the effect of incentives only. The authors did not relate their findings to patient outcomes.

Type of study: Observational study without control group

Purchaser/payer initiatives that reward physicians and other healthcare practitioners for achieving quality goals

Our literature search identified 20 evaluations of programmes designed to reward physicians for quality improvement or for achieving quality benchmarks. One evaluation was based on a randomised design (Christensen et al, 2000), while four employed control groups that were observed before and after the payment programme was implemented for the intervention group. Christensen et al (2000) studied the impact of a financial rewards programme put in place by the State of Washington Medicaid programme that paid pharmacists for providing cognitive services to Medicaid beneficiaries at the time prescriptions were filled. The assumption was that this represented better quality care for these patients. Pharmacists were paid $4 for consultations of under six minutes in length and $6 for longer consultations, as well as $40 per month for documentation costs. The number of cognitive services provided per 100 prescriptions was considerably higher in the group receiving payments (1.59 versus 0.67). Beaulieu and Horrigan (2005) reported the results from a 'natural experiment' in which a managed care organisation instituted a pilot programme that paid physicians a bonus for a 50 per cent improvement in the quality of care for their diabetic patients, or for achieving performance targets. The composite score on which the payment was based reflected both process of care and outcome measures. Significant improvement occurred in five of the six process measures and two of the three outcome measures; however, it was not possible to disentangle the impact of financial incentives from other quality improvement efforts implemented by the managed care organisation. Also, it was not clear whether differences in patient populations and physician characteristics were controlled in the analysis. Levin-Scherz, DeVita and Timbie (2006) studied the impact of financial rewards on diabetes care as well, along with impacts on asthma care. In this payment initiative, a physician network received bonus payments and a return of withholds from its managed care contract depending on performance on a variety of quality measures (with diabetes and asthma results being the only ones included in the study). There were improvements in claims-based diabetes measures from 2001 to 2003, when compared to an index plan. There were smaller improvements in asthma measures, but these measures were at a high level in the baseline year,

leaving less room for improvement. Rosenthal et al (2005) evaluated a broad-based pay-for-performance programme initiated by a managed care plan in California, with enrollees in a neighbouring state serving as a control group. Payments of $0.23 per member per month were made for each target that was achieved; medical groups enrolling 10,000 plan members had the potential to earn $270,000 per year if they achieved all performance goals. This represents about 5 per cent of the group's professional capitation from the plan for this hypothetical set of patients. While the pay-for-performance programme used multiple measures, three were included in the study: cervical cancer screening, mammography and haemoglobin testing. Relative to the comparison group there was improvement only in cancer screening. Because payment was made for achieving benchmarks, groups that started at a high level but showed relatively small improvement received most of the monetary reward distributed by the managed care plan. Felt-Lisk et al (2007) compared outcomes across five health plans that served Medicaid patients in California before and after a pay-for-performance programme was implemented for well-baby visits. They reported favourable results but did not discuss the analytic approach they used to estimate impacts. Their study used qualitative data to assess differences across the plans in implementation experience. They found that the most successful plans with respect to quality improvement paid the highest rewards and did the best job of communicating with physicians about programme details.

Seven studies, none employing a control group, examined the performance of physician pay-for-performance programmes in the UK. A very early study by Langham, Gillam and Thorogood (1995) examined changes in the distribution of financial incentive payments for health promotion, focusing on cardiovascular disease performance payments related to screening and the recording of risk factors. Before implementing the payment programme, practices had been reimbursed for holding health promotion clinics. After the shift to the new approach, funds related to health promotion were found to be more evenly distributed across practices, but practices in areas with the highest measured need lost funds relative to other practices.

Six subsequent evaluations have focused on the more recent implementation of a broad-based pay-for-performance initiative by the UK's National Health Service (NHS). Srirangalingam et al (2006) analysed how referral patterns for diabetes care changed after introduction of the new financial reward system, concluding that there was no significant impact on the total number of referrals to specialty clinics, but that there was a lower threshold for referrals. Sutton and McLean (2006) related physician performance on quality scores to a variety of practice characteristics. The most relevant finding from this study is that practices receiving more income from other sources had lower quality scores. This led the authors to speculate that the impact of the pay-for-performance incentives in these practices is weaker because they are less dependent on these funds to support these practices. Doran et al (2006) took a more comprehensive approach in assessing the impacts of the new pay-for-performance initiative. As they described this programme, the NHS allocated about $3.2bn in new funds over a three-year period to reward primary care physicians on 146 quality indicators relating to ten chronic illnesses, as well as markers of organisation of care and overall patient experience. Points were awarded, with payment limited to $133 per point in 2004/05. In 2004/05, the maximum a physician could receive was $139,400. The NHS programme is much more ambitious in its scope and funding than any pay-for-performance effort implemented to date in other countries, and the Doran et al analysis focused on the first year evidence regarding factors influencing performance of physicians relative to indicators. They found that the median practice achieved 95.5 per cent of the available points (75 per cent was anticipated), with exception reporting having the greatest impact on a practice's performance – that is, physicians who were able to exclude larger proportions of patients from the payment calculations (as allowed under programme regulations) scored higher than other physicians. On average, gross practice income grew by $40,200 under the first year of the programme. Because practice quality was improving prior to the programme, and no baseline data were available for the analysis, the authors hesitated to attribute changes in performance directly to the use of financial incentives, even in light of the substantial payouts that occurred. Campbell et al (2007) addressed the lack of baseline data in their study, focusing on a set of 42 primary care practices for which data on care for asthma, diabetes and coronary heart disease were collected twice prior to the pay-for-performance programme and after the first year of

implementation. They found a significant but modest increase in the trend rate for quality improvement in asthma and diabetes care after implementation of pay-for-performance. However, this improvement in the trend also was observed for quality indicators not included in the pay-for-performance reward structure. They concluded that pay-for-performance could make a 'useful contribution' to quality improvement as part of a comprehensive quality improvement strategy. In a study based on stroke patients in Scotland, Simpson et al (2006) reported that the pay-for-performance programme for GPs improved treatment documentation. In an ethnographic study of two GP practices, McDonald et al (2007) reported that the financial incentives in the pay-for-performance programme in England did not damage the internal motivation of GPs, nor did physicians question the quality targets or their implications.

Seven of the remaining nine studies relating to efforts to reward physicians for quality were set in the USA. As with the UK analyses, they did not employ control groups. Unlike the UK studies, most examined the impact of relatively limited programmes or sets of programmes. The first of these studies, by Morrow, Gooding and Clark (1995), analysed the impact of a multi-faceted intervention designed to improve preventive healthcare provided by member physicians of an independent practice association. Chart audits conducted from 1987 to 1990 demonstrated improvements in virtually all preventive measures. Because the intervention included non-financial, as well as financial, components, it was not possible to disentangle the impact of payment from other factors that might have influenced the findings. In 2003, Chung et al published findings from a qualitative and quantitative assessment of a physician recognition programme implemented by a health plan in 1997. During the first years of the programme, physicians received an average incentive payment of 3.5 per cent of reimbursement, with an additional bonus of $3000 added in 2001 for physicians showing improvement in their scores. Points were calculated for performance on quality, patient satisfaction, business operations and process of care measures. The authors reported results for only three indicators, with improvement demonstrated in ACE inhibitor use and haemoglobin Alc testing for diabetic patients. There was no control group used in this study, suggesting caution needs to be applied in attributing these improvements to the financial reward programme.

Larsen, Cannon and Tower (2003) examined the impact of a disease management effort, which had financial incentives as one component, implemented by an integrated delivery system. The financial incentive represented 0.5 to 1 per cent of physician compensation, and about half the incentive was related to diabetes care. Significant improvement is reported on clinical measures, but there is no control group and it is not possible to attribute improvement to the financial incentives. Greene et al (2004) evaluated a health plan programme (combining education, profiling and financial incentives) to improve physician adherence to treatment guidelines for sinusitis, where 40 per cent of the financial reward was based on quality. The incentive was a reduction in the amount of physician payment normally withheld from the physician, with the top 5 per cent of physicians receiving a reduction from 15 to 10 per cent. There was a reduction in use of ineffective antibiotics, but it was not possible to determine the impact of financial incentives alone on this change. Ettner et al (2006) analysed the association between reimbursement incentives in ten managed care plans from 2000/01 and a variety of measures of treatment of diabetes. They found that care processes were better for providers that were reimbursed on a salary basis, with quality and satisfaction scores determining a portion of physician payment. In this study design it was not possible to determine if financial incentives resulted in better performance or if use of particular payment approaches was more common in situations where physicians performed better on the metrics. Also addressing diabetes care, Curtin et al (2006) calculated the return on investment, from a health plan perspective, of a pay-for-performance programme. Savings in medical care costs for patients with diabetes were compared to the cost of the pay-for-performance programme, with the authors concluding that the programme had a favourable return on investment ranging from 1.6 to 2.5, even though better compliance with treatment guidelines required that physicians provide more services in most cases. Finally, Damberg et al (2005) described the early experience of the Integrated Health Association, a collaboration of health plans and medical groups in California, in implementing a pay-for-performance programme. Improvements were reported in the first programme year on four or more quality indicators for three quarters of the medical groups. However, analytic details were not presented as this study focused primarily on design and implementation issues.

We found only two studies of pay-for-performance efforts conducted outside of the UK or US. Chiang et al (2002) described changes in reporting of tuberculosis (TB) in Taiwan from 1995 to 1999, where clinicians and hospitals received payments for each case reported. Results were compared to reporting prior to implementation of the payment programme. The authors found a 47 per cent increase in reporting in the first year that the programme was implemented. In the two years after that, the number of cases reported declined slightly. The authors stated that the first year increase was significant, but no description of statistical methods was provided and there was no contemporaneous control group used in the study. Gene-Badia et al (2007) examined the impact of a financial incentive programme that rewarded primary care teams in Catalonia, Spain for achieving a set of clinical objectives and also for participating in a professional development programme. They found limited effects on the quality of professional life and on patient satisfaction.

Beaulieu and Horrigan (2005) estimated the impact of a managed care organisation's programme, which combined financial incentives and practice support, on the quality of diabetes care. Physicians who met targets or demonstrated significant improvement received a bonus. The largest payment was equivalent to a 12 per cent increase in their per member per month payment (true for both fee-for-service and capitated physicians). Actual payments ranged from $3000 to $12,000. Data on patient outcomes were self-reported by physicians, with limited audits of medical charts. The control group data were collected as part of the health plan's HEDIS (Healthcare Effectiveness Data and Information Set) reporting. Analysis consisted of statistical comparison of group means before and after the programme. There were significant improvements on five of six process measures. Thirteen of 21 physicians received a financial award. Of the eight not receiving rewards, six improved their scores. There was no evidence that quality declined in areas of care not being measured. Self-selection of physicians into the pilot programme and the small sample size limits the ability to generalise from the results. It also is impossible to determine the marginal effect of the financial incentive because it was implemented along with other practice supports for diabetes care.

Type of study: Controlled observational study

Campbell *et al* (2007) assessed trends in quality of care indicators in physician practices in England before and after introduction of a pay-for-performance programme for GPs in 2004. GPs received payments based on the number of points they garnered in the course of a year. Points were awarded for practice structures supporting quality, process of care measures and access measures. Data from 1998, 2003 and 2005 were collected for 42 primary care practices in England for clinical indicators associated with coronary heart disease, asthma and Type 2 diabetes. Trend analysis was conducted for indicators that were eligible for reward under the pay-for-performance programme and also for some indicators that were not. Several different statistical methods were used to test the robustness of the findings. There was a statistically significant, but modest, increase in the trend rate for asthma and Type 2 diabetes indicators after the introduction of the pay-for-performance programme. The lack of a significant increase for coronary heart disease could be due to the fact that scores on these indicators were high prior to the pay-for-performance programme. There was no difference in the trend rates for indicators subject to pay-for-performance and for those that were not. The authors suggest that their analysis may underestimate the impact of pay-for-performance as practices may have implemented some changes in 2003 in anticipation of programme implementation. The lack of a difference between the trends for the two groups of indicators suggests that increases may not be due to pay-for-performance. Alternatively, practitioner attempts to improve scores on pay-for-performance indicators could have had a beneficial 'spillover' effect on other non-measured components of quality. The authors conclude that their results support the view that pay-for-performance can 'make a useful contribution to improving quality' as 'part of a comprehensive quality improvement program' (p 189).

Type of study: Observational time series

Chiang *et al* (2002) described changes in reporting of TB in Taiwan from 1995 to 1999. Clinicians and hospitals received NT$250 for each confirmed case of TB reported. The authors plotted the number of reported cases from 1995 to 1999. The payment for reporting began in 1997. Changes between various reporting periods (six months) were calculated. There were no tests of significance reported. The incentive programme appeared to have its intended effect. There was a 47 per cent increase in reporting the year that the programme was instituted. However, respectively in 1998 and 1999 the number of reported cases declined slightly (7 per cent and 3 per cent). The study found an impact that was attributed by the authors to incentives. The result was probably significant, but no tests were performed, nor were any data presented concerning the nature of the increased number of reports.

Type of study: Controlled observational study

Christensen *et al* (2000) carried out an evaluation of an intervention among pharmacists in the State of Washington that involved a financial incentive for providing cognitive services to Medicaid recipients at the time prescriptions were filled. Compensation was $4 for interventions up to six minutes and $6 for longer consultations. All pharmacies also received $40 per month for documenting the cognitive services they provided. Pharmacies were randomly assigned to a study (110) or control (90) group. Cognitive services documentation was audited for completeness and consistency. There was a significant difference in the number of cognitive services per 100 prescriptions (1.59 versus 0.67) and 75 per cent of consultations were less than six minutes. The authors do not state expectations directly, but imply that more consultative services results in better quality care, especially in improving patient safety.

Type of study: Randomised controlled trial

Chung *et al* (2003) conducted a qualitative and quantitative assessment of a physician recognition programme employed in the Hawaii Medical Service Association. Physicians received points for achievement relative to quality indicators, patient satisfaction, business operations and utilisation of services. Physicians were ranked and the average incentive reward ranged from 0 to 5.5 per cent, with an average of 3.5 per cent. In 2001 a bonus of up to $3000 was added for practitioners who improved scores. There were payment caps to avoid higher payments to high-fee specialists. Administrative claims data were used to measure the quality indicators and utilisation. A survey was used to collect data on patient satisfaction. Non-parametric tests of statistical significance were conducted for the years 1998–2001. The programme started in 1997. The authors reported results on a subset of indicators (n=3). These are common measures but there is no explanation for why they were chosen for use in this particular case. There was improvement in ACE inhibitor use and in haemoglobin A1c testing. The results were mixed regarding improvement for an immunisation measure, a finding the authors attribute to external factors. There was no control group, so it is not clear if the improvements were due to the compensation programme. Only a subset of results is provided.

Type of study: Observational study without control group

Curtin *et al* (2006) analysed the cost savings from a pay-for-performance programme directed at physicians providing diabetes care. Payments from a health plan to an individual practice association (IPA) withheld dollars which were then returned to the IPA if it met target performance levels. Each year about $15m of these withheld funds were distributed to 3700 participating physicians, specialists as well as generalists. An average primary care physician's distribution ranged from $6000 to $18,000 annually across all performance measures. Diabetes care was one component of the overall performance score on which payout was based. Historical trend data (2000–02) were used to estimate what the costs of care would have been for diabetes patients in 2003/04 in

the absence of the pay-for-performance programme, and this was compared with the cost of the diabetes programme. Claims data provided by the health plan were used in the analysis. Savings were calculated from the perspective of the health plan. The authors found a positive return on investment of 1.6 to 1 in 2003 and 2.5 to 1 in 2004. The most significant cost reductions occurred in the area of hospital care. The authors pointed out that in most instances the pay-for-performance programme essentially rewarded physicians for providing more care for their patients with diabetes, presumably adding to direct treatment costs. Thus, the positive rate of return was more impressive than if achieving the performance goals had required no additional treatment or reductions in treatment.

Type of study: Observational

Damberg *et al* (2005) presented early descriptive findings of the Integrated Health Association's pay-for-performance initiative, with discussion of design and implementation issues. Health plans used a common set of measures drawn from HEDIS to reward physician groups for performance, with public report cards distributed at the same time. Improvements in measures were reported. Tests of significance were referred to but specific results were not provided. Data were from the first reporting year (2003). There was significant improvement on at least four clinical measures for three quarters of the reporting groups. This article focused more on design and implementation issues than on analysis of improvements in quality measures.

Type of study: Observational

Doran *et al* (2006) examined the first year experience of family practice doctors in the UK in achieving targets under the NHS's new pay-for-performance scheme. In 2004 the NHS committed about $3.2b in new funding for three years for a pay-for-performance programme for family practice doctors. Physicians were rewarded for their performance on 146 quality indicators relating to clinical care for ten chronic diseases, organisation of care and patient experience. Points were awarded on a sliding scale within a payment range, with payment limited to $133 per point awarded in 2004/05, adjusted for disease prevalence. In that period, the maximum that a GP could receive from the programme was $139,400. Data were extracted from a national computer database. Data for exception reports was imputed. Linear multiple least-squares regressions with robust estimates of error variance were used to estimate relationships. Fixed effects for practice location were used. The median practice achieved 95.5 per cent of the points available, in comparison to an expected 75 per cent. Achievement was higher in practices with a high ratio of family practitioners to patients, but all significant effects were small and only 20 per cent of the variance was explained by the regression models. The factor with the greatest effect was exception reporting. Physicians who excluded a large proportion of patients from the calculations performed better. The programme increased the gross income of physicians by an average of $40,200. There were no baseline data in the UK to use in the analysis, but there was evidence that quality was improving prior to the programme.

Type of study: Observational study without control group

Ettner *et al* (2006) estimated the association between reimbursement incentives in 10 managed care plans and process measures for quality of care in diabetes treatment. The incentives faced by physicians were measured by proportion of compensation received from salary, capitation, fee-for-service and performance-based payment. A variety of performance measures were used, including receipt of dilated eye exams, foot exams, influenza immunisations, advice to take aspirin, and assessments of glycaemic control, proteinuria and lipid profile. Data were gathered in 2000 and

2001 through patient, provider groups, and health plan surveys and medical records reviews for 6,194 patients with diabetes. The analyses employed multi-level logistic regression techniques with random intercepts for provider groups and health plans. The most significant analytic problem related to high correlation between the payment variables and organisational type. When organisational type was not controlled for in the analysis, care processes were better when physicians were paid on a salary basis, and when quality/satisfaction scores were used to determine a portion of physician payment. The results were confounded by organisational type. Nevertheless, the authors concluded that, regardless of causality, use of quality/satisfaction scores to determine physician compensation 'may indicate delivery of high quality care for diabetes' (p 1222).

Type of study: Observational study without control group

Felt-Lisk *et al* (2007) studied a Medicaid pay-for-performance demonstration involving contracting health plans in California. Providers were rewarded for achieving benchmarks for well-baby visits in the Medicaid population. Four of five plans offering new incentives offered bonuses to contracting entities based on the number of children who met well-baby visit guidelines. The fifth made payments directly to physicians using an existing bonus pool. A difference in difference analysis was used to evaluate impacts where data permitted. Qualitative analysis was used to contrast the approaches taken by the Medicaid plans and the difficulties they encountered. Data covered the period from 2002 to 2005, with the payment years being 2003 to 2005. There were favourable trends overall in the number of well-baby visits, however the experience of the five plans in the study varied. The more successful programmes had greater rewards for providers and had better communication with providers about the incentive programme. There was little information provided in the article relating to the methodology used to estimate quantitative programme impacts.

Type of study: Quasi-experimental design with qualitative implementation analysis

Gene-Badia *et al* (2007) assessed whether implementation of an incentive scheme to improve quality and aid professional development had an impact on quality of professional life and patient satisfaction. Survey data were collected from 257 primary care teams and their patients in Catalonia, Spain in 2002 and 2003. Multivariate regression techniques were used to analyse the impact of financial incentives on 34 measures of quality of professional life and patient satisfaction with care and care facilities. Perception of support from management increased but so did perception of demands on health professionals. There was little evidence of an effect on patient satisfaction.

Type of study: Observational

Greene *et al* (2004) evaluated a health plan programme to increase physician adherence to treatment guidelines for acute sinusitis in an IPA in Rochester (New York State). A scoring system was developed based on 20 per cent patient satisfaction, 40 per cent efficiency and 40 per cent quality measures. From 1999 to 2001 the percentage withhold was 15 per cent. In 2000, the withhold was reduced to 10 per cent for the top 5 per cent of performers and increased to 20 per cent for the bottom 5 per cent. Episodes of care were identified for acute sinusitis among 420,000 HMO patients between 1999 and 2001. Statistical process control charts were used to analyse changes over time, with statistical tests of the magnitude of the observed changes. The 'exception rate' decreased by 20 per cent, with most of the change being a decreased use of ineffective antibiotics. Given the multiple interventions involved, it was not possible to determine the contribution of financial incentives to the change.

Type of study: Controlled observational study

Langham, Gillam and Thorogood (1995) examined changes in the distribution of health promotion financial incentive payments after a programme was implemented in the UK in 1993 that focused payments on cardiovascular disease. Payments were associated with the performance of screening and the recording of risk factors. Previously, GPs were paid for holding health promotion clinics. The study examined the distribution of health promotion payments between health services authorities and between general practices. The retrospective study of payments included the periods before and after the change in payment approach. Payments were analysed for 78 practices in one authority and 85 in another. Changes in payments were calculated for two measures of relative need. Statistical comparisons of means were conducted. Health promotion payments were found to be more evenly distributed after the change. Practices in areas with the highest need lost more. In general, the resulting distribution was unrelated to need or treatment given after the change.

Type of study: Controlled observational study

Larsen, Cannon and Towner (2003) assessed the impact of a disease management process developed by an integrated health system (Intermountain Healthcare – IHC) on diabetes care. The financial incentive was part of a broad-based care improvement effort that included many components. The incentive was not described in detail, but appeared to be relatively small, representing 0.5 to 1 per cent of physician compensation. About half of that incentive related to diabetes care. The authors reported improvement on the key performance measures that were significant and clinically important. Because of the broad-based nature of the programme, it was not possible to determine the impact of the financial incentives by themselves. However, the incentive was small, and therefore it does not seem likely that it was a major influence on behaviour.

Type of study: Observational study without control group

Levin-Scherz, DeVita and Timbie (2006) analysed pay-for-performance contracts with physicians for diabetes and asthma care in Massachusetts. The incentive in the programme was applied at the network level. There was a withhold in provider contracts, often at 10 per cent of fees. In some cases there was the opportunity for bonus payments beyond the fee schedule. Withholds were returned or bonuses earned depending on performance relative to agreed targets. There was a variety of metrics, but the article focused only on performance on diabetes and asthma care. The analysis used claims data taken from the network's multi-year data warehouse. A difference-in-difference analysis was used for the state comparisons. The years included in the study were 2001 to 2003. There was improvement in the network's diabetes measures relative to the index plan in the state and relative to national plans. There was also improvement in asthma measures, but performance in this area started at a relatively high level. The authors noted problems in using claims data to reward performance and the network was looking forward to having access to patient electronic medical records.

Type of study: Observational study without control group

McDonald *et al* (2007) conducted an ethnographic study to assess the impact of the NHS pay-for-performance programme on practice organisation, clinical autonomy and internal motivation of GPs and nurses. Data collection took place in two practices in deprived parts of north-west England. These practices had reputations for high quality care and high scores in the first year of the pay-for-performance programme. Observation was combined with interviews, informal conversations and

document review. The authors concluded that implementation of financial incentives did not damage internal motivation of GPs, although nurses expressed more concern. Most GPs did not question the quality targets or their implications for clinical quality.

Type of study: Qualitative

Morrow, Gooding and Clark (1995) studied associations between a multi-faceted intervention and improvements in the preventive healthcare behaviours of physicians in an IPA. The only information provided by the authors regarding financial incentives was that a good score on preventive services increased reimbursement for physicians in the health plan. Chart audits of practices in a four state area were conducted from 1987 to 1990 (the number of practices was not provided). Confidence intervals were calculated. There were improvements in virtually all of the preventive measures. The authors observed that they could not necessarily attribute the improvements to the plan's programmes, including financial incentives, as there were confounding motivations for change in physician behaviours.

Type of study: Observational study without control group

Rosenthal *et al* (2005) evaluated the impact of a physician pay-for-performance programme implemented by a health plan. Beginning in July 2003, participants received a quarterly bonus of $0.23 per member per month for each performance target met or exceeded. The overall potential for a group with 10,000 health plan patients was $270,000 per year. This represented about 5 per cent of professional capitation paid by the plan and about 0.8 per cent of the group's overall revenue. The evaluation focused on three process measures of clinical quality: cervical cancer screening, mammography and haemoglobin testing. Within the plan, some medical groups received pay-for-performance payments, while groups in another region did not. Data on performance were available before and after the programme was implemented. Generalised least squares techniques were used to estimate a difference-in-difference model. Compared with the groups not receiving a pay-for-performance payment, the groups receiving payment demonstrated greater improvement only in cervical cancer screening. Because payment was made for achieving benchmarks, groups that improved the least, because they started out at a high level, received the most bonus money.

Type of study: Quasi-experimental study

Simpson *et al* (2006) analysed of the impact of a new payment scheme for GPs on recording of quality indicators for patients with stroke. The new payment system, introduced in Scotland in 2004, provided payments to practices that developed an accurate register of stroke patients and for the recording of smoking habits, blood pressure and cholesterol levels. There were also payments for reaching blood pressure control targets and other outcomes. Retrospective data from 310 (self-selected) of Scotland's 850 practices were obtained from a central database in 2005, including data for one year before the new incentive system was introduced and one year after. Binary logistic regression was used to calculate odds ratios for recording of data. Documentation increased from 32.3 to 52.1 per cent. There was a large increase among the oldest patients and most affluent patients. Women had larger increases in documentation than men. The authors noted that inequitable recording still persists, with lower recording for women, older patients and more deprived patients.

Type of study: Observational, before–after

Srirangalingam *et al* (2006) conducted an empirical analysis of how referral patterns for diabetes care changed after introduction of the new financial reward system in the UK. Under the new general medical services contract for primary care in the UK, primary care physicians receive financial rewards for performance on diabetes-related quality indicators. Referrals from primary care to a hospital-based diabetes service before and after implementation of the new incentive system were tracked. The study setting was a deprived area of London. Referrals were tracked from November 2003 to November 2004. Statistical tests of significant differences at the 0.05 level were carried out. There was no significant impact on the total number of referrals to the specialty clinic, but there was a significant increase in referrals for poor glycaemic control. The authors concluded that the new contract led to an increase in referrals for patients with unacceptable glycaemic control along with a lower threshold for referrals.

Type of study: Observational study without control group

Sutton and McLean (2006) assessed factors related to quality scores under a new UK primary medical care contract that pays GPs in part based on quality measures using a relatively complicated formula that the authors do not describe. Data were analysed for 60 practices in two NHS areas in Scotland serving a population of 367,000. Linear regression analysis was used to relate quality scores to various characteristics of the population, GP and GP's practice. The most relevant finding is that practices with higher incomes from other sources had lower quality scores. The authors speculate that the incentive effect of the new contract is weaker when income from other sources makes up a larger portion of practice income.

Type of study: Observational study without a control group

Purchaser/payer initiatives that reward institutions for achieving quality goals

Seven studies addressed the responses of institutional providers to payments for improving quality or reaching quality targets. The earliest of these studies examined impacts of bonus payments to 21 public emergency departments in Victoria, Australia (Cameron, Kennedy and McNeil, 1999). The departments received bonuses at the beginning of each fiscal year and were required to return portions of the payments if they did not achieve targets. Initial payments in 1995 were AUD$7.2m (in total), increasing to AUD$17m by 1997/98. Performance improved on all three metrics, and improvements were sustained for the entire study period. The authors attributed the apparent success of the programme in part to the fact that it was designed collaboratively with local providers of emergency care.

The remaining six studies were set in the US, with three focusing on the CMS/Premier Hospital Quality Initiative. Under this initiative, incentives were paid to approximately 270 non-profit hospitals according to their performance on 34 quality indicators. Hospitals in the top 10 per cent received a 2 per cent bonus in their Medicare payments, while those in a second tier received a 1 per cent bonus. Targets were also set for the lowest 10 per cent and 20 per cent of hospitals, which had their payments reduced by 2 per cent and 1 per cent respectively if they did not achieve targets by the end of the third year. A study by Grossbart (2006) analysed performance improvements for hospitals in a single hospital system in Ohio that participated in the initiative. Hospitals in this system that did not participate were used as a control group. Performance for 28,925 patients in three areas was tracked: acute myocardial infarction (AMI), heart failure and pneumonia. The result was slightly greater improvement on the quality measures for the hospitals participating in the initiative. Lindenauer et al (2007) assessed the impact of the initiative on 213 not-for-profit hospitals that were participating in a public reporting effort as well, comparing improvement with a group of hospitals that participated in public reporting but were not part of the pay-for-performance programme. They reported improvement in both groups of hospitals with respect to composite measures of quality. However, there was significantly greater improvement in hospitals that were part of pay-for-

performance. These results held under a variety of different statistical specifications, but improvement was somewhat smaller when results were corrected for possible 'volunteer bias' due to the self-selection of hospitals. While statistically significant, differences in improvement for the two groups of hospitals were relatively small, leading the authors to speculate regarding whether the benefits of the programme exceeded its costs. The bulk of the payments were received by hospitals that scored the highest at baseline with respect to the quality indicators. Glickman et al (2007) also assessed the impact of the CMS Premier Quality Initiative, focusing their attention on a group of hospitals that were participating in a quality improvement collaboration around care for AMIs. Of these 500 hospitals, 54 were also part of the quality initiative. The authors compared improvements in quality for the two groups of hospitals, focusing on six different process of care measures for AMI that were rewarded by CMS and eight that were not. They found no significant differences and concluded that the pay-for-performance effort did not improve care, but that it also did not have an adverse impact on care that was not rewarded.

Two other studies focused on hospital pay-for-performance efforts in different states. Nahra et al (2006) estimated the quality adjusted life years (QALYs) gained by patients hospitalised for heart treatment, compared with the money spent by a health insurer in performance payments. In a programme involving 85 hospitals in Michigan, an insurer paid $22 million over four years to supplement its diagnosis related group (DRG)-based reimbursements, with the maximum additional payment ranging from 1 to 2 per cent from 2002 to 2003. The cost per QALY gained was between $13,000 and $30,000 which, the authors noted, is well under the consensus value for a QALY, suggesting that the insurer's pay-for-performance initiative was cost effective. Nalli et al (2007) conducted a primarily qualitative study of a collaborative hospital–employer pay-for-performance programme in Maine. Hospitals received payments based on 22 measures encompassing patient satisfaction, patient safety, clinical effectiveness and efficiency. Six of ten hospitals received payments but no data were provided regarding improvement on quality measures.

One study examined the impact of a quality improvement effort in Pennsylvania aimed at reducing bedsores among nursing home residents. Each nursing home staff member received $75 if the targeted reduction in pressure ulcer incidence was achieved. Video education and management feedback were part of the intervention, along with these payments. In the period during which the programme was being implemented there was a reduction in pressure ulcers, but the effect disappeared in subsequent years. The intervention was not designed in a way that allowed isolation of the financial incentive effect.

Cameron, Kennedy, and McNeil (1999) analysed the impact of a programme of bonus payments for 21 hospitals for improved provision of emergency services. Beginning in 1995, 21 public emergency departments in Victoria, Australia were given bonus payments at the beginning of each fiscal year. They were required to return portions of the bonus if they were unable to meet targets for emergency care. The payments started at AUD$7.2m in total, and increased to AUD$17m by 1997/98. The targets related to areas of performance such as ambulance bypass, waiting time for patients with different levels of emergency and access block (patients waiting more than 12 hours for admission to a hospital). The authors used regression analysis to examine performance on the set of payment measures for two years before and three years after the bonus programme was initiated. The data were self-reported by the study hospitals and not audited. There was no explanation regarding how the authors specified the regression equations and carried out their statistical tests. The authors found that performance improved in all areas. All the results were significant except for the reduction in access block. These results were sustained over the three-year post-intervention period. The authors attributed the success of the incentive programme in part to the fact that it was developed collaboratively with local providers of emergency care.

Type of study: Controlled observational study

Glickman *et al* (2007) analysed whether a hospital pay-for-performance programme implemented by Medicare improved care for patients with AMI. Hospitals in the two highest deciles of performance received a reimbursement bonus while those in the lowest decile risked future financial penalties under Medicare's Hospital Quality Incentive Demonstration, which began in 2003. In the first two years, payments totalling $17.55 were made across five clinical conditions, one of which was AMI. In the first year, 123 hospitals received payments; 115 received them in the second year. Data were used for 500 hospitals already participating in a quality improvement initiative (CRUSADE); 54 of these were in the Medicare pay-for-performance initiative, allowing for the creation of a control group of 446 hospitals. Data covered a period before and after the pay-for-performance demonstration. Each hospital collected data and submitted it to a central database. Six different processes of care measures were evaluated as the primary outcome measures. The study also included eight measures of care that were not included in the measures Medicare rewarded as part of the demonstration. There were slightly higher rates of improvement for two of the six measures rewarded by Medicare: aspirin at discharge and smoking cessation counselling. There was no significant difference in a composite that included all six measures, nor was there any significant difference in a composite consisting of all measures not rewarded by Medicare. The hospitals in the analysis were all volunteers and were already committed to improving treatment for patients with AMI. The authors concluded that, while there was no evidence of improvement due to pay-for-performance, neither did they find any adverse effects.

Type of study: Quasi-experimental

Grossbart (2006) evaluated the impact of the CMS (Centers for Medicare and Medicaid Services) demonstration project on performance improvement in hospitals, using hospitals in a single multi-hospital system. Under a three-year demonstration programme instituted in 2003, 278 hospitals were given financial incentives based on 35 quality measures in five clinical areas. For each clinical area, hospitals with composite scores in the top 10 per cent received a 2 per cent bonus payment on top of normal payments. Hospitals in the second decile received a 1 per cent payment. There was a slight downside risk in the third year for hospitals that did not perform above threshold quality scores. The setting was Catholic Healthcare Partners, which has its headquarters in Ohio. Some of its hospitals participated in the pilot, while others did not; hence, these acted as a control group. Analysis was limited to three of the five clinical areas: AMI, heart failure and pneumonia. Performance in the first year (2004) was compared with the previous year using composite scores. The study was based on care provided to 28,925 patients. Data were obtained from the database of the hospital system. A comparison of mean values was conducted. The pace of quality improvement in the pilot hospitals was found to be slightly greater than in the control group.

Type of study: Controlled observational study

Lindenauer *et al* (2007) assessed the impact of a Medicare hospital pay for performance initiative on four composite measures of quality of care. Hospitals performing in the top decile on 33 quality measures relating to five conditions received a 2 per cent bonus payment. Those in the second decile received a 1 per cent bonus and hospitals not performing above the level of hospitals in the lowest two deciles (established in the first year) were penalised from 1 to 2 per cent. Bonuses averaged $71,960 per year. The set of hospitals in the study included 613 hospitals that voluntarily reported information about quality of care through a national public-reporting initiative; 207 of these also were participants in the Medicare pay-for-performance demonstration. Changes in performance were compared for the two groups of hospitals, using multivariate methods to control for differences in hospital characteristics. After adjustment for hospital characteristics and baseline performance, pay-for-performance was associated with improvements from 2.6 to 4.1 per cent over two years. The main share of bonus payments went to hospitals with the highest performance at baseline, but hospitals at all levels of baseline performance improved. The authors view the improvements as modest and acknowledge that the hospitals volunteered and that their attempt to control for 'volunteer bias' may not have been entirely successful. In analyses that did attempt to control for this possible bias, effects were smaller.

Type of study: Quasi-experimental

Nahra *et al* (2006) estimated the QALYs gained in a patient population hospitalised for heart treatment, relative to the money spent by a health insurer in pay-for-performance payments to hospitals. A variety of assumptions needed to be made to generate the estimates in the paper. Eighty-five hospitals in Michigan received about $22 million in a four-year period. The hospitals were paid for achieving minimum levels of compliance with accepted clinical standards for two heart conditions. The incentive payments were add-ons to the hospital's DRG-related payments from BCBS of Michigan (a national health insurer). The maximum add-on for heart care was 1 to 2 per cent between 2000 and 2003. Thresholds for receipt of payment were increased from year to year to encourage continuous improvement. The authors translated the measured improvements into estimated years of life gained relative to cost of the programme. Data on costs were collected from BCBS. Process measures were collected over a four-year period from 2000 to 2003 based on hospital self-reports. The authors used estimates from the literature to convert the process improvements into estimates of QALYs gained. The cost per QALY was estimated to be between about $13,000 and $30,000, which the authors observed is well under the consensus measure for value of a QALY, indicating that the initiative was cost effective.

Type of study: Observational study without control group

Nalli *et al* (2007) conducted a descriptive study of a hospital pay-for-performance programme implemented in Maine in 2005. Hospitals received payments from a fund established by the hospitals for reaching an agreed performance level and then bonus payments from employers based on performance against 22 measures encompassing patient satisfaction, patient safety, clinical effectiveness and efficiency. Qualitative data were collected from programme participants and data on distribution of funds were collected from secondary sources. Six of the ten participating hospitals received payments totalling $89,645. The participants believed that the programme led to care improvements in their hospitals. The programme was not continued, but it was expected that health plans would use the measures developed by the programme in their pay-for-performance efforts.

Type of study: Qualitative (primarily)

Conclusions

In this chapter we reviewed the relatively limited number of studies that assessed the effectiveness of financial incentive programmes intended to improve quality of care. In some studies there was improvement in selected outcome measures. However, in many of the studies financial incentives were combined with other quality improvement efforts, making it impossible to determine the contribution of the incentives in achieving the reported results. This limitation will no doubt be frustrating to policy-makers, payers and managers who wish to use this literature to make evidence-based decisions. Even where there is relatively more research to draw on, such as the use of financial incentives to improve the delivery of preventive services by physicians, the few studies with strong research designs find small, if any, effects on payments to providers.

The findings from broad-scale studies on the effect of payer initiatives that reward providers for quality improvements or the attainment of quality benchmarks are mixed. Relatively few significant impacts were reported, and it was often the case that payer programmes included components in addition to incentive payments, again making it difficult to assess the independent effect of the financial incentives. Insufficient research has been done on the impact of direct payer/purchaser payments to hospitals and long-term care facilities to draw any conclusions.

There are large pay-for-performance programmes underway in the USA and the UK, with more evaluations likely to appear in the peer-reviewed literature in future. Because of the variation in the way in which these programmes have been designed and implemented, synthesising their findings to provide useful guidance to decision-makers will be challenging (Dudley, 2005). It will be especially important to have comprehensive reporting of results in future studies (not limiting results to a subset of quality measures rewarded by payers), accompanied by complete descriptions of study context and possible confounding factors. In the meantime, policy-makers can support and learn from process evaluations of ongoing pay-for-performance efforts, especially by paying particular attention to accurate documentation of costs as well as continued tracking of outcomes.

3. Financial incentives and secondary impacts on quality

The research evidence suggests that financial incentives can influence provider behaviours in ways that alter patterns of utilisation and, depending on their design, reduce costs of care (see Appendix 1). Typically, these studies explored situations in which incentives were restructured with the objective of lowering utilisation of health services and, ultimately, the funds spent on them. The effectiveness of financial incentives at reducing utilisation raises questions about whether the lower levels of utilisation compromise quality of care. This is a difficult issue to address in part because quality in healthcare can be a difficult concept to operationalise. The US Institute of Medicine has defined quality of care as 'the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge' (Lohr and Schroeder, 1990).

Historical context

In environments where providers are paid primarily using fee-for-service, it typically is assumed that there is a substantial amount of excess or unnecessary service use and that reductions in this use are not likely to jeopardise quality in the sense that they would reduce the probability of achieving desired health outcomes. Furthermore, there has been a belief expressed in many quarters that providers, in their roles as decision-making agents for their patients, have ethical standards that make them unlikely to compromise quality of care to secure financial reward. However, as payment approaches based on some form of capitation or prospective payment became more widespread in the USA during the 1980s and 1990s – because of growth in HMO enrolment and implementation by the Medicare programme of a prospective payment system – concern about impacts on quality of care increased. This was reflected in an increasing number of studies that either combined utilisation, cost and quality outcomes in their designs, or focused exclusively on quality or patient outcome measures.

In its early stages the development of this line of research was limited by the rate of progress in developing process measures of quality that:

- had a clear scientific basis
- could be collected in a consistent manner across different types of payment approaches
- were likely to be sensitive to changes in financial incentives over the (typically) restricted time period encompassed by most studies.

At times, the measures used to indicate differences in quality across payment arrangements were somewhat controversial. Measurement of patient outcomes often was limited to comparisons of mortality rates under different payment arrangements or, when inpatient care was the focus, hospital readmission rates.

As the field of quality measurement has progressed so has the breadth and sophistication of the measures used to assess the impact of financial incentives on quality. Many of these measures evolved from collective negotiations between HMOs and large employers in the USA in the early 1990s. The quality indicators established in this process (termed HEDIS measures) have been used to track quality of care in HMOs and other health plans over time with results available to the general public.

Structure of this chapter

In this chapter, we review evidence of the impact of financial incentives on quality, where the incentives were not structured specifically to reward providers for improving quality or meeting quality goals or benchmarks. Instead, the financial incentives considered in these studies typically were designed to influence other outcomes, most commonly utilisation of services or costs. However, study authors either incorporated quality measures alongside measures of cost and utilisation in their study designs or focused only on measuring impacts on quality. The body of literature we review in this chapter has a number of strengths. It addresses impacts of financial incentives on quality of care in a wide variety of settings for a range of medical services over a substantial period of time. Most studies are tied closely to public policy concerns of the time period in which they were conducted. Fundamentally, the literature addresses the question of whether financial incentives intended to influence costs and utilisation have a negative secondary impact on quality. This contrasts with Chapter 2 where we addressed the impact of financial incentives that explicitly reward providers for improving quality of care.

The remainder of this chapter is organised into six parts. It begins with a summary of the overall evidence regarding whether payment arrangements that attempt to reduce cost and utilisation increases have a negative impact on measures of quality and patient outcomes. We then discuss the findings of previous literature reviews that address this topic to varying degrees. In the next four parts, we discuss the evidence by study orientation, including studies that focus on insurance arrangements, on institutional providers, on physicians and other healthcare staff, and on health system comparisons. In the final part, we assess the strengths and limitations of the literature and, especially, what can be extracted from the study findings that can inform our assessment of efforts to reward quality, as discussed in Chapter 2.

Evidence summary

We reviewed 46 published articles that addressed, in some manner, financial incentives and their secondary impact on quality of care. We draw the following conclusions from that review:

- The evidence regarding the impact of financial incentives on quality of care is not compelling. There are several possible explanations. First, the incentives studied were designed, for the most part, to reduce utilisation of services. Generally, the hypothetical link between service reduction and quality in the studies is not clear, especially where utilisation may have been ‘excessive’ prior to the introduction of different payment arrangements. Second, the literature reports results for a wide range of quality and outcome measures, making it difficult to detect patterns in the findings. The most commonly used outcome measure – mortality – may not be sensitive to the relatively modest changes in financial incentives found in many studies. Also, mortality can be influenced by a host of factors, many unrelated to medical care, making it difficult to isolate the marginal effects of provider incentives.
- How incentives are transmitted to the level at which decisions about treatment are actually made is not clear in most studies. Typically, information is lacking concerning other efforts using the health plan, or hospital or physician practices relating to quality management. It seems likely that these efforts would interact with financial incentives for providers to influence quality of care. Most studies do not control for quality management efforts at the care delivery level when drawing conclusions about the impact of financial incentives.
- The use of multiple quality of care and patient outcome indicators in a single research study enables a richer interpretation of findings. However, when results are conflicting, no clear overall picture of the impact of incentives emerges. Also, it is not clear in most of these studies if the authors adjusted their statistical tests to account for the multiple comparisons undertaken in their analyses.

- The exact nature of provider payment arrangements often is not clearly described in the studies. This is true in particular for comparisons of quality of care under different insurance arrangements. Because the relationships between payment arrangements and quality are likely to be more subtle than the links between payment and service utilisation, the absence of a description of provider payment incentives makes interpretation of findings even more difficult.

Description of findings

Literature reviews

Our literature search identified 14 literature reviews that addressed the relationship between financial incentives and quality of care. Three of these reviews pertained to the impact of payment reforms in the UK, nine addressed the impact of managed care on quality in the USA, one focused on how moving to prospective payment for hospitals in the USA affected quality of care, and one took a cross-national perspective. These reviews generally were completed in the 1990s and seven of the 14 could be classified clearly as systematic reviews. All of the reviews considered a variety of outcome measures, some related to quality of care and others related primarily to cost and utilisation. In this respect they reflected the structure of individual studies, which typically included other measures, along with quality indicators, in their analyses. None of the reviews covered studies published after 2000.

The primary conclusion of the three UK reviews is that the literature on the relationship between payment reforms and quality of care is sparse. In 1996, Silcock and Ratcliffe summarised the evidence concerning the impact of payment changes instituted in 1990 on services provided by GPs. The authors concluded that the ‘small but growing’ body of evidence was not sufficient to draw definitive conclusions. Smith and Wilton (1998) examined the impact of NHS reforms designed to create an internal market for services, with GPs purchasing care on behalf of their populations. Using a systematic search process, the authors concluded that the evidence was incomplete or mixed. They suggested that procedural aspects of care delivery seemed to have improved, but that there was little evidence regarding impacts on the health of patients. Mays, Mulligan and Goodwin (2000) also reviewed studies relating to this topic. They concluded that few studies attempted to measure quality directly but that there was support for the notion that GP fundholders had used their funds to make services more accessible and waiting times shorter. These are not clinical measures of quality, nor health status, but do suggest more timely service provision for patients. In general, these three reviews find little evidence regarding impacts of UK payment reforms for GPs on quality of care.

The nine reviews of the impact of managed care in the USA on quality reach similar conclusions. Five of these reviews were conducted by Luft and, for the last three, co-authors (Miller and Luft, 1994; Miller and Luft, 1997; Dudley et al, 1998). In his first review, Luft (1980) compared quality of care in HMOs versus fee-for-service arrangements and concluded that HMOs compared favourably on structure and process measures of quality. With respect to outcome measures, he concluded that HMO enrollees did as well as, and in some cases better than, enrollees in fee-for-service plans. However, he also noted that data were limited and studies focused on well-established HMOs, leading him to observe that the issue was largely ‘unresolved’. Luft (1988) updated his review eight years later, with similar conclusions. At that time, he cautioned that these findings might not hold up as the insurance industry in the USA moved into a new era of competition between HMOs and other plans, and as new forms of HMOs proliferated. Miller and Luft published a further review of this literature in 1994 that concluded that quality of care in HMOs and fee-for-service indemnity plans was ‘roughly comparable’. In 14 of 17 studies, quality measures were better or equivalent in HMOs. Three years later, based on a systematic search of the literature, Miller and Luft (1997) found equal numbers of statistically significant positive and negative results on quality measures for HMOs compared with non-HMO plans. They observed that results unfavourable to HMOs were found for populations with chronic illnesses and suggested that more research was

needed on quality of care for this group. One year later, Dudley et al (1998) reviewed the entire literature from 1980 to 1997 that compared quality of care in HMOs versus other insurance arrangements. Their innovation was the use of a scoring system to evaluate the relative strength of the evidence. They found no clear pattern favouring better or worse quality of care in HMOs versus fee-for-service plans. Dudley et al (1998) noted the relative lack of studies that assess the link between explicit incentives for quality and measures of quality of care. Two other authors conducted literature reviews on the same topic at approximately the same time, reaching similar conclusions as Luft and colleagues. Cangialose et al (1997) found no differences in clinical quality or outcomes in patients in managed care plans versus traditional insurance arrangements. Steiner and Robinson (1998), based on a systematic review of the literature, found that managed care increased the use of screening and had a neutral impact on patient outcomes. Christianson et al (2001) found no evidence of less outpatient care being provided in managed care plans compared to fee-for-service ones for people with chronic illness, nor were there differences in health outcomes. The single exception was a study that reported greater declines in health status for managed care enrollees who were elderly or poor, and who had poor health status.

Hodgkin and McGuire (1994) conducted a systematic review of the impact of Medicare's movement to prospective payment for hospitalised beneficiaries based on DRGs. In general, they concluded that the strength of hospital responses to prospective payment was related to the percentage of hospital revenue derived from Medicare.

Chaix-Couturier et al (2000) reviewed the cross-national literature from 1993 to 1999 on the impact of financial incentives on physician practices, including costs, use of services, and measures of processes and outcomes of care. They found substantial evidence that incentives can alter physician behaviour, but did not summarise the evidence relating specifically to quality of care.

Chaix-Couturier *et al* (2000) used MEDLINE, Enbase, Health Planning and Administration, Pascal, International Pharmaceutical Abstracts and the Cochrane Library to search for articles from 1993 to 1999 that reported on the impact of financial incentives on medical practice. Outcome measures included visits, patient satisfaction, health ratings, immunisation rates and compliance with guidelines. The authors concluded that financial incentives could be used to reduce use of health resources and improve compliance with practice guidelines. They did not synthesise the impact of incentives on quality of care across the studies in their review.

Type of study: Systematic review

Christianson *et al* (2001) reviewed the literature on the impact of managed care on costs of care and patient outcomes for people with chronic illness. They found no evidence of less outpatient care being used by people with chronic illnesses in managed care plans compared to traditional insurance arrangements ones, nor were there differences in health status. The single exception was a finding in one study that elderly and poor people with initially poor health status had greater declines in health status in managed care plans. The authors noted that the number of studies addressing these issues was relatively small and that much more work was needed.

Type of study: Systematic review

Dudley *et al* (1998) provided an extended discussion of the measurement of quality and the ways in which financial incentives could affect quality of care. The second section of the paper reviewed the literature comparing quality in fee-for-service plans and HMOs based on literature from 1980 to August 1997. The authors discussed the incentives under capitation payment to HMOs versus fee-for-service payment to providers under traditional plans. They searched for articles with structure,

process or outcome measures of quality, but found only studies with process and outcome measures that fitted their selection criteria. Four different search engines were used to identify articles for possible inclusion in the review: articles had to have ending dates of 1980 or later, there had to be a comparison group, there had to be statistical adjustment for differences in enrollee characteristics and the articles had to pertain to plans in the USA and be peer reviewed. The authors developed a scoring system to weight the evidence in each article. Evidence was assigned a number ranging from one to seven, with one indicating strong evidence supporting quality differences in favour of HMOs and seven indicating strong evidence of quality differences in favour of traditional fee-for-service. The majority of studies using outcomes as quality measures found no difference between fee-for-service and HMO plans. Studies using process measures also did not reveal a dominant pattern.

Type of study: Systematic review

Hodgkin and McGuire (1994) developed a theoretical model of the impact of prospective payment for Medicare patients on hospitals and then conducted a literature review in the context of that model. A variety of outcomes were discussed, most related to utilisation of hospital services. The authors found that many studies do not separate the effects of changes in average and marginal reimbursement levels on hospital utilisation. Studies that have separated out the impact of average payment level have found it to be large. One study found that a 10 per cent decrease in payment was associated with an increase of 10 to 15 deaths per 1000 AMI admissions. The authors concluded that the direction of the effects was not always in accordance with predictions. Responses seemed to be related to the percentage of revenue the hospital derived from Medicare.

Type of study: Systematic review

Mays, Mulligan, and Goodwin (2000) reviewed literature that evaluates the NHS quasi-market reforms introduced in 1991/92. The incentives of the three foundations of the quasi-market reforms, as they might relate to quality, were not described in detail. Outcomes included measures relating to efficiency, equity, quality, choice and responsiveness, and accountability. The authors reviewed all published and unpublished studies that included data on the impact of the reforms. Studies were identified using electronic databases and a variety of other sources. Very few studies attempted to measure quality directly. There was some evidence that fundholders used their budgets to provide more accessible services and negotiated shorter wait times for hospitalisations.

Type of study: Systematic review

Miller and Luft (1997) built on their earlier (1994) review of HMO performance but focused only on studies that compare quality measures for HMOs and indemnity plans. HMOs are reimbursed on a prepaid basis, creating incentives for lower use of services compared with fee-for-service plans. These incentives could have a detrimental effect on quality of care. A variety of quality of care measures were employed, reflecting the quality measures used in the studies that were reviewed. The authors conducted a MEDLINE search to identify studies published in peer-reviewed journals in the last quarter of 1993 through to the last half of 1996 and early 1997. They excluded articles covered in their previous review and required statistical adjustments for differences between managed care and non-managed care groups. They also included results for Medicaid enrollees (unlike the prior review article). They identified 68 studies that had been published since October, 1993 that had not been included in the previous review, with 35 studies meeting the inclusion criteria. Fifteen of these 35 studies included results on quality of care, with many reporting on multiple measures of quality. The authors found that there were ‘equal numbers of statistically significant positive and negative results for HMO performance, compared with non-HMO plans’. They pointed out that the studies showing

negative findings with respect to quality of care in HMOs tend to focus on persons with chronic illness, suggesting that more research is needed in this area. They also suggested that the results should be interpreted with caution for a variety of reasons.

Type of study: Systematic review

Smith and Wilton (1998) reviewed the evidence regarding the NHS reforms that established an internal market with health authorities and GP fundholders purchasing care on behalf of their populations. The goal of the reforms was to improve cost-containment, cost-effectiveness, quality and patient choice. There was a variety of different incentives instituted as part of the changes. Fundholding gave GPs an opportunity to retain surpluses and therefore an incentive to contain costs. The specific outcomes varied with the study being reviewed but included costs, referrals and healthcare procedures. A key word search of two databases was conducted: MEDLINE and EconLit, from 1990 to 1996. Papers were selected from abstracts that seemed to address the objectives of fundholding. Additional papers referenced in the papers identified by the search also were reviewed. Overall, the authors concluded that the evidence was incomplete and mixed. There was evidence of cost-savings in prescribing but the evidence was less clear with respect to referrals. The prescribing improvements were very costly administratively to achieve. Some procedural aspects of care improved but there was little evidence regarding effects on health outcomes. There was very little discussion of the exact nature of the incentive changes, making it difficult to interpret the review findings.

Type of study: Systematic review

Steiner and Robinson (1998) conducted a literature review of the 'high quality' US evidence on performance of managed care organisations and on the impact of different managed care techniques, including financial incentives. Outcomes included use of services, quality of care, consumer satisfaction, and equity. The authors attempted to determine if research on managed care in the USA has produced findings that could be used to inform care delivery in the UK. Literature searches were conducted using MEDLINE, EMBASE, Social Sciences Citation Index, and the National Health Service Centre for Reviews and Dissemination (York) library. Seventy articles were identified for systematic review, covering 18 dimensions of performance. Also, 19 articles relating to specific managed care techniques were identified for review. The authors concluded that managed care reduces hospitalisation and the use of high-cost discretionary services, increases screening, reduces patient satisfaction, and is neutral with respect to patient outcomes. The authors also concluded that the results were not terribly useful for the UK because of limitations in research design and because the UK already practises population-based care management and gatekeeping.

Type of study: Systematic review

Insurance-level comparisons

We identified 13 studies that compared select quality or outcome measures for members of managed care plans versus fee-for-service plans. Seven of these studies analysed the experience of Medicaid recipients in the USA in managed care plans versus fee-for-service plans. The impact of increased reliance of Medicaid programmes (where states share funding with the federal government but administer programmes themselves) on managed care plans has been controversial. One concern has been that Medicaid recipients may not be able obtain the care they need when managed care plans have incentives to restrain costs. In response, proponents of managed care in Medicaid suggest that care will be better co-ordinated and that the plans can be held accountable for provision of preventive services, leading to higher quality of care overall for Medicaid recipients. In 1993, Langa and Sussman

studied rates of revascularisation in California, comparing Medicaid, HMO and fee-for-service insurance arrangements. They found that rates of revascularisation increased more rapidly in HMOs and private arrangements than in Medicaid, and more rapidly in private plans than in HMOs. They suggested that limits to access to care in Medicaid could explain the results (there were no Medicaid HMO plans during the study period). Hanson et al (1998) compared immunisation rates for Medicaid children in New York City in a managed care plan with those in fee-for-service Medicaid. There were no differences in the degree to which physicians performed preventive care under the two financial arrangements, but children in the managed care plan were more likely to be up-to-date on their immunisations. The study is noteworthy in that the researchers reported information on how physicians actually were reimbursed, with primary care physicians in the managed care plan receiving a capitated payment for preventive and acute services.

The remaining five Medicaid studies addressed various aspects of care for newborns and mothers. Levinson and Ullman (1998) compared the adequacy of prenatal care and also infant birthweights in Wisconsin counties under managed care versus fee-for-service Medicaid, finding no differences in birthweights and weak support for a conclusion that prenatal care was better in the managed care plans. Oleske et al (1998) also addressed obstetric care in Medicaid managed care and fee-for-service Medicaid in California, examining a wide range of newborn outcomes and pregnancy outcomes. The only significant difference was that there were fewer low birthweight babies under managed care. In a later study, Duggan (2004) revisited birth outcomes under Medicaid managed care in California versus fee-for-service, finding no difference in outcomes, but an increase in expenditures under managed care. In another California-based study, Aizer, Currie and Moretti (2004) compared a variety of quality of care and outcome measures for mothers and infants in Medicaid managed care versus fee-for-service, finding a negative impact of managed care on birthweight, gestation period and neonatal death rates. However, the quality of hospitals and intensity of care were higher in managed care plans. The method by which providers were paid under managed care was not described. In another Medicaid managed care versus fee-for-service analysis, Mitchell and Gaskin (2005) reported that children in fee-for-service Medicaid were more likely to have unmet needs for health services, prescription drugs and dental care. In this study, the authors noted that providers in managed care plans were paid on a fee-for-service basis at rates slightly higher than under fee-for-service Medicaid.

The other seven studies that compared quality and outcomes across insurance plans addressed a variety of issues. In an early study of this type, Udvarhelyi et al (1991) found that the quality of ambulatory care for HMO patients was equal to or better than care received by fee-for-service patients and suggested that lower out-of-pocket costs for HMO patients might have been part of the explanation. Potosky et al (1999) studied use of different therapies for treatment of prostate cancer, as well as ten-year survival, comparing two HMOs with fee-for-service Medicare. They reported that HMO patients received more aggressive therapy in one metropolitan area. Overall mortality was approximately equal, but mortality due to prostate cancer was higher among HMO patients. Mukamel, Zwanziger and Tomoszewski (2001) also compared outcomes for Medicare fee-for-service versus HMO patients, focusing on risk-adjusted mortality for hospitalised patients. Specifically, this study examined whether higher HMO market penetration was associated with better outcomes for fee-for-service Medicare patients, concluding that it was. Bundorf (2004) addressed a similar question using Medicare data from 1994 to 1996, focusing on hospitalised patients with AMI. She found that rates of revascularisation and cardiac catheterisation were lower in areas with relatively large numbers of HMO enrollees. Haile and Stein (2002) examined complications during a hospital stay in California as a measure of quality of care, finding significantly worse outcomes for HMO patients, especially patients involved in major surgery. They did not report the financial incentives facing providers under managed care. Luft (2003) compared rates of revascularisation and mortality for fee-for-service versus HMO patients in California after hospital treatment for AMI. He found that risk-adjusted death rates were slightly higher for fee-for-service patients, but that there was considerable variation in performance among the study HMOs. Finally, Mitchell and Gaskin (2007) concluded, based on a survey of caregivers to special needs children,

that caregivers to children enrolled in fee-for-service plans were more likely to rate access to care as fair or poor on four measures, in comparison with caregivers to children in managed care plans. They suggested that this could reflect better case management and care co-ordination in the managed care plans.

Aizer, Currie and Moretti (2004) conducted a comparison of patient outcomes in fee-for-service Medicaid with managed Medicaid plans, including technical quality of care received in the hospital by mothers and babies and outcomes for babies, such as low birth weight, short gestation and neonatal death. The authors estimated fixed effects statistical models using a longitudinal data base from 1990 to 2000 constructed from the California Birth Statistical Masterfile and the Birth Cohort files. They concluded that the introduction of managed care plans resulted in an increase in the incidence of low birthweight, short gestation and neonatal death. The quality of hospitals and intensity of care increased under managed care, but this was offset by a decline in access. There was no discussion of how payment to providers was structured under managed care. The changes observed did not seem related to provider payments.

Type of study: Quasi-experimental study

Bundorf (2004) conducted an empirical analysis of the impact of HMO activity in a geographic area on treatment, costs and outcomes for fee-for-service Medicare beneficiaries living in the same area. The incentives in managed care may change treatment processes employed by physicians. These same processes may then be applied to patients who are not HMO enrollees, affecting their costs and outcomes. This has been termed a 'spillover effect'. The author investigated patient treatment measures, cost rates of revascularisation and cardiac catheterisation for Medicare patients aged 65 or older discharged from hospitals between January 1994 and February 1996 with a diagnosis of AMI. The study sample consisted of patients aged 65 and older who were admitted to a non-rural hospital in 1994 or 1995 with a confirmed AMIU diagnosis. Measures of managed care penetration and managed care competition were the critical independent variables. Dependent variables included six different patient treatments and tests. Claims data were used to determine revascularisation and cardiac catheterisation during 30-day and one-year post admission periods. Logistic regressions and ordinary least squares regression was used to estimate the models. Rates of revascularisation and cardiac catheterisation were lower in areas with relatively high HMO penetration. In high HMO competition markets, patients were more likely to receive cardiac catheterisation.

Type of study: Observational study without a control group

Duggan (2004) estimated the effect of the move by Medicaid in California to contracting with managed care plans on Medicaid spending and infant health outcomes. Managed care plans have an incentive to hold costs down as they are at risk for the costs in excess of their capitated payments. Under fee-for-service Medicaid providers are rewarded for providing more services. There was no discussion of the financial incentives for providers in managed care plans. However, Medicaid may be able to negotiate lower prices with providers, suggesting that costs could increase by contracting with managed care plans. Outcomes include spending per Medicaid beneficiary, percentage of infants born prematurely and percentage of infants who died in the hospital. Complete claims data were used for a 20 per cent sample of AFDC-linked Medicaid recipients in California covering the period from January 1993 to December 1999, combined with California's hospital discharge data set. The author employed a fixed-effects model specification with county-specific time trend variables to estimate effects on Medicaid expenditures. He aggregated to the county level to estimate impact on outcome measures. The average effect of the switch in enrolment to managed care plans was an increase in Medicaid expenditures of 17 per cent. There was no significant improvement in health outcomes.

Type of study: Quasi-experimental

Haile and Stein (2002) analysed the relationship between type of managed care coverage and complications during a hospital stay, which was taken to be a measure of quality. The exact incentive was not described. Patients were grouped by their insurance coverage and it was assumed that physicians providing care in some type of managed care plan (HMO) may have incentives that could result in lower quality than physicians providing care to patients in other plans. The outcome measure was defined as the odds of experiencing a complication during a hospitalisation episode. California hospital discharge data from hospitals in Sacramento and San Diego from 1995 and 1996 formed the basis for the analysis. Only patients covered by commercial insurance were considered. A computer algorithm identifying presence of possible complications associated with admissions was used to construct the dependent variable. The authors used a regression model to estimate odds ratios. The authors found evidence of significantly worse outcomes for HMO patients, especially major surgery patients. Patients were treated similarly within hospitals, so the differences arose across hospitals. The authors addressed or discussed possible problems of unobserved heterogeneity and of patient selection and provider selection. However, the usefulness of the findings was limited by the fact that the authors did not know the actual incentives (type or degree) in place in the managed care plans they studied.

Type of study: Observational study without control group

Hanson *et al* (1998) carried out a cross-sectional analysis of immunisation rates for poor children who were covered under a managed care plan versus fee-for-service Medicaid. Fee-for-service is presumed to encourage provision of a higher volume of services, particularly acute services, which are reimbursed at a higher rate than preventive services. Preventive services take longer to deliver and are reimbursed at below market value. Most primary care physicians in the managed care plans are reimbursed at a capitated rate that includes both preventive and acute services. This encourages physicians to have large caseloads and to limit services. But patients 'belong' to a specific physician and plans must report performance on preventive care and other measures. The plans have an interest in encouraging providers to comply with preventive care guidelines. Measured outcomes include physician coverage for immunisations and lead and anaemia screenings, along with process measures of care, child receipt of immunisations and screenings. In 1995 a random sample of children was drawn from 60 physician practices that served New York City's poorest neighbourhoods.

A cross-sectional analysis was used to compare outcomes for children cared for by managed care and non-managed care physicians. In 1996 an independent sample of children from the same practices was used to compare outcomes for children enrolled in managed care versus those under traditional Medicaid. Logistic regression and chi-squared analysis was used to estimate differences. There were no differences between the two groups of physicians regarding their performance with respect to immunisations or screenings, or in process measures of care. Children in managed care plans were more likely to be up-to-date with immunisations. The authors suggested that physicians receiving compensation under two payment systems may treat children differently depending on each child's mode of reimbursement.

Type of study: Controlled observational study

Langa and Sussman (1993) examined rates of revascularisation in Medicaid, HMO and privately insured patients in California in 1983, 1985 and 1988. Rates were compared over three years for patients who received care in non-federal hospitals in California. Logistic regression models were used to analyse the data. The frequency of revascularisation increased across all of the groups, but more rapidly in the fee-for-service and HMO groups, compared with the Medicaid group. The rate of

increase in the HMO group was less than the fee-for-service group. The authors speculated that the difference in financial incentives may explain these findings. Also, reimbursement of physicians in Medicaid was constrained, leading to lower physician participation rates, especially for cardiologists. Limits on access to care for Medicaid patients could help to explain the results. Patients may present with more severe disease, where there is less opportunity for revascularisation.

Type of study: Controlled observational study

Levinson and Ullman (1998) conducted an empirical analysis of the effect of Medicaid managed care on the adequacy of prenatal care and on infant birthweights. Outcomes for Medicaid beneficiaries in three counties in Wisconsin who were enrolled in managed care plans were compared with outcomes in other counties where beneficiaries received services under fee-for-service Medicaid. It was not clear how providers were reimbursed in the managed care counties. Outcomes included infant birthweights and the Kotelchuck index of prenatal care, which combines the month prenatal care begins with the number of prenatal visits. Data were taken from 1994 birth records and Medicaid eligibility files. The authors controlled for potential selection bias in their analysis and used ordinary least squares and probit regression models to estimate relationships. The study found no differences in birthweights between the managed care and fee-for-service groups. The authors also concluded that women in the managed care group were 'somewhat more likely' to receive adequate prenatal care. Overall, they stated that their results supported the notion that savings generated by enrolling Medicaid beneficiaries in managed care plans were not coming at the expense of lower quality care or worse outcomes.

Type of study: Controlled observational study

Luft (2003) investigated whether fee-for-service and HMO patients in California had different rates of revascularisation and mortality after hospital treatment for AMI. HMOs are paid on a capitated basis and presumably have an incentive to manage care to contain costs, while providers serving fee-for-service health plan members are paid by fee-for-service. There was no information concerning how physicians were paid in the HMOs. Hospitals were paid on a prospective DRG basis for fee-for-service patients, with no information concerning how they were paid by HMOs. Hospital discharge data and death certificate data from California were linked with Medicare enrolment files for patients discharged from California hospitals from 1994 to 1996. Observed versus expected values of the outcome measures were calculated and compared. Risk-adjusted death rates were slightly higher for fee-for-service patients. There was wide variation in length-of-stay and revascularisation rates, but these rates were not associated with outcomes. Not all HMOs performed equally well, so it was suggested that more research is needed to explain differences in outcomes among HMOs.

Type of study: Observational study without control group

Mitchell and Gaskin (2005) carried out an empirical analysis of unmet needs for healthcare among children with disabilities, comparing those receiving care under a capitated managed care plan with those receiving care in the fee-for-service system. A health plan received a capitated payment from government to provide all needed services to children with special healthcare needs that qualify them for social security coverage. The plan was at risk for administrative, case management and outreach costs, but only partially at risk for medical expenses. Providers in the plan were reimbursed at fee-for-service rates 'slightly higher' when compared with rates of fee-for-service Medicaid. Several outcome

measures were used, all related to the concept of unmet need, as judged by the child's caregiver. They included unmet needs for services or equipment, physician or hospital services, medical equipment, prescription drugs and dental care. The study setting was the District of Columbia's Medicaid programme. Data for the study were collected through telephone interviews conducted in 2002 with 1088 caregivers. A statistical 'selectivity correction' approach was used to adjust for possible differences in children in the two plan options. Probit analysis was used to examine the determinants of having any unmet need in each category. Children in the fee-for-service option were 9.9 per cent more likely to have an unmet need for health services and 4.5 per cent more likely to have an unmet need for medical equipment. They also were more likely to have unmet needs for prescription drugs and dental care, although this difference was not statistically significant.

Type of study: Controlled observational study

Mitchell and Gaskin (2007) analysed survey data to determine if caregivers for special needs children believed access to different kinds of medical care was different under partially capitated managed care versus fee-for-service reimbursement. A two-step estimation procedure was used to correct for possible 'selection effects' associated with choice of options. The data were collected through telephone interviews conducted in 2002 with 1088 caregivers of children with special needs. Caregivers for children in the fee-for-service option were significantly more likely to rate access as fair or poor for four measures. The authors suggested that the observed differences could be due to the presence of more case management and care co-ordination in the managed care option.

Type of study: Observational

Mukamel, Zwanziger and Tomaszewski (2001) investigated whether the quality of care for fee-for-service Medicare beneficiaries improved as HMO penetration increased due to 'spillover' effects. The authors hypothesised that there were several different ways in which HMOs may influence the quality of care for non-members, including changing practice styles that are applied to HMO and non-HMO patients alike, reducing discretionary resources that providers can use to invest in new technologies, limiting the availability of some community providers for non-HMO patients and causing some providers to leave communities. The study outcome was risk-adjusted mortality rates for hospitalised patients. Hospital data from 1990 were used for 134 cities where there were five or more hospitals. Data sources included the American Hospital Association, Medicare and Interstudy. Regression models were used to estimate relationships between HMO measures and other variables and risk-adjusted 30-day post-admission mortality rates. Higher HMO market penetration was associated with better mortality rates for fee-for-service Medicare enrollees. The authors noted that the study was not able to determine the specific pathway by which the better rates were achieved.

Type of study: Controlled observational study

Oleske *et al* (1998) compared quality and patient outcomes of obstetric care for Medicaid women under managed care versus traditional Medicaid. The authors compared outcomes in two counties where care was delivered under an arrangement in which health plans received capitated payments to three counties where providers were reimbursed directly by fee-for-service. The authors did not indicate how providers in the managed care plans were paid. Two types of outcomes were studied. Newborn outcomes included birthweight, foetal alcohol syndrome, drug withdrawal syndrome, congenital rubella, meconium aspiration syndrome, birth injury, hyaline membrane disease, seizures and admission to an intensive care unit after 24 hours. Adverse pregnancy outcomes included eclampsia, convulsions during labour, maternal death, anaesthetic complications, excessive bleeding,

febrility and unsuccessful vaginal birth after caesarean section. Caesarean birth was viewed as a negative outcome because of potential risks to the mother and her newborn baby. For comparison purposes, counties were matched by geography and sociodemographic characteristics. Data were obtained from the State of California's birth certificate records. Multiple logistic regression analysis was used to estimate odds ratios and confidence intervals. The likelihood of low birthweight was lower for births taking place in counties where the capitated plan existed, controlling for other characteristics. There were no other significant differences. The authors noted that there were factors in the operations of the capitated plans that could explain the relatively favourable findings, including the assignment of patients to a primary care physician who served as a case manager and programmes in these counties that encouraged pregnant mothers to seek care early. The authors noted that reliability and validity of birth certificate data could be a limitation of the study.

Type of study: Controlled observational study

Potosky *et al* (1999) conducted an empirical analysis that compared treatment of prostate cancer and outcomes in two group practice models with fee-for-service. The two managed care plans in the study were paid by Medicare on a capitated basis. In one of the plans – Group Health of Puget Sound – the physicians were paid on a salary basis. There was no discussion of how physicians were paid in the other plan (Kaiser-Northern California). In traditional Medicare, physicians were paid on a fee-for-service basis. Outcomes included use of prostatectomy or radiation in treatment and mortality over a ten-year period. The study tracked a cohort of men diagnosed with prostate cancer between 1985 and 1992, and followed them into 1994. Data from Medicare claims files and from two population-based tumour registries covering San Francisco/Oakland and Seattle/Puget Sound were used in the analysis. Multivariate logistic regression models were employed to estimate relationships involving treatments, and Cox proportional hazard regression models were used to estimate survival relationships. Patients in Kaiser were more likely to receive aggressive therapy relative to fee-for-service patients. There was no difference for patients in Group Health. Overall, mortality was equivalent over ten years, but prostate cancer mortality was higher for the HMO patients. The authors speculated that differences in unmeasured patient characteristics may explain observed mortality differences.

Type of study: Controlled observational study

Udvarhelyi *et al* (1991) compared the receipt of services by HMO and non-HMO patients in the same four group practices. Physicians were paid based on productivity for fee-for-service patients. In one group, the physicians were capitated for their HMO patients, while in the other three they received productivity-based payments with utilisation review. In general, the incentives as described by the authors would not have provided clear signals to physicians because they were seeing both types of patients in their practices. Measures related to hypertension control, preventive care and resource use were used. Equal numbers of HMO and fee for-service patients were selected by randomly choosing medical records from the four group practices, resulting in 246 patients with chronic uncomplicated hypertension and 250 women without chronic diseases who received preventive care. The quality and quantity of ambulatory care received by HMO patients was equal to or better than that for fee-for-service patients. HMO patients received more visits. The authors noted that there were lower costs borne by the HMO patients themselves when receiving care and that this might explain the results. Also, the difference in incentives may not have been very large and physicians may not have been able to practise a different type of care for their HMO patients.

Type of study: Controlled observational study

Hospital-level comparisons

Eight studies explored the relationship between hospital payment arrangements and various quality or outcome measures. The USA was the setting for three of these studies, all of which examined the movement to a prospective, DRG-based payment system. In a very early study, Mushlin et al (1988) assessed the impact of moving to a community-wide (Rochester, New York State) prospective payment approach for hospital care, featuring a cap on hospital expenditures. They found that outcomes of care were not affected by the change. Davis et al (1995) studied the impact of Medicare's prospective payment system on care for patients with depression, concluding that length of stay was unchanged but intensity of services received by hospitalised patients increased. In a recent study, Dafney (2005) analysed how hospitals responded to price changes for some DRGs, in comparison to DRGs where prices did not change. They found that the price changes did not result in increases in intensity of care or quality. Instead, they concluded that the extra money flowing to hospitals was associated with longer stays, more surgeries, more intensive care unit days and 'possibly worse outcomes', although evidence was weak on this point. In research not related to DRG payment rates, Hamilton (1993) studied the responsiveness to per day payment levels of hospice decisions to become certified by Medicare, a proxy quality measure. Each dollar increase in daily home care rates increased the probability of certification by 1.7 per cent.

The five non-US studies were set in different countries. Hamilton and Bramley-Hasher (1999) assessed the impact of introducing a DRG-based, per admission payment approach in the UK on care given to hip fracture patients, concluding that in-hospital mortality fell. Milcent (2005) examined the impact of an ownership and reimbursement approach on mortality rates of hospitals for treatment of AMI. The private sector hospitals were paid fee-for-service and demonstrated a lower 'spontaneous risk of dying' but with considerable variation across hospitals. In Norway, Kjerstad (2003) assessed the impact of replacing global budgets for hospital care with DRG reimbursement. The magnitude of the impacts was small and no direct quality measures were available. In Sweden, Forsberg, Axelsson and Arnetz (2001) surveyed physicians about the impact of the introduction of diagnosis-related payment for hospital services. Physicians reported that quality of care had deteriorated. Shmueli, Intrator and Israeli (2002) studied the first year impact of changing hospital payment in Israel from a daily to a payment per procedure approach. Quality of care was measured by readmission and mortality rates for five procedures in four medical centres. They found that readmission rates increased significantly in three of the procedures, but there was no impact on mortality.

Dafney (2005) analysed how hospitals responded to large price changes in 1988 for 43 per cent of Medicare admissions. Price changes occurred for some DRGs but not others. Many different data sources related to the Medicare programme were used to assemble data on price changes, hospital admissions and hospital characteristics. American Hospital Association data were also used to construct measures of hospital characteristics. Various multiple regression techniques were employed to estimate the impacts of the price changes. Hospitals coded patients to diagnoses with the largest price changes, especially for-profit hospitals. There was little evidence of volume increases. More importantly, the price increases did not result in increases in intensity of care or quality. There was some evidence that hospitals spent the additional money on more care in all DRGs, not just where price increases occurred. Extra money was associated with longer stays, more surgeries, more intensive care unit (ICU) days and 'possibly worse outcomes', although the evidence on the latter was weak. The results suggested that policy-makers may not be able to 'buy' better quality in a particular service area simply by increasing the price for services of that type.

Type of study: Controlled observational study

Davis *et al* (1995) conducted a methodologically sophisticated analysis of the impact of the introduction of Medicare's prospective payment system on service use in the treatment of elderly depressed patients. Payment for hospital care per admission based on diagnosis was compared with reimbursement on a flat payment per case based on historical charges. The expectation was that length of stay and intensity of service use after DRG implementation would both decline. Data were collected from 297 acute care general hospitals, with and without psychiatric units, in five states. The hospitals were chosen using a stratified selection criteria. The final sample included 2746 depressed Medicare patients aged 65 and older who had been hospitalised during the before or after period. Data were collected through a review of medical records. Fourteen measures of use of clinical services were developed. Multiple regression analysis was used to analyse the data, controlling for covariates. Implementation of Medicare prospective payment was not associated with a decrease in length of stay but was associated with an increase in intensity of services. Although the authors did not explicitly say so, the results seemed to imply that quality of care, measured by intensity of treatment, was not affected by the switch to prospective payment-based DRGs. A limitation of the study was that the strength of the difference in financial incentives under the two payment arrangements, and the direct impact on treatment providers (not hospitals), was difficult to determine.

Type of study: Quasi-experimental study

Forsberg, Axelsson and Arnetz (2001) conducted a study of the type of incentive created by performance-based reimbursement in Sweden, specifically the introduction of diagnosis-related payments for hospital services. The study was a cohort design involving the five largest hospitals in Stockholm and ten other county councils. Data were collected from physician surveys in the late autumn of 1994 and 1998. There was a 77 per cent response rate in 1998. Principal components analysis was used to create factor scores and regression analysis was used to analyse relationships. Most physicians considered the prospective payment system a good tool to improve efficiency and said they had not experienced financial pressure. However, physicians felt that the quality of care had deteriorated during the study period.

Type of study: Observational study without control group

Hamilton (1993) conducted an empirical analysis of the impact of reimbursement rate on decisions of hospices to become certified by Medicare and thus available to Medicare beneficiaries. Data on reimbursement rates and labour costs were from the Medicare programme, and various other sources of data were used to construct control variables. A switching regression model with endogenous switching was estimated with maximum likelihood techniques. Each dollar increase in the daily home care rate raised the probability of certification by 1.7 per cent. However, if rates are not set to reflect the real costs of certification in any given area, disparities in the availability of hospice care for Medicare beneficiaries would exist.

Type of study: Observational study without control group

Hamilton and Bramley-Hasher (1999) assessed the impact of payment reforms on length of hospital stay for hip fracture patients in the UK. NHS reforms introduced a DRG-based, per admission payment system for hospitals with the intent of reducing length of stay and queues for hospital services. Length of stay in the hospital for hip fracture patients, waiting times, inpatient mortality and discharge destination were measures of impact used in the study. Data were taken from the hospital

and 1994/95. The authors used survival analysis to examine the timing of death or discharge as a function of wait time for surgery, controlling for other factors. A duration model with multiple destinations was used to model where the patient went after discharge. The authors addressed the possibility of unobserved heterogeneity caused by unobserved differences in patients at the time of their admission. Length of stay declined after the implementation of payment reforms, but only a small part of the decline was due to decreased waiting times (and presumably more timely surgery). The decline was primarily due to discharges to other less costly facilities. In-hospital mortality fell as well.

Type of study: Observational study with data from before and after payment reform

Kjerstad (2003) assessed the impact of replacing global budgets with DRG-based reimbursement for hospital care in Norway. The purpose of this payment reform was to increase the number of patients seen in hospitals in Norway. A DRG-based system was expected to result in more patients being seen more quickly without reducing quality of care, relative to the existing block grant system. Data were collected from 59 hospitals from 1995 to 1998. A difference-in-difference model was estimated using two time periods: the year before and the year after the reform was introduced. The reform had a significant impact on the number of patients seen and on DRG points, but the magnitude of the effects was small. The authors noted that they were not able to address quality directly, due to a lack of appropriate data. However, under the payment system, an increase in readmissions led to more DRG points. If that were the driving factor in the results, it could indicate reduced quality of care.

Type of study: Quasi-experimental

Milcent (2005) examined the effect of ownership and a system of reimbursement on mortality rates in France. Hospitals in the private sector in France are paid fee-for-service which creates different incentives than those that exist for public hospitals; in addition, private hospitals can select patients. Mortality rates in hospitals for treatment of AMI, and the level of hospital investment, were used as measures of quality. A 1997 national database on hospitals served as the primary source of data. A duration model with multiple destinations was used for the analysis, avoiding the assumption that length of stay and mortality were independent. A piece-wise constant exponential model also was employed. To appropriately rank hospitals by mortality rate, the composition of patients needs to be taken into account. The author found that in for-profit hospitals patients had, on average, a lower spontaneous probability of dying. But there was more heterogeneity within this group, making inferences about the overall quality of the hospital more difficult. Private hospitals providing more innovative procedures had lower mortality rates. The author did not discuss the explicit role of the different payment incentives in explaining differences in mortality rates.

Type of study: Observational study without control group

Mushlin *et al* (1988) assessed the impact of moving to a prospective payment approach for hospital care on a community-wide basis in Rochester, New York State, to a prospective payment approach for hospital care. Hospital payment took place under a community-wide cap on all patient care revenues. Reimbursement to individual hospitals was based on historic costs adjusted for wage changes. If a hospital generated a surplus, it could retain it. The authors tracked measures of access, admissions and outcomes of care. Data were collected for five years (1980–84) from a uniform database on hospitalised patients in Rochester. Access to care was maintained and admissions

increased in some areas. However, rates of admission for elective surgery declined. Outcomes of care – including neonatal deaths, ischaemic heart disease deaths, deaths from five selected surgical conditions and rates of adverse outcomes from sentinel conditions – remained stable. The experiment in Rochester was unique. It was the first attempt to implement comprehensive provider payment reform at the community level in the USA, and has not been repeated in other communities.

Type of study: Observational study without control group

Shmueli, Intrator and Israeli (2002) examined the first year effect of a payment change in 1990 for hospital care in Israel. Hospital payment changed from a per day arrangement to prospective payment per procedure. The study tracked volume of hospital activity, length of stay, quality of care and hospital income. The focus of the study was on five procedures in the four largest medical centres in Israel. The dataset included 17,000 hospitalisations occurring during the two years prior to the change (1988–1990) and one year after (1990/91). Quality of care was measured by the 60 day post-discharge readmission rate and in-hospital, 60, and 365 days post-discharge mortality. Logit analysis was used to estimate impact on readmissions and mortality. Readmission rates increased significantly for three of the five procedures. There was no impact on mortality. Thus, the findings were mixed with respect to quality of care.

Type of study: Controlled observational study

Physician-level comparisons

Of the 11 studies we reviewed relating to payment of physicians and quality of care, five used care-based measures of quality in their analysis. The other six were based on physician surveys or interviews in which physicians were asked whether they thought quality had been affected by a specific payment change, or might be affected under a specified scenario. In one of the five studies based on actual care delivery data, Bateman et al (1996) examined the impact of reimbursement on the prescribing behaviour of GPs in a single region in the UK during 1993/94. GPs were allowed to keep part of the savings in their prescribing expenditures, relative to a predetermined target, and could receive awards for prescribing practices judged to be of good quality as well. The primary finding of the study was that the incentives to conserve prescribing expenditures did not have a negative impact on quality. Gaynor, Rebitzer and Taylor (2001) examined the impact of different types of GP contracts in group practices in the USA on expenditures and measures of quality. They found that a combination of incentives rewarding lower costs and achievement of quality standards led to improved performance on preventive care measures at the same time that expenditures declined by about 5 per cent for outpatient care. In a study with similar objectives, Wee et al (2001) examined the impact of physician productivity incentives for 11 academically-affiliated internist practices in Boston on the provision of preventive care. The results of this study were somewhat contrary to those reported by Gaynor, Rebitzer and Taylor (2001). Wee et al (2001) found that patients cared for by physicians with financial productivity incentives were less likely than other patients to receive Pap smears and cholesterol screening, but that there was no significant effect on mammography or influenza vaccination rates. Madison (2004) examined the relationship between the type of affiliation that a physician had with a hospital and outcomes for myocardial infarction treatment, with different affiliations having different financial implications. They found little evidence of an impact on patient treatments and outcomes. Goodman (2006) analysed the relationship between the financial risk experienced by primary care physicians and emergency department use. Use of emergency rooms was less for patients of physicians exposed to higher levels of risk, but this result was not tied to specific measures of quality or patient outcomes.

Among the studies adopting a survey or interview approach, physicians typically were queried on how they would respond, or had responded, to particular incentives. Grumbach et al (1998), in his survey of Californian physicians, found that those with productivity incentives were more likely to report that such incentives compromised care. Hadley et al (1999) found that most of the physicians in their survey (67.8 per cent) reported that the financial incentives they faced were neutral in their impact on care. Physicians who reported moderate or strong incentives to reduce services were more likely to report that they didn't have the freedom to care for patients as they wanted to. In a second study of Swedish physicians, Forsberg, Axelsson and Arnetz (2000) reported that physicians felt that a new prospective payment system had precipitated deterioration in quality of care and increased premature discharges from hospitals. Tufano et al (2001), using interviews of physicians and administrators, found that respondents believed different compensation schemes had little impact on the technical quality of care. In a quite different approach, Pantilat, Chesney and Lo (1999) examined the responses of physicians when confronted in a survey with different types of financial incentives and asked if they would refer patients under four scenarios. The idea was to determine if financial incentives would cause physicians to withhold services that were indicated according to established guidelines. The authors concluded that their study provided 'modest support' for concerns that financial incentives could influence physicians in ways that would reduce quality of care. Shen et al (2004) took a similar approach, presenting physicians with various scenarios regarding financial incentives, with the scenarios distinguishing capitated from fee-for-service care. In the scenarios, treatments that promised large, undeniable benefits for patients were unaffected by financial incentives, but treatments with smaller expected benefits were affected. In general, when faced with incentives to conserve resources, physicians tried to do so, in the scenarios, in ways that would preserve quality of care for their patients.

Bateman (1996) assessed the effects of reimbursement on the prescribing behaviour of GPs in the UK who are not fundholders. Physicians could retain a part of any savings they generated when their prescribing expenditures were less than a targeted amount. Twenty-three per cent of practices achieved the incentive target, and they received incentive payments of £420,000. Quality awards were made to 27 practices where prescribing was judged to be of good quality (£43,000). The setting was non-fundholding practices (459) in the northern region in 1993/94, which were grouped into three bands according to the ratio of their prescribing amount to the local average. A statistical analysis of means was conducted. There were no controls for practice or practitioner characteristics. Twenty-three per cent of practices achieved targeted savings. There was no reduction in the quality of prescribing practices.

Type of study: Controlled observational study

Forsberg, Axelsson and Arnetz (2000) attempted to determine the changes in physician attitudes and behaviour related to efficiency and quality of care that occurred after a new prospective payment system was introduced in Stockholm County in January, 1992. Information on a variety of measures related to efficiency and quality of care was collected through physician questionnaires. Data were collected from physicians during one month each year from 1992 to 1994 in Stockholm County and in ten other counties in 1994. Official statistics were used as a source of utilisation data. A comparison of means was conducted. Physicians in the hospitals with the new reimbursement approach believed their clinics had become more efficient and that the number of premature discharges had increased. They were pessimistic about the future prospects for providing good quality care and believed that quality had deteriorated more rapidly than physicians in other areas. The authors observed that research was needed to determine the quality implications of earlier discharges.

Type of study: Controlled observational study

Gaynor, Rebitzer and Taylor (2001) analysed the impact of physician payment incentives on use of services and some quality indicators. Incentives included group-based contracts for physicians in which financial rewards were made for average patient expenditure below a target level and bonuses were paid by the HMO for meeting specified quality standards patterned after HEDIS standards. Physician income under the incentive structure increased by about $0.10 for every $1 reduction in expenditure. The study focused on a specific HMO and its physician payment policy in place between 1994 and 1996. Regression analysis was used to control for other factors that could influence outcomes. The incentive reduced expenditures by about 5 per cent, predominantly in the outpatient area. Data were available for only one year to address quality outcomes. The authors reported that quality improved for the preventive care measures in response to the incentives.

Type of study: Observational study without control group

Goodman (2006) conducted an empirical analysis of cross-sectional data from an HMO to determine if HMO patients under the care of primary care physicians with greater financial risk have lower rates of use of the emergency rooms. Emergency departments may also have financial incentives to welcome relatively easy to treat patients. The exact nature of the financial risk assumed by the primary care physician is not known, as typically the contracts were between the health plan and the medical group. Administrative data from the year 2000 on 217,298 commercial members enrolled in a statewide (Michigan) HMO were used for the analysis. The physicians had various levels of financial risk. Ordinary least squares and binary logistic models were used to estimate relationships. Emergency room use was lower by 33 visits per 1000 and 51 visits per 1000 for patients of physicians with medium to high financial risk compared with physicians with low financial risk. There was no attempt to relate these findings specifically to quality of care measures.

Type of study: Observational study without control group

Grumbach *et al* (1998) analysed data from a survey conducted in 1996 of physicians (766) in the largest urban counties in California. Physicians were asked about how they had responded to financial incentives of varying types. Logistic regression models were used to analyse the data. Fifty-seven per cent of respondents felt pressure to limit referrals and 75 per cent felt pressure to see more patients per day. Physicians with incentives based on refferals were more likely to believe that pressure to limit referrals compromised care. Physicians with productivity incentives were more likely to believe that incentives to see more patients compromised care. Physicians with incentives related to the quality of care or patients satisfaction were more likely to be ‘very satisfied’.

Type of study: Observational study without control group

Hadley *et al* (1999) used physician survey data to explore how physician financial incentives affected three measures of satisfaction and practice style: overall practice satisfaction, degree to which prior expectations about professional autonomy and ability to practise good quality medicine are met and specific measures of practice style. In 1997, the authors conducted a telephone survey of 1549 physicians. These physicians were located in the 75 largest MSAs, as measured using 1991 population figures. Eligible physicians were under 52 years of age, had 8–17 years of post residency experience and were in patient care 20 hours per week. There were 1549 responses, for a response rate of 74 per cent. Logit regression and ordered logit regression analysis were used to estimate relationships. Most physicians (67.8 per cent) reported that their financial incentives were neutral.

Fifteen per cent said they had a moderate or strong incentive to reduce services. These physicians were more dissatisfied with their practices, and more likely to report that their expectations were not being met and that they didn't have the freedom to care for patients the way they wanted to.

Type of study: Observational study without control group

Madison (2004) analysed how hospital affiliation of physicians affected treatments, expenditures and patient outcomes. The authors hypothesised that, if affiliation made the reputations of affiliated physicians interconnected, physicians might spend more time overseeing the care provided by other physicians and care measures might improve as a result. The outcomes in the study related to treatment of myocardial infarction and included whether a patient was readmitted, received a catheterisation or angioplasty or bypass surgery, or died within 90 days of initial admission. Expenditures also constituted an outcome variable. Data were taken from the Medicare Provider Analysis and Review dataset, AHA data and the Area Resource File. The analysis focused on Medicare patients with AMI admitted to general medical–surgical hospitals between 1994 and 1998. Relationships were estimated using multivariate regression models. The authors found that an integrated salary model of hospital-physician affiliation was associated with slightly higher procedure rates and higher patient expenditures. However, there was little evidence that affiliation had an effect on patient treatments or outcomes.

Type of study: Observational study without control group

Pantilat *et al* (1999) conducted a survey of physicians who were told they had different types of financial incentives (fee-for-service, capitation with bonuses and full capitation) and asked about whether they would refer patients under four scenarios, where a referral was clearly indicated in each case. Questionnaires were sent to 1009 randomly selected internists in areas with 30 per cent or greater managed care penetration; there was a 70 per cent response rate. Chi-squared tests were used to conduct the comparisons. The researchers found that the more restrictive the financial incentive, the less likely that physicians would say they would order the referral or test. Compared to fee-for-service, capitation with utilisation review caused a small decrease of 3–11 per cent. However, even among physicians randomised to fee-for-service, 6–39 per cent failed to say they would order the test for indicated services. According to the authors, the results provided 'modest support' for concerns that financial incentives could influence physicians in a way that resulted in lower quality care.

Type of study: Randomised controlled trial

Pantilat *et al* (1999) conducted a survey of physicians who were told they had different types of financial incentives (fee-for-service, capitation with bonuses and full capitation) and asked about whether they would refer patients under four scenarios, where a referral was clearly indicated in each case. Questionnaires were sent to 1009 randomly selected internists in areas with 30 per cent or greater managed care penetration; there was a 70 per cent response rate. Chi-squared tests were used to conduct the comparisons. The researchers found that the more restrictive the financial incentive, the less likely that physicians would say they would order the referral or test. Compared to fee-for-service, capitation with utilisation review caused a small decrease of 3–11 per cent. However, even among physicians randomised to fee-for-service, 6–39 per cent failed to say they would order the test for indicated services. According to the authors, the results provided 'modest support' for concerns that financial incentives could influence physicians in a way that resulted in lower quality care.

Type of study: Randomised controlled trial

Shen *et al* (2004) asked physicians to respond to different scenarios regarding financial incentives. The incentives were hypothetical, embedded in patient scenarios, and distinguished between capitated and fee-for-service payment. Outcomes included performance of discretionary care, performance of life-saving care, and physician discomfort regarding decisions. The purpose of the study was to test the null hypotheses that physician decision-making would not be affected by financial incentives and that physicians would have equal concerns about decisions under different types of incentives. The authors surveyed a representative random sample of family physicians in direct patient care. Data from 601 physicians (a 72 per cent response rate) were analysed. In addition to comparison of means, multiple regression and logistic regression techniques were used. Physicians on average said they would conserve discretionary resources under capitated arrangements compared with fee-for-service. They felt more distressed when making decisions under capitation. Treatments that offered large undeniable benefits to patients were not affected by payment. Treatments offering small or questionable benefits were affected. Physicians tried to conserve resources when faced with financial incentives to reduce utilisation but also tried to maintain quality.

Type of study: Randomised controlled trial

Tufano *et al* (2001) conducted a qualitative study based on interview data collected from physicians and administrators in 1998. Respondents were asked their perceptions regarding how compensation affects productivity and the behavioural effects of individual financial incentives. Interviews were conducted with 146 physicians, medical leaders and administrators in 46 groups in four states. Compensation method was perceived to have a significant impact on physician productivity. The vast majority of respondents said that there was no impact of their compensation schemes on technical quality of care.

Type of study: Observational study

Wee *et al* (2001) studied the impact of physician productivity incentives on the receipt of preventive care by patients in 11 academically-affiliated practices in Boston. Four measures from HEDIS were used in the study: Pap smears, mammograms, cholesterol screening and influenza vaccinations. Chart reviews of 4473 patients of 169 internists from 11 primary care practices were carried out, along with a survey of the physicians and interviews with practice managers. Multiple logistic regression methods were used, with adjustment for clustering. Patients cared for by physicians with financial productivity incentives were significantly less likely to receive pap smears and cholesterol screening, but there was no impact on receipt of the other two types of preventive care.

Type of study: Controlled observational study

System-level comparisons

We identified only one study that could be characterised as a comparative system-level analysis: a non-systematic literature review regarding the impact of payment reforms in Taiwan, South Korea and China (Eggleston and Hsieh, 2004). In Taiwan, the impacts on quality of care of reimbursement for hospital care on a case basis were mixed. In South Korea, the change to a DRG-based system had no effect on quality measures. Likewise, no impacts on quality were reported that were related to payment reforms in China.

Conclusions

In this chapter we reviewed the literature regarding the impact of financial incentives, designed primarily to affect utilisation, on quality of care. We found no consistent evidence that financial incentives employed to reduce utilisation have had a negative impact on quality. In this respect our findings are consistent with those of the numerous reviews comparing quality of care received by patients enrolled in HMOs with patients in traditional insurance arrangements.There are several possible explanations for this result:

- measures used in many of the studies may not have been sensitive enough to detect quality impacts
- it was not always clear that incentives at the practitioner level were substantially different for the groups of patients or health plan members being compared, especially in the 'insurance-level' comparisons
- many of the studies were cross-sectional, making it difficult to infer causality in the estimated relationships between incentives and quality measures
- there may have been considerable variability in provider quality of care irrespective of financial incentives that made it difficult for researchers to detect the influence of incentives on quality.

The relatively few studies in this literature on the impact of financial incentives on individual physicians is particularly problematic given the current widespread interest in designing incentives that influence quality of care at the individual provider level.

References

Aas IHM (1995). 'Incentives and financing methods'. *Health Policy,* vol 34, pp 205–20.

Achat H, McIntyre P and Burgress M (1999). 'Health care incentives in immunization'. *Australian and New Zealand Journal of Public Health,* vol 23, pp 285–88.

Aizer A, Currie J and Moretti E (2004). *Competition in Imperfect Markets: Does it help California's Medicaid mothers?* NBER Working Paper 10429. National Bureau of Economic Research.

Amundson G, Solberg LI, Reed M, Martini EM and Carlson R (2003). 'Paying for quality improvement: compliance with tobacco cessation guidelines'. *Joint Commission Journal on Quality & Safety,* vol 29, pp 59–65.

Anderson KK, Sebaldt RJ, Lohfeld L, Burgess K, Donald FC and Kaczorowski J (2006). 'Views of family physicians in southwestern Ontario on preventive care services and performance incentives'. *Family Practice,* vol 23, pp 469–71.

Armour BS, Pitts MM, Maclean R, Cangialose C, Kishel M, Imai H and Etchason J (2001). 'The effect of explicit financial incentives on physician behavior'. *Archives of Internal Medicine,* vol 161, pp 1261–66.

Ashworth M, Armstrong D, de Freitas J, Boullier G, Garforth J and Virji A (2005). 'The relationship between income and performance indicators in general practice: a cross-sectional study'. *Health Services Management Research,* vol 18, pp 258–64.

Audit Commission (2005). *Early Lessons From Payment By Results.* Audit Commission.

Badgett T and Rabalais GP (1997). 'Prepaid capitation versus fee-for-service reimbursement in a Medicaid population'. *The American Journal of Managed Care,* vol 3, pp 277–282.

Bateman DN, Campbell M, Donaldson LJ, Roberts SJ and Smith JM (1996). 'A prescribing incentive scheme for non-fundholding general practices: an observational study'. *British Medical Journal,* vol 313, pp 535–38.

Beaulieu ND and Horrigan DR (2005). 'Organisational processes and quality: putting smart money to work for quality improvement'. *Health Services Research,* vol 40, 1318–34.

Benson T (2002). 'Why general practitioners use computers and hospital doctors do not – part 1: incentives'. *British Medical Journal,* vol 325, pp 1086–89.

Berwick DM (1995). 'The toxicity of pay for performance'. *Quality Management in Health Care,* vol 4, pp 27–33.

Bjorndal A, Arntzen E and Johansen A (1994). 'Salaried and fee-for-service general practitioners: is there a difference in patient turnover?' *Scandinavian Journal of Primary Health Care,* vol 12, pp 209–13.

Bodenheimer T, May JH, Berenson RA and Coughlan J (2005). *Can Money Buy Quality? Physician Response to Pay for Performance.* Center for Studying Health System Change Issue Brief No. 102.

Bokhour BG, Burgess Jr JF, Hook JM, White B, Berlowitz D, Guldin MR, Meterko M and Young GJ (2006). 'Incentive implementation in physician practices: a qualitative study of practice executive perspectives on pay for performance'. *Medical Care Research and Review,* vol 63, pp 73S–95S.

Bradlow J and Coulter A (1993). 'Effect of fundholding and indicative prescribing schemes on general practitioners' prescribing costs'. *British Medical Journal,* vol 307, pp 1186–89.

Bundorf MK, Schulman KA, Stafford JA, Gaskin D, Jollis JG and Escarce JJ (2004). 'Impact of managed care on the treatment, costs, and outcomes of fee-for-service Medicare patients with acute myocardial infarction'. *Health Services Research,* vol 39, pp 131–52.

Burns LR, Geller SE and Wholey DR (1995). 'The effect of physician factors on the cesarean section decision'. *Medical Care,* vol 33, pp 365–82.

Cameron PA, Kennedy MP and McNeil JJ (1999). 'The effects of bonus payments on emergency service performance in Victoria'. *The Medical Journal of Australia,* vol 171, pp 243–46.

Campbell S, Reeves D, Kontopantelis E, Middleton E, Sibbald B and Roland M (2007). 'Quality of primary care in England with the introduction of pay for performance'. *The New England Journal of Medicine,* vol 357, pp 181–90.

Cangialose CB, Cary SJ, Hoffman LH and Ballard DJ (1997). 'Impact of managed care on quality of healthcare: theory and evidence'. *The American Journal of Managed Care,* vol 3, pp 1153–70.

Casalino LP, Alexander GC, Hin L and Konetzka RT (2007). 'General internists' views on pay-for-performance and public reporting of quality scores: a national survey'. *Health Affairs,* vol 26, pp 492–99.

Chaix-Couturier C, Durand-Zaleski I, Jolly D and Durieux P (2000). 'Effects of financial incentives on medical practice: results from a systematic review of the literature and methodological issues'. *International Journal for Quality in Health Care,* vol 12, pp 133–42.

Chan L, Koepsell TD, Deyo RA, Esselman PC, Haselkorn JK, Lowery JK and Stolov WC (1997). 'The effect of Medicare's payment system for rehabilitation hospitals on length of stay, charges, and total payments'. *The New England Journal of Medicine,* vol 337, pp 978–85.

Chiang C-Y, Enarson DA, Yang S-L, Suo J, Lin T-P (2002). 'The impact of national health insurance on the notification of tuberculosis in Taiwan'. *The International Journal of Tuberculosis and Lung Disease,* vol 6, pp 974–79.

Christensen DB, Neil N, Fassett WE, Smith DH, Holmes G and Stergachis A (2000). 'Frequency and characteristics of cognitive services provided in response to a financial incentive'. *Journal of the American Pharmaceutical Association,* vol 40, pp 609–17.

Christianson JB, Knutson DJ and Mazze RS (2006). 'Physician pay-for-performance: Implementation and research issues'. *Journal of General Internal Medicine,* vol 21 (Suppl 2), S9–S13.

Christianson JB, Reidel AA, Abelson DJ, Hamer RL, Knutson DJ and Taylor RA (2001). *Managed Care and the Treatment of Chronic Illness.* Sage Publications Inc.

Chung RS, Chernicoff HO, Nakao KA, Nickel RC and Legorreta AP (2003). 'A quality-driven physician compensation model: Four-year follow-up study'. *Journal for Healthcare Quality,* vol 25, pp 31–37.

Coleman T, Wynn AT, Stevenson K and Cheater F (2001). 'Qualitative study of pilot payment aimed at increasing general practitioners' antismoking advice to smokers'. *British Medical Journal,* vol 323, pp 432–35.

Conrad DA and Christianson JB (2004). 'Penetrating the 'black box': financial incentives for enhancing the quality of physician services'. *Medical Care Research and Review,* vol 61 (3 Suppl), pp 37S–68S.

Conrad DA, Maynard C, Cheadle A, Ramsey S, Marcus-Smith M, Kirz H, Madden CA, Martin D, Perrin EB, Wickizer T, Zierler B, Ross A, Noren J and Liang SY (1998). 'Primary care physician compensation method in medical groups: does it influence the use and cost of health services for enrollees in managed care organisations?' *Journal of the American Medical Association,* vol 279, pp:853–58.

Conrad DA, Sales A, Liang S-Y, Chaudhuri A, Maynard C, Pieper L, Weinstein L, Gans D and Piland N (2002). 'The impact of financial incentives on physician productivity in medical groups'. *Health Services Research,* vol 37, pp 885–906.

Conrad DA, Saver BG, Court B and Heath S (2006). 'Paying physicians for quality: evidence and themes from the field'. *Joint Commission Journal on Quality and Patient Safety,* vol 32, pp 443–51.

Curtin K, Beckman H, Pankow G, Milillo Y and Greene RA (2006). 'Return on investment in pay for performance: a diabetes case study'. *Journal of Healthcare Management,* vol 51, pp 365–76.

Dafney LS (2005). 'How do hospitals respond to price changes?' *The American Economic Review,* vol 95, pp 1525–47.

Damberg CL, Raube K, Williams T and Shortell SM (2005). 'Paying for performance: implementing a statewide project in California'. *Quality Management in Health Care,* vol 14, pp 66–79.

Davidson SM, Manheim LM, Werner SM, Hohlen MM, Yudkowsky BK and Fleming GV (1992). 'Prepayment with office-based physicians in publicly funded programmes: Results from the Children's Medicaid Program'. *Pediatrics,* vol 89, pp 761–67.

Davis LM, Wells KB, Rogers WH, Benjamin B, Norquist G, Kahn K, Kosecoff J and Brook R (1995). 'Effects of Medicare's prospective payment system on service use by depressed elderly inpatients'. *Psychiatric Services,* vol 46, pp 1178–84.

Department of Health (2006). *Practice Based Commissioning: Practical implementation.* Department of Health.

Department of Health (2002). *Reforming NHS Financial Flows: Introducing payment by results.* Department of Health.

Dhanani N, O'Leary JF, Keeler E, Bamezai A and Melnick G (2004). 'The effect of HMOs on the inpatient utilisation of Medicare beneficiaries'. *Health Services Research,* vol 39, pp 1607–1726.

Docteur E and Oxley H (2003). *Health-Care Systems: Lessons from the reform experience.* OECD Health Working Papers.

Doran T, Fullwood C, Gravelle H, Reeves D, Kontopantelis E, Hiroeh U and Roland M (2006). 'Pay-for-performance programmes in family practices in the United Kingdom'. *The New England Journal of Medicine,* vol 335, pp 375–84.

Dudley RA (2005). 'Pay-for-performance research. How to learn what clinicians and policy makers need to know'. *Journal of the American Medical Association,* vol 294, pp 1821–23.

Dudley RA, Frolich A, Robinowitz DL, Talavera JA, Broadhead P, Luft HS and McDonald K (2004). *Strategies to Support Quality-Based Purchasing: A review of the evidence.* Technical Review Number 10. AHRQ Publication No. 04-0057, Agency for Healthcare Research and Quality.

Dudley RA, Miller RH, Korenbrot TY and Luft HS (1998). 'The impact of financial incentives on quality of health care'. *The Milbank Quarterly,* vol 76, pp 649–86.

Duggan M (2004). 'Does contracting out increase the efficiency of government programmes? Evidence from Medicaid HMOs'. *Journal of Public Economics,* vol 88, pp 2549–72.

Dusheiko M, Gravelle H, Jacobs R et al (2003). *The Effects of Budgets on Doctors' Behaviour: Evidence from a natural experiment.* CMPO Working Paper Series No. 03/064. The Centre for Market and Public Organisation.

Eddy D (2005). 'Evidence-based medicine: a unified approach'. *Health Affairs,* vol 24, pp 9–17.

Eggleston K and Hsieh C-R (2004). 'Healthcare payment incentives: a comparative analysis of reforms in Taiwan, South Korea and China'. *Applied Health Economics and Health Policy,* vol 3, pp 47–56.

Eisenberger R and Cameron J (1996). 'Detrimental effects of reward: reality or myth?' *American Psychologist,* vol 51, pp 1153–66.

Eisenhardt KM (1989). 'Agency theory: an assessment and review'. *Academy of Management Review,* vol 14, pp 57–74.

Ellis RP and McGuire TG (1996). 'Hospital response to prospective payment: moral hazard, selection, and practice-style effects'. *Journal of Health Economics,* vol 15, pp 257–77.

Epstein AM (2006). 'Paying for performance in the United States and abroad'. *The New England Journal of Medicine,* vol 355, pp 406–08.

Ettner SL, Thompson TJ, Stevens MR, Mangione CM, Kim C, Steers WN, Goewey J, Brown AF, Chung RS, Vankat Narayan KM and the TRIAD Study Group (2006). 'Are physician reimbursement strategies

associated with processes of care and patient satisfaction for patients with diabetes in managed care?' *Health Services Research,* vol 41, pp 1221–41.

Fairbrother G, Hanson KL, Friedman S and Butts GC (1999). 'The impact of physician bonuses, enhanced fees, and feedback on childhood immunization coverage rates'. *American Journal of Public Health,* vol 89, pp 171–75.

Fairbrother G, Siegel MJ, Friedman S, Kory PD and Butts GC (2001). 'Impact of financial incentives on documented immunization rates in the inner city: results of a randomised controlled trial'. *Ambulatory Pediatrics,* vol 1, pp 206–12.

Felt-Lisk S, Gimm G and Peterson S (2007). 'Making pay-for-performance work in Medicaid'. *Health Affairs – Web Exclusive,* June 26, w516–w527.

Ferman JH (2004). 'Pay for performance: obstacles/implications. Despite challenges, pay-for-performance programmes are here to stay'. *Healthcare Executive,* vol 19, pp 44, 46.

Fisher ES (2006). 'Paying for performance: risks and recommendations'. *The New England Journal of Medicine,* vol 355, pp 1845–47.

Flynn KE, Smith MA and Davis MK (2002). 'From physician to consumer: the effectiveness of strategies to manage health care utilisation'. *Medical Care Research and Review,* vol 59, pp 455–81.

Forsberg E, Axelsson R and Arnetz B (2001). 'Financial incentives in health care. The impact of performance-based reimbursement'. *Health Policy,* vol 58, pp 243–62.

Forsberg E, Axelsson R and Arnetz B (2000). 'Effects of performance-based reimbursement in healthcare'. *Scandinavian Journal of Public Health,* vol 28, pp 102–10.

Gagné M and Deci EL (2005). 'Self-determination theory and work motivation'. *Journal of Organisational Behaviour,* vol 26, pp 331–62.

Galvin R (2006). 'Pay-for-performance: too much of a good thing? A conversation with Martin Roland'. *Health Affairs – Web Exclusive,* w412–w419.

Gaynor M, Rebitzer JB, Taylor LJ (2001). *Incentives in HMOs.* NBER Working Paper 8522. National Bureau of Economic Research.

Gené-Badia J, Escaramis-Babiano G, Sans-Corrales M, Sampietro-Colom L, Aguado-Menguy F, Cabezas-Peña C and Gallo de Puelles P (2007). 'Impact of economic incentives on quality of professional life and on end-user satisfaction in primary care'. *Health Policy,* vol 80, pp 2–10.

Gerdtham UG, Löthgren M, Tambour M, Rehnberg C (1999). 'Internal markets and health-care efficiency: a multiple-output stochastic frontier analysis'. *Health Economics,* vol 8, pp 151–64.

Gerdtham UG, Rehnberg C and Tambour M (1999). 'The impact of internal markets on health-care efficiency: Evidence from health-care reforms in Sweden'. *Applied Economics,* vol 32, pp 935–45.

Gerhart B and Rynes SL (2003). *Compensation. Theory, Evidence, and Strategic Implications.* Sage Publications.

Glickman SW, Ou F-S, DeLong ER, Roe MT, Lytle BL, Mulgund J, Rumsfeld JS, Gibler WB, Ohman EM, Schulman KA and Peterson ED (2007). 'Pay for performance, quality of care, and outcomes in acute myocardial infarction'. *Journal of the American Medical Association,* vol 297, pp 2373–80.

Goodman RM (2006). 'Relationship between primary care physician financial risk and member emergency department use in a commercial HMO population'. *American Journal on Managed Care,* vol 12, pp 329–40.

Grabowski DC, Stevenson DG, Huskamp HA and Keating NL (2006). 'The influence of Medicare home health payment incentives: does payer source matter?' *Inquiry,* vol 43, pp 135–49.

Grady KE, Lemkau JP, Lee NR and Caddell C (1997). 'Enhancing mammography referral in primary care'. *Preventive Medicine,* vol 26, pp 791–800.

Greene RA, Beckman H, Chamberlain J, Partridge G, Miller M, Burden D and Kerr J (2004). 'Increasing adherence to a community-based guideline for acute sinusitis through education, physician profiling and financial incentives'. *The American Journal of Managed Care,* vol 10, pp 670–78.

Grootendorst P and Stewart D (2006). 'A re-examination of the impact of reference pricing on anti-hypertensive drug plan expenditures in British Columbia'. *Health Economics,* vol 15, pp 735–42.

Grossbart SR (2006). 'What's the return? Assessing the effect of 'pay-for-performance' initiatives on the quality of care delivery'. *Medical Care Research & Review,* vol 63 (1 Suppl), pp 29S–48S.

Gruber J and Owings M (1994). *Physician Financial Incentives and Cesarean Section Delivery.* NBER Working Paper 4933. National Bureau of Economic Research.

Grumbach K, Osmond D, Vranizan K, Jaffe D and Bindman AB (1998). 'Primary care physicians' experience of financial incentives in managed-care systems'. *New England Journal of Medicine,* vol 339, pp 1516–21.

Hadley J, Mandelblatt JS, Mitchell JM, Weeks JC, Guadagnoli E, Hwang Y-T and OPTIONS Research Team (2003). 'Medicare breast surgery fees and treatment received by older women with localized breast cancer'. *Health Services Research,* vol 38, pp 553–573.

Hadley J, Mitchell JM, Sulmasy DP and Bloche MG (1999). 'Perceived financial incentives, HMO market penetration, and physicians' practice styles and satisfaction'. *Health Services Research,* vol 34, pp 307–21.

Haile PA and Stein RM (2002). 'Managed care incentives and inpatient complications'. *Journal of Economics & Management Strategy,* vol 11, pp 37–79.

Hamilton V (1993). 'The Medicare hospice benefit: The effectiveness of price incentives in health care policy'. *RAND Journal of Economics,* vol 24, pp 605–24.

Hamilton BH and Bramley-Hasher RE (1999). 'The impact of the NHS reforms on queues and surgical outcomes in England: Evidence from hip fracture patients'. *The Economic Journal,* vol 109, pp 437–62.

Hanson KL, Fairbrother G, Kory P, Butts GC and Friedman S (1998). 'The transition from Medicaid fee-for-service to managed care among private practitioners in New York City: effect on immunization and screening rates'. *Maternal and Child Health Journal,* vol 2, pp 5–14.

Hellinger FJ (1996). 'The impact of financial incentives on physician behaviour in managed care plans: a review of the evidence'. *Medical Care Research and Review,* vol 53, pp 294–314.

Hillman AL, Pauly MV and Kerstein JJ (1989). 'How do financial incentives affect physicians' clinical decisions and the financial performance of health maintenance organisations?' *New England Journal of Medicine,* vol 321, pp 86–92.

Hillman AL, Ripley K, Goldfarb N, Nuamah I, Weiner J and Lusk E (1998). 'Physician financial incentives and feedback: failure to increase cancer screening in Medicaid managed care'. *American Journal of Public Health,* vol 88, pp 1699–1701.

Hillman AL, Ripley K, Goldfarb N, Weiner J, Nuamah I and Lusk E (1999). 'The use of physician financial incentives and feedback to improve pediatric preventive care in Medicaid managed care'. *Pediatrics,* vol 104, pp 931–35.

Hodgkin D and McGuire TG (1994). 'Payment levels and hospital response to prospective payment'. *Journal of Health Economics,* vol 13, pp 1–29.

Hofer TG, Hayward RA, Greenfield S, Wagner EH, Kaplan SH and Manning WG (1999). 'The unreliability of individual physician 'report cards' for assessing the costs and quality of care of a chronic disease'. *Journal of the American Medical Association,* vol 281, pp 2098–105.

Huang CC, Chung K-P, Kuo N-C and Hung CL (2005). 'Effectiveness of coping strategies used by hospitals in response to implementation of a case-based payment system by the National Health Insurance Program'. *Journal of the Formosan Medical Association,* vol 104, pp 468–75.

Hussey PS, Anderson GF, Osborn R, Feek C, McLaughlin V, Millar J and Epstein A (2004). 'How does the quality of care compare in five countries?' *Health Affairs,* vol 23, pp 89–99.

Hutchison B, Birch S, Hurley J, Lomas J and Stratford-Devai F (1996). 'Do physician-payment mechanisms affect hospital utilisation? A study of Health Service Organisations in Ontario'. *Canadian Medical Association Journal,* vol 154, pp 653–61.

Inkelas M (2005). 'Incentives in a Medicaid carve-out: impact on children with special health care needs'. *Health Services Research,* vol 40, pp 79–102.

Institute of Medicine (2001). *Crossing the Quality Chasm: A new health system for the 21st century.* National Academy Press.

Institute of Medicine, *Medicare Strategy for Quality Assurance,* vol. 1, K. Lohr, editor, National Academy Press, Washington, DC, 1990.

Intrator O and Mor V (2004). 'Effect of state Medicaid reimbursement rates on hospitalizations from nursing homes'. *Journal of the American Geriatrics Society,* vol 52, pp 393–98.

Iversen T (1998). 'Clinical response of salaried consultants to economic incentives'. *Developments in Health Economics & Public Policy,* vol 6, pp 137–58.

Kahn III CN, Ault T, Isenstein H, Potetz L and Van Gelder S (2006). 'Snapshot of hospital quality reporting and pay-for-performance under Medicare'. *Health Affairs,* vol 25, pp 148–62.

Keating NL, Landon BE, Ayanian JZ, Borbas C and Guadagnoli E (2004). 'Practice, clinical management, and financial arrangements of practicing generalists: are they associated with satisfaction?' *Journal of General Internal Medicine,* vol 29, pp 410–18.

King's Fund (2005). *Payment By Results.* King's Fund.

Kjerstad E (2003). 'Prospective funding of general hospitals in Norway: incentives for higher production H?' *International Journal of Health Care Finance and Economics,* vol 3, pp 235–51.

Kouides RW, Bennett NM, Lewis B, Cappuccio JD, Barker WH and LaForce FM (1998). 'Performance-based physician reimbursement and influenza immunization rates in the elderly. The primary-care physicians of Monroe County'. *American Journal of Preventive Medicine,* vol 14, pp 89–95.

Kristiansen IS and Holtedahl K (1993). 'Effect of the remuneration system on the general practitioner's choice between surgery consultations and home visits'. *Journal of Epidemiology & Community Health,* vol 47, pp 481–84.

Kuykendall DH, Johnson ML and Geraci JM (1995). 'Expected source of payment and use of hospital services for coronary atherosclerosis'. *Medical Care,* vol 33, pp 715–28.

Langa KM and Sussman EJ (1993). 'The effect of cost-containment policies on rates of coronary revascularization in California'. *The New England Journal of Medicine,* vol 329, pp 1784–89.

Langham S, Gillam S and Thorogood M (1995). 'The carrot, the stick and the general practitioner: how have changes in financial incentives affected health promotion activity in general practice?' *British Journal of General Practice,* vol 45, pp 665–68.

Larsen DL, Cannon W and Towner S (2003). 'Longitudinal assessment of a diabetes care management system in an integrated health network'. *Journal of Managed Care Pharmacy,* vol 9, pp 552–58.

Lave JR and Frank RG (1990). 'Effect of the structure of hospital payment on length of stay'. *Health Services Research,* vol 25, pp 327–47.

Levin-Scherz J, DeVita N and Timbie J (2006). 'Impact of pay-for-performance contracts and network registry on diabetes and asthma HEDIS measures in an integrated delivery network'. *Medical Care Research & Review,* vol 63(1 Suppl), pp 14S–28S.

Levinson A and Ullman F (1998). 'Medicaid managed care and infant health'. *Journal of Health Economics,* vol 17, pp 351–68.

Lindenauer PK, Remus D, Roman S, Rothberg MB, Benjamin EM, Ma A and Bratzler DW (2007). 'Public reporting and pay for performance in hospital quality improvement'. *The New England Journal of Medicine,* vol 356, pp 486–96.

Lindrooth RC, Bazzoli GJ and Clement J (2007). 'The effect of reimbursement on the intensity of hospital services'. *Southern Economic Journal,* vol 73, pp 575–87.

Lindrooth RC and Weisbrod BA (2007). 'Do religious nonprofit and for-profit organizations respond differently to financial incentives? The hospice industry'. *Journal of Health Economics,* vol 26, pp 342–57.

Liu X and Mills A (2003). 'The influence of bonus payments to doctors on hospital revenue: results of a quasi-experimental study'. *Applied Health Economics and Health Policy,* vol 2, pp 91–98.

Luft HS (2003). 'Variations in patterns of care and outcomes after acute myocardial infarction for Medicare beneficiaries in fee-for-service and HMO settings'. *Health Services Research,* vol 38, pp 1065–79 and S1–S12.

Luft HS (1988). 'HMOs and the quality of care'. *Inquiry,* vol 25, pp 147–56.

Luft HS (1980). 'Assessing the evidence on HMO performance'. *Milbank Memorial Fund Quarterly,* vol 58, pp 501–36.

Madison K (2004). 'Hospital-physician affiliations and patient treatments, expenditures, and outcomes'. *Health Services Research,* vol 39, pp 257–78.

Mannion R and Street A (2005). *Payment by Results and Demand Management: Learning from the South Yorkshire laboratory.* Centre for Health Economics, University of York.

Mays N, Mulligan J-A, Goodwin N (2000). 'The British quasi-market in health care: a balance sheet of the evidence'. *Journal of Health Services Research & Policy,* vol 5, pp 49–58.

McCleary K, Asubonteng J and Munchus G (1995). 'The effect of financial incentives on physicians' behaviour in health maintenance organisations'. *Journal of Management in Medicine,* vol 9, pp 8–26.

McDonald R, Harrison S, Checkland K, Campbell SM and Roland M (2007). 'Impact of financial incentives on clinical autonomy and internal motivation in primary care: ethnographic study'. *British Medical Journal,* vol 334, pp 1357–62.

McElduff P, Lyratzopoulos G, Edwards R, Heller RF, Shekelle P and Roland M (2004). 'Will changes in primary care improve health outcomes? Modeling the impact of financial incentives introduced to improve quality of care in the UK'. *Quality and Safety in Health Care,* vol 13, pp 191–97.

McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A and Kerr EA (2003). 'The quality of health care delivered to adults in the United States'. *New England Journal of Medicine,* vol 348, pp 2635–45.

McMahon Jr LF, Hofer TP and Hayward RA (2007). 'Physician-level P4P–DOA? Can quality-based payment be resuscitated?' *The American Journal of Managed Care,* vol 13, pp 233–36.

Meltzer D and Chung J (2001). *Effects of Competition Under Prospective Payment on Hospital Costs Among High and Low Cost Admissions: Evidence from California.* NBER Working Paper 8069. National Bureau of Economic Research.

Mikkola H (2003). 'Hospital pricing reform in the public health care system: an empirical case study from Finland'. *International Journal of Health Care Finance and Economics,* vol 3, pp 267–86.

Milcent C (2005). 'Hospital ownership, reimbursement systems and mortality rates'. *Health Economics,* vol 14, pp 1151–68.

Milgate K and Chen SB (2006). 'Pay-for-performance: the MedPAC perspective'. *Health Affairs,* vol 25, pp 413–19.

Milgrom P and Roberts P (1992). *Economics, Organisation, and Management.* Prentice Hall.

Miller RH and Luft HS (1997). 'Does managed care lead to better or worse quality of care?' *Health Affairs,* vol 16, pp 7–25.

Miller RH and Luft HS (1994). 'Managed care plan performance since 1980: a literature analysis'. *The Journal of the American Medical Association,* vol 271, pp 1512–19.

Mitchell JM (1996). 'Physician joint ventures and self-referral: an empirical perspective' in Spece RG, Shimm DS and Buchanan AE eds *Conflicts of Interest in Clinical Practice and Research.* Oxford University Press.

Mitchell JM and Gaskin DJ (2007). 'Caregivers' ratings of access. Do children with special health care needs fare better under fee-for-service or partially capitated managed care?' *Medical Care,* vol 45, pp 146–53.

Mitchell JM and Gaskin DJ (2005). 'Factors affecting plan choice and unmet need among supplemental security income eligible children with disabilities'. *Health Services Research,* vol 40, pp 1379–99.

Morrow RW, Gooding AD and Clark C (1995). 'Improving physicians' preventive health care behaviour through peer review and financial incentives'. *Archives of Family Medicine,* vol 4, pp 165–69.

Mukamel DB, Zwanziger J and Tomaszewski KJ (2001). 'HMO penetration, competition, and risk-adjusted hospital mortality'. *Health Services Research,* vol 36, pp 1019–35.

Munoz E, Sterman H, Cohen J, Goldstein J, Margolis IB and Wise L (1988a). 'Financial risk, hospital cost, complications, and comorbidities in surgical noncomplication- and noncomorbidity-stratified diagnostic related groups'. *Annals of Surgery,* vol 207, pp 305–09.

Munoz E, Sterman H, Goldstein J, Chalfin D, Cohen J and Wise L (1988b). 'Financial risk, hospital cost, and complications and comorbidities (CCs) in the non-CC-stratified pulmonary medicine diagnostic-related group Medicare hospital payment system'. *American Review of Respiratory Disease,* vol 137, pp 998–1001.

Munoz E, Sterman H, Patel P, Chalfin D, Mulloy K and Wise L (1988c). 'Financial risk, hospital cost, and complications and comorbidities (CCs) in non-CC stratified neurosurgical diagnostic related groups (DRGs)'. *Neurosurgery,* vol 22, pp 955–60.

Munoz E, Sterman H, Friedman R, Goldstein J, Gerold T and Wise L (1988d). 'Financial risk, hospital cost, complications and comorbidities (CCc) in non-CC stratified urology diagnostic related groups'. *Urology*, vol 32, pp 380–84.

Munoz E, Thies H, Maesaka JK, Angus G, Goldstein J and Wise L (1988e). 'Diagnosis related groups, resource utilization, age, and outcome for hospitalized nephrology patients'. *American Journal of Kidney Diseases,* vol 11, pp 481–88.

Murkofsky RL, Phillips RS, McCarthy EP, Davis RB and Hamel MB (2003). 'Length of stay in home care before and after the 1997 Balanced Budget Act'. *Journal of the American Medical Association,* vol 289, pp 2841–48.

Mushlin AI, Panzer RJ, Black ER, Greenland P and Regenstreif DI (1988). 'Quality of care during a community-wide experiment in prospective payment to hospitals'. *Medical Care,* vol 26, pp 1081–91.

Nahra TA, Reiter KL, Hirth RA, Shermer JE, Wheeler JRC (2006). 'Cost-effectiveness of hospital pay-for-performance incentives'. *Medical Care Research and Review,* vol 63, pp 49S–72S.

Nalli GA, Scanlon DP and Libby D (2007). 'Developing a performance-based incentive program for hospitals: a case study from Maine'. *Health Affairs,* vol 26, pp 817–24.

Nassiri A and Rochaix L (2006). 'Revisiting physicians' financial incentives in Quebec: a panel system approach'. *Health Economics,* vol 15, pp 59–64.

Nichols L and O'Malley A (2006). 'Hospital payment systems: will payers like the future better than the past?' *Health Affairs,* vol 25, pp 81–93.

Nicholson S and Song D (2001). 'The incentive effects of the Medicare indirect medical education policy'. *Health Economics,* vol 20, pp 909–33.

Norton EC, Van Houtven CH, Lindrooth RC, Normand S-LT and Dickey B (2002). 'Does prospective payment reduce inpatient length of stay?' *Health Economics,* vol 22, pp 377–87.

O'Malley AJ, Frank RG, Kaddis A, Rothenberg BM and McNeil BJ (2006). 'Impact of alternative interventions on changes in generic dispensing rates'. *Health Services Research,* vol 41, pp 1876–94.

Oleske DM, Branca ML, Schmidt JB, Ferguson R and Linn ES (1998). 'A comparison of capitated and fee-for-service Medicaid reimbursement methods on pregnancy outcomes'. *Health Services Research,* vol 33, pp 55–73.

Pantilat SZ, Chesney M and Lo B (1999). 'Effect of incentives on the use of indicated services in managed care.' *The Western Journal of Medicine,* vol 170, pp 137–42.

Petersen LA, Woodard LD, Urech T, Daw C and Sookanan S (2006). 'Does pay-for-performance improve the quality of health care?' Annals of Internal Medicine, vol 145, pp 265–72.

Pink GH, Brown AD, Studer ML, Reiter KL and Leatt P (2006). 'Pay-for-performance in publicly financed healthcare: some international experience and considerations for Canada'. *Healthcare Papers,* vol 6, pp 8–26.

Pockney P, George S, Primrose J, Smith H, Kinley H, Little P, Lattimer V, Lowy A and Kneebone R (2004). 'Impact of the introduction of fee for service payments on types of minor surgical procedures undertaken by general practitioners: observational study'. *Journal of Public Health,* vol 26, pp 264–67.

Potosky AL, Merrill RM, Riley GF, Taplin SH, Barlow W, Fireman BH and Lubitz JD (1999). 'Prostate cancer treatment and ten-year survival among group/staff HMO and fee-for-service Medicare patients'. *HSR: Health Services Research,* vol 34, pp 525–46.

Pratt W and Zeckhauser RJ (1985). *Principals and Agents: The structure of business.* Harvard University Press.

Prendergast C (1999). 'The provision of incentives in firms'. *Journal of Economic Literature,* vol 37, pp 7–63.

Ranson SB, McNeeley SG, Kruger ML, Doot G and Cotton DB (1996). 'The effect of capitated and fee-for-service remuneration on physician decision making in gynecology'. *Obstetrics & Gynecology,* vol 87, pp 707–10.

Reschovsky JD and Hadley J (2007). 'Physician financial incentives: use of quality incentives inches up, and productivity still dominates'. Issue Brief No. 108. Center for Studying Health System Change. Accessed 5 January 2007 at www.hschange.org/CONTENT/905/?PRINT=1

Rice T (1997). 'Physician payment policies: impacts and implications'. *Annual Review of Public Health,* vol 18, pp 549–69.

Ritchie LD, Bisset AF, Russell D, Leslie V and Thomson I (1992). 'Primary and preschool immunization in Grampian: progress and the 1990 contract'. *British Medical Journal,* vol 34, pp 816–19.

Robinson JC (2001). 'Theory and practice in the design of physician payment incentives'. *The Milbank Quarterly,* vol 79, pp 149–77.

Robinson JC (1999). 'Blended payment methods in physician organisations under managed care'. *Journal of the American Medical Association,* vol 282, pp 1258–63.

Roland M (2004). 'Linking physicians' pay to the quality of care: a major experiment in the United Kingdom'. *The New England Journal of Medicine,* vol 251, pp 1448–54.

Rosenthal MB (2007). 'P4P: rumors of its demise may be exaggerated'. *The American Journal of Managed Care,* vol 13, pp 238–39.

Rosenthal MB, Fernandopulle R, Song HR and Landon B (2004). 'Paying for quality: Providers' incentives for quality improvement'. *Health Affairs,* vol 23, pp 127–41.

Rosenthal MB and Frank RG (2006). 'What is the empirical basis for paying for quality in health care?' *Medical Care Research & Review,* vol 63, pp 135–57.

Rosenthal MB, Frank RG, Li Z and Epstein AM (2005). 'Early experience with pay-for-performance: from concept to practice'. *Journal of the American Medical Association,* vol 294, pp 1788–93.

Rosenthal MB, Landon BE, Normand S-LT, Frank RG and Epstein AM (2006). 'Pay for performance in commercial HMOs'. *The New England Journal of Medicine,* vol 335, pp 1895–902.

Rynes SL, Gerhart B and Parks L (2005). 'Personnel psychology: performance evaluation and pay for performance'. *Annual Review of Psychology,* vol 56, pp 571–600.

Safran DG, Rogers WH, Tarlov AR, Inui T, Taira DA, Montgomery JE, Ware JE and Slavin CP (2000). 'Organisational and financial characteristics of health plans: are they related to primary care performance?' *Archives of Internal Medicine,* vol 160, pp 69–76.

Sappington DEM (1991). 'Incentives in principal-agent relationships'. *Journal of Economic Perspectives,* vol 5, pp 45–66.

Schoen C, Osborn R, Huynh PT, Doty M, Davis K, Zapert K and Peugh J (2004). 'Primary care and health system performance: adults' experiences in five countries'. *Health Affairs,* Suppl Web Exclusives, pp W4-487–503.

Schoen C, Osborn R, Huynh PT, Doty M, Zapert K, Peugh J and Davis K (2005). 'Taking the pulse of health care systems: experiences of patients with health problems in six countries'. *Health Affairs – Web Exclusive* 1 Nov. 2005:W5–509–W5–525.

Schoenman JA (2005). 'Rural implications of Medicare's post-acute-care transfer payment policy'. *The Journal of Rural Health,* vol 21, pp 122–30.

Scott A and Hall J (1995). 'Evaluating the effects of GP remuneration: problems and prospects'. *Health Policy,* vol 21, pp 183–95.

Shen J, Andersen R, Brook R, Kominski G, Albert PS and Wenger N (2004). 'The effects of payment method on clinical decision-making: physician responses to clinical scenarios'. *Medical Care,* vol 42, pp 297–302.

Shmueli A, Intrator O and Israeli A (2002). 'The effects of introducing prospective payments to general hospitals on length of stay, quality of care, and hospitals' income: the early experience of Israel'. *Social Science & Medicine,* vol 55, pp 981–89.

Shorten B and Shorten A (2004). 'Impact of private health insurance incentives on obstetric outcomes in NSW hospitals'. *Australian Health Review,* vol 27, pp 27–38.

Silcock J and Ratcliffe J (1996). 'The 1990 GP contract: meeting needs?' *Health Policy,* vol 36, pp 199–207.

Simpson CR, Hannaford PC, Lefevre K and Williams D (2006). 'Effect of the UK incentive-based contract on the management of patients with stroke in primary care'. *Stroke,* vol 37, pp 2354–60.

Sliwka D (2007). 'Trust as a signal of a social norm and the hidden costs of incentive schemes'. *The American Economic Review,* vol 97, pp 999–1012.

Smith P and York N (2004). 'Quality incentives: the case of UK general practitioners – an ambitious UK quality improvement initiative offers the potential for enormous gains in the quality of primary health care'. *Health Affairs,* vol 23, pp 112–18.

Smith RD and Wilton P (1998). 'General practice fundholding: progress to date'. *British Journal of General Practice,* vol 48, pp 1253–57.

Spooner A, Chapple A and Roland M (2001). 'What makes British general practitioners take part in a quality improvement scheme?' *Journal of Health Services & Research Policy,* vol 6, pp 145–50.

Srirangalingam U, Sahathevan SK, Lasker SS and Chowdhury TA (2006). 'Changing pattern of referral to a diabetes clinic following implementation of the new UK GP contract'. *British Journal of General Practice,* vol 56, pp 624–26.

St Jacques PJ, Patel N and Higgins MS (2004). 'Improving anesthesiologist performance through profiling and incentives'. *Journal of Clinical Anesthesia,* vol 16, pp 523–28.

Stanton S and Shortt SED (2003). 'The influence of payment method on patterns of physician practice: Experience at a Canadian academic health center'. *Research in Healthcare Financial Management,* vol 8, pp 43–58.

Steiner A and Robinson R (1998). 'Managed care: US research evidence and its lessons for the NHS'. *Journal of Health Services Research and Policy,* vol 3, pp 173–84.

Sutton M and McLean G (2006). 'Determinants of primary medical care quality measured under the new UK contract: cross sectional study'. *British Medical Journal,* vol 332, pp 389–90.

Tai-Seale M, Rice TH and Stearns SC (1998). 'Volume responses to Medicare payment reductions with multiple payers: a test of the McGuire-Pauly model'. *Health Economics,* vol 7, pp 199–219.

Town R, Kane R, Johnson P and ButlerM (2005). 'Economic incentives and physicians' delivery of preventive care: a systematic review'. *American Journal of Preventive Medicine,* vol 28, pp 234–40.

Town R, Wholey DR, Kralewski J and Dowd B (2004). 'Assessing the influence of incentives on physicians and medical groups'. *Medical Care Research and Review,* vol 61, pp 80S–118S.

Tufano J, Conrad DA, Sales A, Maynard C, Noren J, Kezirian E, Schellhase KG and Liang SY (2001). 'Effects of compensation method on physician behaviours'. *American Journal of Managed Care,* vol 7, pp 363–73.

Udvarhelyi IS, Jennison K, Phillips RS and Epstein AM (1991). 'Comparison of the quality of ambulatory care for fee-for-service and prepaid patients'. *Annals of Internal Medicine,* vol 115, pp 394–400.

Wee CC, Phillips RS, Burstin HR, Cook EF, Puopolo AL, Brennan TA and Haas JS (2001). 'Influence of financial productivity incentives on the use of preventive care'. *American Journal of Medicine,* vol 110, pp 181–87.

Wennberg J and Gittelsohn A (1973). 'Small area variations in health care delivery'. *Science,* vol 82, pp 1102–08.

White C (2003). 'Rehabilitation therapy in skilled nursing facilities: effects of Medicare's new prospective payment system'. *Health Affairs,* vol 22, pp 214–23.

Williams S and Buchan J (2006). *Assessing the New NHS Consultant Contract: A something for something deal?* King's Fund.

Wodchis WP, Fries BE and Pollack H (2004). 'Payer incentives and physical rehabilitation therapy for nonelderly institutional long-term care residents: evidence from Michigan and Ontario'. *Archives of Physical Medicine & Rehabilitation,* vol 85, pp 210–17.

Wolfe BL and Detmer DE (1988). 'Diagnosis related groups (DRGs): efficiency and/or quality producing? Some evidence from surgical patients'. *Health Services Management Research,* vol 1, pp 19–28.

Xirasagar S and Lin HC (2006). 'Effects of payment incentives, hospital ownership and competition on hospitalization decisions for ambulatory surgical procedures'. *Health Policy,* vol 76, pp 26–37.

Yip W and Eggleston K (2001). 'Provider payment reform in China: the case of hospital reimbursement in Hainan Province'. *Health Economics,* vol 10, pp 325–39.

Young GJ and Conrad DA (2007). 'Practical issues in the design and implementation of pay-for-quality programs'. *Journal of Healthcare Management,* vol 52, pp 10–19.

Young GJ, Meterko M, White B, Bokhour BG, Stautter KM, Berlowitz D and Burgess Jr JF (2007). 'Physician attitudes toward pay-for-quality programs: perspectives from the front line'. *Medical Care Research and Review,* vol 64, pp 331–43.

Zinn JS, Mor V, Intrator O, Feng Z, Angelelli J and Davis JA (2003). 'The impact of the prospective payment systems for skilled nursing facilities on therapy service provision: a transaction cost approach'. *Health Services Research,* vol 36, pp 1467–85.

APPENDIX 1
Financial incentives, utilisation and costs

As a general matter, financial incentives are designed by payers to influence the utilisation of healthcare services by consumers. Typically, the utilisation of healthcare services is modelled as the result of the joint decisions of providers and consumers. In an economic paradigm, consumers demand medical care and providers supply it. However, an unusual characteristic of this exchange is that patients expect their providers to play the role of trusted patient 'agent' in the decision process. Providers possess better information about the consequences of treatment alternatives, and the consumer trusts the provider to convey that information in an unbiased manner, typically accompanied by a treatment recommendation.

Financial incentives have the potential to influence the recommendations of providers as agents of their patients. Indeed, the intent of different payment arrangements employed by payers often is to alter the interaction of providers with consumers in reaching a decision about the utilisation of services. For example, if a primary care physician is 'at risk' for the cost of referrals to specialists (such as through a capitation payment that includes both primary and specialty services), the utilisation of specialty care will be less likely than if the same physician is at risk only for primary care. In this case, the primary care physician may be more likely to suggest to patients that she/he can provide the service (specialty care is not needed) or to recommend that the patient seek specialty care from a relatively low-cost provider.

The degree to which either type of provider behaviour is observed depends on many factors in addition to the amount of financial reward at stake, including the willingness of patients to accept the advice of their primary care physicians and the ability of primary care physicians to reconcile their recommendations with their role of patient advocate, or agent, for the patient's best interest. In any event, if either action were taken, it might or might not have an impact on costs or quality. Providing services in the primary care physician's office could result in lower costs and have no effect on quality. Alternatively, the lower cost specialist may provide higher quality care than delivered by more expensive specialists who might otherwise receive the referral.

The greater the degree to which the introduction of new financial payment arrangements for providers, or the altering of existing incentives, affects utilisation, the more likely that research will be able to detect changes in costs, quality or both. However, not all research studies are designed to capture cost and quality impacts of utilisation changes. This may reflect a lack of data, especially related to quality of care, or the research may focus, by design, on the outcome of most interest to policy-makers. Given the sometimes dramatic increases in healthcare costs experienced by purchasers/payers, the impact of payment systems on costs often has been the primary, or only, focus of research studies. However, in many studies the analysis has been content to stop at documenting and explaining utilisation changes, assuming that decreases in utilisation of the service in question will ultimately mean lower costs. The obvious flaw in this reasoning is that use of other services may increase as a consequence of lower use of the service targeted by the change in incentives, with uncertain consequences for overall costs.

Historical context

Interest in the impact of financial incentives on provider behaviour is longstanding, but the volume of literature on this topic began to grow in parallel to the growth of HMOs in the USA during the 1970s and 1980s. During that period, HMOs were organisations that contracted with providers (creating provider networks) and accepted capitation payments from private and public purchasers. The HMOs agreed to provide specified services to enrolled groups, typically for a year, in return for these payments. The problem faced by the HMO was how to manage the providers (hospitals and physicians) in its network so that first, its costs did not exceed its capitated payments and second, the performance of the plan and, especially, the network providers satisfied purchasers, resulting in contract renewal from year to

year. Network providers were paid in a wide variety of ways, including elements of capitation, fee-for-service and salary as HMO managers sought to achieve their objectives. As HMOs grew in number and in the population they served, policy attention focused on how HMO payment methods influenced service utilisation. Of particular interest in early studies was whether HMOs delivered fewer services to their members than these same members would have received in traditional insurance plans. Often these comparisons were incorrectly characterised as 'capitated' versus 'fee-for-service'. There was a strong, if sometimes unstated, assumption that the reliance of traditional insurance plans on fee-for-service payment of providers had encouraged the use of unnecessary services, especially hospital care, and that reduction in provision of inpatient services could save costs without any impact on quality. Therefore, studies of this type often focused on documenting impacts on service utilisation, comparing HMO members with enrollees in traditional insurance arrangements. They paid relatively little attention to the actual financial arrangements between HMOs and contracted providers and, typically, did not attempt to measure the impact of differences in service utilisation on measures of quality or on patient outcomes. To be fair, early research of this type was limited in its focus by the relatively underdeveloped status of quality measurement in the health services research field, the costs of abstracting data on treatment processes and patient outcomes from paper medical records, and the unwillingness of many HMOs to allow researchers access to the data needed to construct measures of quality. However, even in this environment the variety of different types of HMOs – and the combinations of financial incentives and managerial techniques they used to affect utilisation decisions of their contracted providers – offered ample opportunity for simple but potentially misleading comparisons of capitation and fee-for-service, with policy-makers providing a receptive audience for the findings.

Also of note during the early 1980s was the adoption by the US Medicare programme (a government-funded and administered insurance programme for people 65 years and older) of a new payment arrangement for hospitals: prospective payment. Under this bundled payment arrangement, hospitals were reimbursed a fixed amount for each Medicare beneficiary inpatient stay, with the amount varying by type of diagnosis. Prospective payment replaced existing hospital payment arrangements based loosely on costs, and it created an incentive for hospitals to reduce length of stay for Medicare patients. (More precisely, the prospective payment system adopted by Medicare was in effect a 'blended payment' approach because hospitals received additional amounts for patients with stays exceeding predetermined guidelines.) Again, the assumption was that existing lengths of stay were longer than necessary and that reductions could occur with little or no threat to quality. Therefore, as with HMOs, initial research focused on determining the impact of the new payment system on utilisation, with very few attempts to measure quality impact directly.

UK context: Payment by Results

In October 2002 the Department of Health published Reforming NHS Financial Flows: Introducing payment by results, which described a new pricing and payment mechanism for health services which is similar to the US Medicare prospective payment approach. Prior to the introduction of Payment by Results (PbR), payments to hospitals were determined by locally negotiated block contracts. Under the block contract system there was considerable variation in prices for operations across the country and hospitals were often paid even when they under-performed in terms of activity by failing to carry out the number of operations required of them.

Introduced into the NHS in 2004 for selected procedures and organisations, PbR is a system which rewards providers for the type and volume of activity performed. The level of payment is dictated by a fixed national tariff which is a schedule of prices for Healthcare Resource Groups (HRGs). There are more than 1,000 HRG codes and they cover a wide range of clinical procedures, treatments and diagnoses. The tariff is applied to an entire spell of hospital care, from admission to discharge for individual patients, with case mix adjustments calculated locally within the commissioning process. In terms of incentives, if the costs to a hospital for treating patients are less than the national tariff price, the hospital can retain that surplus. Similarly, if a hospital is able to attract larger numbers of patients it

can increase its income. Hospitals that are relatively high-cost are provided with incentives to make cost savings in order to remain financially viable.

According to the OECD (Docteur and Oxley, 2003), tariff-based systems have several favourable characteristics:

- they allocate budgetary resources on the basis of output
- they can give purchasers some control over treatment intensity (for example, through price)
- they encourage hospitals to increase output where there is demand and capacity, and look for ways to reduce costs per sickness episode.

A study cited by the OECD found that tariff-based systems can save 10 per cent or more compared with other financing methods (Gerdtham et al, 1999; Gerdtham, Rehnberg and Tambour, 1999). However, they also can pose a significant risk to financial performance of the system, placing the risk of cost over-runs on the provider. This can lead to undesirable provider behaviour – such as preferential selection of patients, up-coding, discouraging collaboration and sharing best practice across providers; and compromising quality of care.

Payment by Results: early evidence

There are, as yet, no comprehensive evaluations of PbR. To date, studies have primarily focused on gathering qualitative evidence from stakeholders and small-scale uncontrolled observational studies. In brief:

1. A study conducted in South Yorkshire (Mannion and Street, 2005) found that following early implementation of PbR there was a rise in acute elective activity with a marked increase in consultant to consultant referrals. The authors suggest a number of possible explanations:
    - pent-up latent demand because of lower waiting times
    - providers adjusting admission and discharge decisions in response to tariff incentives
    - GPs inflating referral rates in the base year of practice-based commissioning.
2. A study conducted by the Audit Commission (2005) found that the early implementers of PbR were, on balance, positive about the change. The perceived benefits of PbR included:
    - a clearer framework for planning and management
    - strengthened planning, financial management arrangements, information systems and overall performance management
    - encouragement for commissioners to focus on demand management and improving clinical pathways
    - greater clarity of roles and responsibilities
    - a positive change in culture and accountability
    - improved understanding of the local health economy.

The study found that in its first year early implementers found PbR to be a more complex, time-consuming and challenging process than they had anticipated. It required considerable investment of time and resources – a sum calculated to be around £50m nationally. It also required clinical engagement, better planning and reporting arrangements, careful negotiation and close attention to detail. This complexity meant that time and energy was devoted to the system and its mechanics, and at the time of the study efforts had not yet broadened into the desired concentration on quality of care and performance improvement. PbR also exposed existing weaknesses in terms of underlying financial difficulties, inadequate financial management arrangements and problems with data quality.

Quantitative analysis of hospital activity data for 2004/05 showed little difference in activity growth or efficiency between early implementers (that is, foundation trusts) and other acute trusts. On the basis of these data, the report suggests that the reported growth in activity across the NHS had been driven by initiatives other than PbR (for example, national targets) and that expected efficiency gains were not immediately realised by early implementers.

3. A preliminary report from a national study of PbR undertaken by OHE (2006) outlined qualitative data drawn from interviews with senior managers involved in and affected by PbR. Overall the interviewees supported PbR as representing a sound principle – paying hospitals for what they do – but they had concerns about the specifics, including the way that prices are fixed in the national tariff:
    - The dominant impression from both provider trusts and PCTs is of NHS commissioners and providers of acute hospital care working in a largely co-operative way to resolve problems that arise under PbR.
    - Hospital activity was generally observed to be increasing, but interviewees were reluctant at such an early stage to attribute this to the effect of PbR. Interviewees highlighted two main reasons for their perception of a lack of impact attributable to PbR: lack of spare capacity and a desire among trusts, including foundation trusts, to co-operate with local commissioners rather than increase activity in an unplanned way.
    - Most hospital respondents said they did not misrepresent case-mix, giving reasons relating to NHS culture and mature co-operative relationships between trusts and PCTs. More pragmatic reasons given for not misrepresenting case-mix included shortage of coding staff and inadequacy of management systems to respond in that way. However, increases in the recording of activity are expected as a result of the incentive to code care more quickly and accurately.
    - There were few comments by interviewees about whether and how PbR might affect the efficiency of current activities. The general view is that existing pressures for efficiency are already great in an environment of tight budgets, growing demand and full capacity.
    - No interviewee considered that PbR was yet stimulating trusts to compete on quality.

Overall, the study found that respondents did not perceive that a more competitive environment was developing as a consequence of PbR, but some recognised the possibility of this happening in future.

Structure

In this appendix we review the evidence on the impact of financial incentives on utilisation and costs. Our focus is on studies that do not measure quality of care as a study outcome. However, even in these studies, authors occasionally speculate about the likely relationship between utilisation changes and

quality. The studies summarised in this appendix, which are set largely in the USA, provide a context for our review of literature in Chapters 2 and 3 that does incorporate explicit measures of quality. If financial incentives under different payment arrangements have had no significant impact on utilisation, which is certainly possible given the variety of other factors that can influence utilisation, then the likelihood that they will affect quality is questionable. On the other hand, if provider financial incentives are important factors in explaining utilisation, there is a basis for expecting that they may affect quality as well.

The remainder of this appendix is organised into six parts. In the next section we briefly summarise the overall evidence on the impact of provider financial incentives on healthcare utilisation and, typically, costs. We also discuss the findings from prior literature reviews. In the next four parts, we discuss the evidence disaggregated by studies that focus on:

- comparisons of insurance arrangements
- comparisons of incentive arrangements for institutional providers
- comparison of financial incentives pertaining to physicians and other healthcare personnel
- analyses of incentive changes in the health system level.

In a final section we draw conclusions about the strengths and limitations of this literature, as well as its implications for studies that focus explicitly on the impact of financial incentives on quality, as reviewed in the text.

Evidence summary

We reviewed 48 published articles that examined the relationship between the financial incentives of healthcare providers and utilisation and costs of medical care. We drew four general conclusions from that review.

- There is clear evidence that medical service utilisation differs under different provider payment arrangements. It is reasonable to conclude that provider behaviour in the patient–provider decision-making process can be influenced by financial incentives.
- The exact nature of the provider payment arrangement often was not discussed in the studies reviewed in this appendix. This was especially true of studies where comparisons of utilisation and costs under different insurance plans were made. The reader was left to infer the causal link between the general incentives in insurance arrangements and provider behaviour. Consequently, these studies are limited in their ability to inform policy-makers about the impact of specific financial incentives.
- There were very few studies that described how incentives in payment arrangements between payers and provider organisations were actually transmitted to practitioners. This was a particular problem in interpreting the findings of studies that compared different payment arrangements for hospitals. Although payments are directed at hospitals, the primary decision-maker with respect to utilisation of hospital services is the physician. The degree to which physicians were exposed in their decisions to financial incentives under different hospital payment approaches was not clear.
- In general, while the literature was wide-ranging and many of the study approaches were innovative, findings were reported at a very general level. Beyond supporting the idea that incentives can be used to influence provider behaviour, the literature provides little useful direction for decision-makers regarding the design of incentive systems that would reward quality of care.

Description of findings

We found six published literature reviews in our literature search, none of which could be classified as systematic reviews. McCleary, Asubonteng, and Munchus (1995) reviewed the early literature on the effect of financial incentives on physician behaviour in HMOs. They found that service utilisation was generally less for physicians in HMOs. They also observed that little research had addressed the impact of financial incentives on quality of care, and that more research was needed in this area. A literature review by Aas, also published in 1995, assessed changes in behaviour resulting from changes in inpatient payment policies for hospitals and physicians. They concluded that physicians paid by capitation had lower admission rates than physicians paid fee-for-service, that payment of hospitals on a per day basis resulted in longer lengths of stay (compared with other payment methods) and that salaried physicians used fewer diagnostic procedures than physicians paid fee-for-service. Like McCleary, Asubonteng and Munchus, Hellinger (1995) reviewed the impact of financial incentives on physician behaviour in managed care plans. He found that financial incentives were a key element in the success of managed care plans in reducing utilisation of services, but also identified several sources of bias in the studies reviewed. Flynn, Smith and Davis (2002) reviewed the literature regarding different approaches used to manage healthcare utilisation, describing the different types of financial incentives employed to influence provider behaviour. They found no conclusive evidence of behaviour impacts, with the exception that fee-for-service payment tended to be associated with the delivery of more primary care services by physicians. Mitchell (1996) reviewed the evidence on the relationship between physician ownership of facilities and their referral patterns, concluding that ownership was associated with increased utilisation of services. A sixth paper, authored by Rice (1997), while not based on a comprehensive literature review, did assess the impact of changes in financial incentives in the US Medicare programme in the context of a theoretical model. He concluded that the literature on fee-for-service payment provided evidence that physicians often respond to payment changes, sometimes increasing volume of services provided in response to payment reductions. None of the authors described a systematic approach for identifying the studies included in their reviews. Hellinger's review (1996) provided a rigorous critique of research designs and analytic methods. He observed that information about how plans actually reimbursed providers was typically missing in the articles reviewed, making the usefulness of these studies in drawing inferences concerning the impact of financial incentives on provider behaviour quite limited.

Insurance-level comparisons

The managed care era in the USA, and particularly the emergence of HMOs as a major health insurance option, spawned studies that compared outcomes in managed care plans to outcomes in more traditional plans (which typically were labelled fee-for-service plans). We reviewed five studies of this type. Badgett and Rabalais (1997) found that prescriptions per year, hospital admissions and hospital days per 1000 Medicaid beneficiaries were substantially less in the contracting HMO. Unlike many studies of this nature, the financial incentives actually facing physicians in the HMO were described and seemed relatively powerful in their structure. Primary care physicians received capitated payments that included specialty and emergency room care, and hospital budget surpluses were shared with physicians. In a second Medicaid study, Inkelas (2005) reported that the incentives in Medicaid contracts with managed care plans led to more frequent referral of children with special healthcare needs to specially-trained providers. An inference could be made that quality of care improved as a result, but direct measures of quality were not examined. In a comparison of use of services for atherosclerosis in private HMOs, Medicaid, Medicare and private insurance in California, Kuykendall, Johnson and Geraci (1995) found that privately insured patients received more services. They attributed this to differences in financial incentives under the different payment arrangements. Dhanani (2004) found that Medicare beneficiaries joining HMOs had fewer hospital days than beneficiaries remaining in fee-for-service Medicare, but most of the difference was attributable to favourable selection into HMOs. In a comparative study of a different type, Safran et al (2000) collected patient assessments of their primary care under five different models

of managed care. The nature of the incentive differences across these managed care plans was not described in detail. However, the authors concluded that paying physicians by capitation was negatively associated with most outcomes, especially the quality of the patient–physician relationship.

Badgett and Rabalais (1997) assessed the use of services by Medicaid enrollees in Kentucky who were in a capitated HMO in 1983/84 versus those in fee-for-service Medicaid. In the HMO, primary care physicians received capitated payments that included specialty care and ER care, with unspent funds returned to them. Surpluses in the hospital budget also were shared with the primary care physician. Utilisation data were collected for a 12-month period. The authors stated that, because they had complete data, no statistical testing was necessary. Physician visits were similar for the two groups, but there were large differences in prescriptions per year (1.9 versus 4.9), hospital admissions (0.11 versus 0.22), and hospital days per 1000 (461 versus 909). The authors speculated that utilisation differences could reflect prior differences in use, or confusion on the part of recipients during the first programme year. The authors acknowledged that the lack of outcome measures meant that the impact on recipients could not be detected.

Type of study: Controlled observational study

Dhanani *et al* (2004) conducted an empirical analysis of inpatient utilisation by Medicare beneficiaries before and after switching from fee-for-service to HMOs. It was presumed, based on past research, that HMOs create incentives for providers to reduce hospitalisation, although no specific incentive at the provider level is described. Total inpatient days per thousand Medicare beneficiaries per year constituted the outcome variable. The authors compared before-and-after utilisation by beneficiaries joining HMOs from 1991 to 1995 in California with a comparison group that did not join HMOs and with a second comparison group whose members were in HMOs for the entire time period. Data from Medicare and from the State of California databases were used in the analysis. A two-part model was estimated: the probability of using hospital care, along with the number of hospital days for those with some utilisation. For beneficiaries joining a group or staff model HMO, hospital days were 18 per cent lower than if they had remained in fee-for-service. Reductions in days for beneficiaries in IPA model HMOs were 11 per cent. In group/staff model HMOs favourable selection explained 70 per cent of the observed difference. Selection explained 85 per cent of the difference in IPA model HMOs.

Type of study: Quasi-experimental study

Inkelas (2005) analysed the impact of a Medicaid carve-out programme to determine the impact for Title V special needs children on the number of children classified as Title V as well as the expenditures on them. The different start-up dates for the programme created a quasi-experimental design. The California Medicaid programme provided data for the study. Multiple regression was used with correction for heteroskedasticity. The number of children receiving Title V services increased and expenditures for some ambulatory services increased, but not overall expenditures. There were quality of care implications in the findings. It appears that Title V standards were applied to the care of more children. The likelihood that they would receive care from Title V providers increased as well as early initiation of care.

Type of study: Quasi-experimental study

Kuykendall, Johnson and Geraci (1996) compared the provision of services for coronary atherosclerosis across different types of payers to examine associations between payer and provision of services. The data consisted of 24,424 discharges from California acute care hospitals during 1989. Multiple regression analysis was used to analyse the data. Privately insured patients were more likely to receive most services than patients covered by other payers. The authors attributed their findings to differences in financial incentives among payers. The authors suggested that further studies were needed to help interpret their findings and observed that recent percutaneous transluminal coronary angioplasty (PTCA) clinical trials could change use of this modality.

Type of study: Controlled observational study

Safran *et al* (2000) assessed patient evaluations of their primary care in different types of health plans where primary care physicians operated under different financial incentives. The authors compared primary care received by patients in five models of managed care and identified characteristics of health plans associated with performance differences. A telephone survey was conducted of 6018 Massachusetts state employees in 1996 who reported a personal physician and for whom plan type was known. The response rate was 68.5 per cent. Health plans provided information about the financial and non-financial features of their contracts with physicians. The managed indemnity system performed the best across the measures, and staff model plans performed the worst. In IPA/network plans, paying physicians capitation was negatively associated with most outcomes and the relationship was significant for quality of patient–physician relationship. When IPA physicians were paid based in part on patient satisfaction there was a positive relationship with the measures.

Type of study: Controlled observational study

Institutional-level comparisons

Our search efforts identified 24 studies that examined the impact of payment incentives for hospitals or other institutional providers on utilisation or costs, 20 of which were set in the USA. Thirteen of the US studies were related in some form to hospital responses to the introduction of a prospective payment methodology, or a change in payments. However, five of these studies utilised the same database, drawn from a single hospital. All five focused on documenting patient heterogeneity within payment category, suggesting that hospitals have incentives to avoid sicker patients in a category because they require more resources to treat (Munoz et al, 1988a–e). The remaining eight U S studies covered a range of issues. Ellis and McGuire (1996) found a reduction in the length of stay for psychiatric patients, as expected, when prospective payment was introduced in a Medicaid programme. Meltzer and Chung (2001) determined that competition among hospitals in California had a different effect on costs before and after the introduction of prospective payment by Medicare, with competition associated with lower costs after introduction of prospective payment. Norton et al (2002) reported an increase in length of stay for psychiatric admissions when payment in a Medicaid programme was changed from per day to per admission, which was attributed to an increase in payment that accompanied this change. Lave and Frank (1990) found that per case payments and negotiated payments for hospitals resulted in shorter lengths of stay, and that the hospital response was stronger for psychiatric than for medical or surgical admissions. Wolfe and Detmer (1988) examined charges and length of stay for patients with inguinal hernias and acute appendicitis, finding that a change to prospective payment per case resulted in lower costs. Yip and Eggleston (2001) found that when a province in China changed from a fee-for-service to a prospective payment approach for hospital care, the rate of expenditure growth for inpatient services slowed.

In addition to studies that addressed changes in length of stay or costs resulting from payment change, four other hospital-based studies focused on different outcomes. Chan et al (1997) examined whether rehabilitation hospitals increased charges in a base year in order to obtain bonus payments for reducing charges in subsequent years. He concluded that they did, resulting in an overall increase in Medicare

costs and number of hospital days for patients in these hospitals. Nicholson and Song (2001) studied how hospitals responded to the financial incentives of the Medicare programme with respect to graduate medical education. They found that hospitals, as expected, hired more residents and closed beds, but did not increase Medicare admissions or change nurse staffing. Schoenman (2005) examined hospital responses to Medicare payment incentives that theoretically could change discharge policies, but found no effect. Lindrooth and Weisbrod (2007) concluded that not-for-profit hospitals with relatively large shares of Medicare patients were more likely to reduce treatment intensity when confronted with a payment reduction by Medicare. Finally, in a recent study set in Taiwan, Xirasagar and Lin (2006) examined whether the incentives for not-for-profit and for-profit hospitals differed from those for public hospitals such that they would favour outpatient treatment for hernia and cataract surgery patients. The study found higher rates of inpatient treatment in large teaching hospitals and, to a lesser degree, in public hospitals.

Intrator and Mor (2004) analysed the relationship between how much nursing homes are reimbursed and the likelihood that their residents will be hospitalised, concluding that a 10 per cent increase in Medicaid reimbursement above the mean reduced the risk of hospitalisation by 9 per cent.

Three other studies addressed the impact of payment changes on utilisation in long-term care facilities. White (2003) assessed the response of skilled nursing facilities to the transition to prospective payment, examining the impact on provision of rehabilitation therapy in these facilities. There was a large drop in rehabilitation charges, as well as changes in service use. Zinn et al (2003) also addressed provision of rehabilitation therapy after a change to prospective payment, concluding that facilities either quit providing these services or exercised greater control over service provision. The latter was especially the case in facilities that were part of chains or were for-profit in nature. Relating to home healthcare, Murkofsky et al (2003) found that a reduction in payment was associated with a decrease in the length of time patients received home health visits. Grabowski et al (2006) reported that restrictions on payments for home healthcare by Medicare were associated with a decrease in care for the sickest patients, less agency assistance with activities of daily living and a shorter length of service use, but without impacts on patient outcomes. Lindrooth et al (2007) found that not-for-profit hospices were less likely than for-profit hospices to respond to Medicare payments that rewarded admitting patients with shorter, less profitable lengths of stay.

Two relatively complicated studies recently were conducted that related to hospital response to payment change in Finland and in Taiwan. In Finland, there was a change in hospital pricing from a per diem model to case-based or fee-for-service models. Mikkola (2003) followed the impact of the change on three discretionary surgical procedures. The expectation was that volume of procedures would increase, and it did for discectomies and knee replacements. Importantly, Mikkola (2003) used information about the context of the change, especially limits on the number of available surgeons, in interpreting the findings. Huang et al (2005) identified strategies that hospitals used in responding to implementation of a case-based payment system, one of which was financial incentives imposed on medical staff members. These were found to affect overall hospital revenues.

Chan *et al* (1997) analysed possible gaming of Medicare payments by rehabilitation hospitals to increase base year payments in order to receive bonus payments for reducing charges in subsequent years to determine if rehabilitation hospitals responded to payment incentives by increasing charges in a base year. Medicare data from more than 190,000 discharges from 69 rehabilitation hospitals from 1987 to 1994 were analysed. Multivariate analysis was used that controlled for patient characteristics, temporal trends, inflation and hospital characteristics. The new payment method increased Medicare costs and numbers of hospital days for the average patients. There was a 28 per cent increase in payments in the base year. There is no information provided regarding quality.

Type of study: Controlled observational study

Ellis and McGuire (1996) conducted an econometric analysis of the impact of implementation of DRG payments for psychiatric admissions to determine the nature of the impact of instituting DRG payments for psychiatric admissions. The major data source was New Hampshire Medicaid claims data from 1987 to 1992. This encompassed a period of 1.5 years prior to implementation of the change. Results were compared with general trends in the commercial sector. The authors found a 14 per cent reduction in length of stay. There also was some indication of quality-related competition for patients likely to be profitable to the hospital.

Type of study: Quasi-experimental

Grabowski *et al* (2006) assessed the impact of the institution of a new payment system in Medicare on home healthcare service provision. Medicare implemented its interim payment system (IPS) in October 1997, which created a natural experiment. The IPS reduced the per visit payment maximum and introduced a per beneficiary payment cap. Probit models were used to estimate difference-in-difference equations for a variety of outcome measures. Data were taken from a survey that was administrated in 1994, 1996, 1998 and 2000. The IPS was associated with a decrease in care for the sickest patients, less agency assistance with activities of daily living and a shorter length of service use. There were no differences in patient outcomes. The authors noted that the payment arrangements had some negative impact on non-Medicare patients as well as Medicare patients.

Type of study: Quasi-experimental design

Huang *et al* (2005) identified 'coping strategies' used by hospitals in response to implementation of a case-based payment system and then assessed the impact of these strategies on medical revenues and patient satisfaction. Data were collected from a survey of Taiwan hospitals in October 2000. The survey generated a 69 per cent response rate. Factor analysis was used to identify clusters of coping strategies, with one of the clusters consisting of a set of financial incentives. These factor scores, together with control variables, were then used in a stepwise linear regression. The financial factor was related to medical revenues but not to patient satisfaction.

Type of study: Observational study without control group

Intrator and Mor (2004) tried to establish empirically the relationship between the amount that nursing homes are reimbursed and the likelihood that nursing home residents will be hospitalised to determine the factors that affect hospitalisation of nursing home residents. A sample of non-governmental nursing homes with 25 or more beds in one community in each of ten states was drawn in 1993, resulting in 253 nursing homes. Then, eight to 16 residents were randomly selected from each facility, totalling 2080. Nurses completed a baseline assessment for these residents. They were tracked for six months. A multinomial logit analysis was performed, with the purpose of seeing which factors predicted hospitalisation during the six-month period. A $10 increase in 1993 Medicaid reimbursement above the mean rate of $75 resulted in a 9 per cent reduction in the risk of hospitalisation. The authors argued that the additional funding was used to increase staff and quality, thereby reducing the need for hospitalisation.

Type of study: Observational study without control group

Lave and Frank (1990) examined the effect of different payment methods on length of stay for Medicaid patients. Per case, negotiated contracts and prospective per diem payments were examined. Data were from 1670 hospitals in 1984 participating in the Professional Activity Study, supplemented

by data from a variety of other sources. Ordinary least squares regression was used for two groups of hospitals and three classes of diagnoses. Per case payment systems and negotiated contracts were associated with decreased lengths of stay for all groups. Prospective per diem with limits in most cases led to decreases in length of stay. The response was stronger for psychiatric than for medical and surgical patients. The authors stated that they could not draw conclusions about the impact of their findings on patients.

Type of study: Controlled observational study

Lindrooth *et al* (2007) examined how hospital treatment intensity was affected by an exogenous change in average reimbursement for an admission. Implementation of the Balanced Budget Act of 1997 lowered hospital reimbursement for Medicare inpatient days by slowing the update factor for the standard dollar payment rate used in Medicare's prospective payment system. The primary dataset used in the study was the Healthcare Cost and Utilization Project State Inpatient Database for 1996 and 2003, covering 11 states. The analysis was limited to non-federal urban, short-term general hospitals and 16 disease categories encompassing 50 DRGs. Quantile regression was used to estimate a difference-in-differences model with the expectation that hospitals with high Medicare shares would be more likely to respond to the payment change. The authors found that not-for-profit hospitals cut treatment intensity at the 50th, 75th and 95th quantiles only for generously reimbursed services. There was no statistically significant response for public or for-profit hospitals. This may have been because for-profit hospitals already were providing services at a low treatment intensity.

Type of study: Quasi-experimental design

Lindrooth and Weisbrod (2007) estimated the differential response of religious not-for-profit hospices and for-profit hospices to an exogenous Medicare reimbursement incentive that encouraged hospices to maximise length of stay. Given that the cost function for hospice care is non-linear, with greater expenses in the initial days, and payment is on a per diem basis that generates profits on days of care provided beyond the initial period, hospices have an incentive to select patients with a long expected length of stay. For-profit hospices may be more likely to exploit this opportunity. The authors used Medicare data on all admissions to hospices reimbursed by Medicare in 1993, with patients followed until 1996, combined with date-of-death data from other sources. Two-stage regression techniques were used to estimate relationships. For-profit hospices were significantly less likely to admit patients with shorter, less profitable expected lengths of stay, but ownership status did not influence the timing of admissions.

Type of study: Observational

Meltzer and Chung (2001) analysed the impact of competition on high- and low-cost admissions to examine the effects before and after the establishment of the Medicare PPS. Data from California state government were used to construct cost measures. Census data also were used. Herfindahl indices at the county level were used as measures of competition. Analysis was limited to the 12 highest volume DRGs. Quantile regression analysis was employed to examine relationships. Competition was associated with higher costs for both age groups before DRGs and lower costs after DRGs were implemented. The authors concluded that their analysis raised questions about the impact of competition and DRGs on patient outcomes, especially for the sickest patients, and that future research needed to address these questions.

Type of study: Controlled observational study

Mikkola (2003) conducted an empirical analysis of the response of hospitals in Finland to a change in pricing from a bed day to case-based and to fee-for-service prices. The expectation was that volume would increase due to the change in payment methods. Data on procedures was taken from the Finnish Healthcare Register. Data on other variables were taken from the Database for Social and Health Statistics in Finland. Panel models were used to estimate effects. The basic model was a two-way fixed effects model. Other approaches used were OLS, one-way fixed effects and random effects. The rates of use for discectomies increased by 8 per cent under case-based prices and decreased by 11 per cent for hip replacements. Significant increases were observed only for knee replacements (21 per cent) under mixed bed day and fee-for-service prices. The size of the impacts was less than in Sweden and in Norway when these countries switched to new payment approaches. The author suggested this may be because of limits on the number of surgeons and thus the capacity to respond to reimbursement changes. The increased use of discectomies could be an indicator of poorer quality care, although the author did not speculate on this, instead commenting only on the fact that discectomies were of questionable cost-effectiveness.

Type of study: Controlled observational study

Munoz *et al* (1988d) conducted a study that focused on hospital resource consumption for patients with 17 urology non-complicating condition DRGs to examine how costs and service use varied across urology DRGs with varying numbers of non-complicating conditions. Data pertained to 185 Medicare urology patients at the Long Island Jewish Medical Center in 1985 and 1986. The data were analysed through comparison of means with tests for statistical significance. Patients were placed in four categories depending on number of non-complicating conditions. More conditions per patients correlated positively with all outcomes, including mortality. The authors made no quality inferences. This was a very limited study. There were relatively few data points, the analysis was unsophisticated and the findings were not relevant for assessing impact of incentives on quality.

Type of study: Observational study without control group

Munoz *et al* (1988e) conducted an empirical analysis of differences in resource use among nephrology patients in a single hospital. The authors used hospital records to examine resource use for 784 patients admitted to Long Island Jewish Medical Center in 1985 and 1986. Statistical tests of differences in group means were conducted. Hospital costs and mortality increased with age. Patients with diabetes and with other complications had higher resource consumption. The authors concluded that the current DRG scheme could contain financial incentives for poorer access and care.

Type of study: Observational study without control group

Munoz *et al* (1988a) conducted an empirical analysis of resource use in 147 surgical DRGs. The authors analysed data from 2,647 Medicare surgical patients admitted to Long Island Jewish Medical Center in 1985 and 1986. ANOVA was used to test for differences between groups of patients, defined by number of complications. Patients with greater numbers of complications had higher costs, longer lengths of stay, more procedures and higher mortality rates. The authors suggested that DRG payments, because they didn't adjust for number of complications, may have had financial incentives for hospitals to avoid sicker patients or deliver lower quality, less expensive care.

Type of study: Observational study without control group

Munoz *et al* (1988b) examined hospital resource consumption for pulmonary admissions, focusing on varying use with number of complications, to examine differences in hospital resource use and mortality for pulmonary patients. The authors analysed 427 patients admitted to Long Island Jewish

Medical Center in 1985 and 1986. Statistical tests were conducted for differences in means. Resource consumption and mortality were higher for patients with more complications. The authors speculated that heterogeneity within DRG categories could jeopardise access and quality of care.

Type of study: Observational study without control group

Munoz *et al* (1988c) examined the impact on hospital resource consumption of complications in neurosurgical DRGs. The authors analysed the effect of complications on resource consumption in nine neurosurgical DRGs. The authors used data on 148 Medicare patients admitted to the Long Island Jewish Medical Center in 1985 and 1986. Analysis of variance was used to test for differences among four groups defined by number of complications. Patients with more complications had higher resource consumption and poorer outcomes. The authors concluded that this suggested the DRG reimbursement scheme was inequitable and hospitals might have disincentives to care for these patients.

Type of study: Observational study without control group

Murkofsky (2003) conducted an empirical analysis of the length of stay in home healthcare before and after a reduction in Medicare payments. Data from the 1996 and 1998 National Home and Hospice Care Surveys, which included 1053 home health agencies in 1996 and 1088 in 1998, were used in the analysis. There were 4127 patients in 1996 and 4051 in 1998. A Cox proportional hazards model was used to analyse the data. There was an unadjusted decrease of 16 days, from 60 to 44 days. The effect of the change in reimbursement was significant in multivariate analysis. The authors observed that little is known about quality of care before and after the change.

Type of study: Controlled observational study

Nicholson and Song (2001) conducted an empirical analysis of the responses of hospitals to changes in the Medicare indirect medical payment policy to determine how hospitals have responded to the financial incentives associated with graduate medical education. The authors employed a panel dataset of 3900 hospitals, including 900 teaching hospitals, with data from 1984, 1986 and 1991. The dataset was constructed using Medicare data and American Hospital Association data. They estimated a difference-in-difference model (taking advantage of the fact that hospitals in Maryland and New Jersey were not eligible for payments between 1983 and 1991) and hospital-specific difference regressions. They found that teaching hospitals did hire more residents and close beds in response to Medicare payment policies but did not increase Medicare admissions or alter their use of nurses. The resident effect was relatively small. Other literature suggested that there was a positive relationship between quality of care and the registered nurse to patient ratio in hospitals. Implicitly, the finding of no impact on use of nurses combined with the small impact on use of residents could suggest that the graduate medical education policies had limited impact on quality of care.

Type of study: Quasi-experimental study

Norton *et al* (2002) conducted an empirical analysis of the impact on length of stay of changing the reimbursement model for inpatient care for severely mentally ill patients from per diem to per episode. Data were taken from a sample of 8509 severely mentally ill patients in the Massachusetts Medicaid programme covering the period from 1991 to 1993. The change to a per episode basis occurred in late 1991, with a change back to a per diem arrangement about a year later. Regression techniques were used, with hospital fixed effects (random effects models were estimated as well). A sub-analysis

was run for a subset of patients with at least one admission in each period. The authors found that the marginal price elasticity was zero and that the average price elasticity was between 0.16 and 0.20. Thus, on average, length of stay increased overall as payment increased. A 50 per cent increase in average price led to an 8 per cent increase in length of stay.

Type of study: Quasi-experimental study

Schoenman (2005) analysed the effect of a change in Medicare reimbursement policy on hospitals. Medicare data from 1998 to 2001 were used for the analysis. Regression analysis was employed. There were no significant changes in hospital behaviours.

Type of study: Controlled observational study

White (2003) analysed the response of skilled nursing facilities (SNFs) to the introduction of prospective payment to determine if SNFs responded to prospective payment by reducing their costs through provision of less rehabilitation therapy. SNFs made the transition to a new payment system in 1998. Claims data for 1997 to 2000 were used in the analysis. It was not clear if statistical methods were used to analyse the data versus simple tabulations. Residents were less likely to receive substantial rehabilitation, but also less likely to receive no rehabilitation. The author noted concerns about whether some patients were getting the intensive therapy they needed.

Type of study: Controlled observational study

Wolfe and Detmer (1988) analysed charges and length of stay for two types of patients – inguinal hernia and acute appendicitis – in a university hospital. The purposes of the analysis were to ask whether there were differences in medical care – in terms of charges and length of stay – among identical surgical patients with identical positive outcomes and whether DRGs controlled sufficiently for severity differences to make the categories meaningful in terms of incentives, yet avoid pressure to select against severely ill patients. Data were from operations in 1981 and 1982, and were gathered through a computer search of a university hospital's medical records. There were 183 observations in total. The authors concluded that DRGs could lead to a reduction in costs without a reduction in patient outcomes. The quality conclusion was reached based on inferences from constructing physician practice profiles. The results were based on a very limited number of patients in a setting that is probably not possible to generalise.

Type of study: Observational study without control group

Xirasagar and Lin (2006) conducted a study in Taiwan which explored whether for-profit and not-for-profit hospitals were less likely than public hospitals to accept patients under prospective payment, favouring outpatient over inpatient care for surgical patients. This was a retrospective, population-based, cross-sectional study using data from Taiwan's national health insurance database. Diagnosis-wise logistic regression analysis was used. Large teaching hospitals were more likely to admit patients than for-profit district hospitals. Public hospitals also were more likely to admit, but the differences were not as great. The differences were less in areas of greater competition. The authors noted that they could not address whether the lower admissions represented a more efficient use of resources or poorer quality of care.

Type of study: Controlled observational study

Yip and Eggleston (2001) assessed the impact of payment reform for hospitals in Hainan Province, China. The authors used a difference-in-difference approach and regression analysis to assess the impact of the change. The study covered two years (1995–1997) and data were provided by the Department of Health Insurance, Hainan Social Protection Bureau. The change to a new payment approach slowed expenditure growth for inpatient services (average expenditure per admission) and also co-payments per inpatient admission. While not addressed in the study design, the author noted that there was some anecdotal evidence of under-provision including quality problems at some hospitals. Some hospitals were fined as a result.

Type of study: Quasi-experimental study

Zinn *et al* (2003) Transaction cost theory was used to structure an empirical analysis of skilled nursing facility decisions relating to provision of rehabilitation therapy. The authors used Online Survey Certification and Reporting data from 1992 to 2001 on a sample of 10,241 freestanding urban SNFs. They estimated a longitudinal multinomial logit regression model. Following the institution of prospective payment by Medicare, SNFs acted to limit their transaction costs by either exiting the market or exerting greater control by managing services in-house. For-profit status and chain affiliation were associated with greater control. The authors expressed concern that bringing therapy in-house could result in lower staffing levels, given the incentives under prospective payment, and therefore poorer quality of care. They suggested that monitoring of quality effects is needed.

Type of study: Controlled observational study

Physician-level comparisons

We identified 23 studies that examined the impact of physician payment incentives on physician behaviour as evidenced in utilisation and cost measures. Eleven of these studies were set in the USA, four in Canada, four in the U K, two in Norway, one in Ireland and one in China. Of the studies set in the USA, half investigated the impact of financial incentives on care provided to women. Burns, Geller and Wholey (1995) found no relationship between the performance of Caesarean sections and the percentage of a physician's patients enrolled in traditional insurance plans versus HMOs. Hadley et al (1999) examined the effect of being an HMO enrollee on treatment choice for breast cancer and found that HMO enrollees were less likely to receive 'breast conserving' therapy and, for HMO patients receiving a mastectomy, lengths of stay were shorter in comparison to other patients. In a later (2003) study, Hadley et al found that an increase in fees increased the probability of receiving breast conserving therapy, as did a reduction in fees received for performing the alternative treatment of mastectomy. The authors speculated that the relatively high degree of responsiveness to fees in their study may be related to the fact that the procedures were expected to have similar outcomes. Ransom et al (1996) reported a 15 per cent overall decrease in the number of gynaecologic surgical procedures done under capitation versus fee-for-service payment, and Gruber and Owings (1994) found that the likelihood that physicians would perform Caesarean sections increased as their overall income levels were threatened.

Of the remaining six studies set in the USA, two by Conrad and colleagues are particularly noteworthy. These studies examined how physician production levels in medical group practices responded to financial incentives. In their first study (1998) they found that compensation methods did not affect costs and utilisation. In a subsequent analysis, published in 2002, they found that group level incentives did not affect productivity but individual incentives were related to higher productivity levels. Hillman and colleagues (1989) studied the impact of the use of different physician payment arrangements by HMOs on utilisation. Their findings indicated that hospital utilisation rates were lower under capitation or salaried arrangements. Other incentive arrangements placing physicians at risk reduced the number of physician visits. Davidson et al (1992) reported on the results of an experiment in which physicians were randomly assigned to a group that received higher than previous fees for managing patient utilisation

versus a group in which physicians held some degree of financial risk. They found that, in the second group, there was a lower rate of referral of children in the Medicaid programme to specialists. Tai-Seale, Rice and Stearns (1998) examined the response of physicians serving Medicare patients to fee reductions, with inconclusive results. O'Malley et al (2006) found that financial incentives provided by a health plan to its network physicians did not increase physician prescribing of generic drugs.

Of the four studies set in Canada, Hutchison et al (1996) found that paying physicians an additional incentive for lowering hospital utilisation was not effective. However, the number of physicians in the study was small and there were a limited number of control variables employed. Nassiri and Rochaix (2006) reported that physicians changed the type of reported consultations performed in response to changes in relative prices, while Stanton and Shortt (2003) concluded that a change from fee-for-service to block funding for clinical faculty in a medical centre did not alter practice patterns. Grootendorst and Stewart (2005) found that a reference pricing scheme for drugs implemented in British Columbia, in which reimbursement was set equal to low-priced alternatives to encourage physicians to prescribe generics, resulted in savings.

In the UK Pockney et al (2004) evaluated a programme in which GPs were given financial incentives to increase the number of procedures they performed in order to decrease the number of referrals of patients to specialists. The number of procedures did increase; this was driven in large part by greater use of cryotherapy for wart removal. Kristianson and Holtedahl (1993) reported that physicians who were paid fee-for-service (rather than salary) provided more home visits. And Dusheiko et al (2003) found that when GPs changed from holding their own budget to becoming members of trusts, elective hospital admission rates fell. Bradlow and Coulter (1993) concluded that prescribing expenditures were reduced when GP fundholders could retain part of the savings from their prescribing budget. A study set in Ireland also analysed the prescribing behaviour of GPs under different financial incentives. These incentives were found to have little impact on GPs who had high initial spending on prescriptions.

The two Norway studies focused on different topics. Bjorndal, Arntzen and Johansen (1994) reported that salaried GPs spent more time in consultations outside of their offices, while fee-for-service GPs had more consultations per hour of patient contact. However, the general conclusion was that financial incentives had little impact. The results of this study were based on a relatively small number of physicians. Iversen (1998) studied salaried physicians' responses to a change in the way in which their hospitals were reimbursed. He found that these salaried physicians did respond to the reimbursement change in their choice of inpatient versus outpatient therapy for patients with the same diagnosis.

Liu and Mills (2003) examined the impact on hospital revenues of bonus payments of different strength to hospital-based physicians in China. They found relationships between bonus schemes and revenues using longitudinal and cross-sectional data.

Bjorndal, Arntzen and Johansen (1994) compared the characteristics of physician practices in Norway under different financial arrangements to determine if different contractual arrangements for GPs affect their practise. Data were collected in March 1991 from all GPs in Oslo, with 83 per cent participating. The methodology employed was a statistical comparison of means. Salaried GPs spent more time in consultation-based work outside of their offices. Fee-for-service physicians had more consultations per hour of patient contact, while salaried GPs had more telephone consultations. The conclusion was that financial incentives had little impact, if any, on physician behaviour.

Type of study: Controlled observational study

Bradlow and Coulter (1993) conducted an empirical analysis of the impact of GP fundholding in the Oxford region of England on prescribing behaviour of physicians. Prescribing and cost information were analysed for two six-month periods in 1991 and 1992. Included in the study were three dispensing fundholding practices, five non-dispensing fundholding practices and seven non-dispensing, non-fundholding practices. The analytic technique was a comparison of means. Five of the eight fundholding practices made savings in their fundholding budgets. The authors concluded that costs were less in fundholding practices. The authors also noted that these practices likely had some flexibility that allowed them to reduce costs without compromising care.

Type of study: Quasi-experimental study

Burns, Geller and Wholey (1995) conducted an empirical analysis using cross-sectional data from one state and one year to examine the importance of non-clinical factors (physician characteristics) in explaining observed variation in Caesarean section rates. Physician convenience measures were more important than other physician characteristics, such as training, in explaining variation in Caesarean section rates. The log odds of performing a Caesarean section increased with the physician's rate of Caesarean sections in the prior year, delivery on a Friday and delivery between 6am and 6pm. If patients had traditional insurance they were no more likely to have Caesarean sections than if they were covered by HMOs or the state's Medicaid programme. Physicians with heavier caseloads of patients with commercial insurance were no more likely to perform Caesarean sections. However, women without insurance were significantly less likely to have Caesarean sections.

Type of study: Observational study without control group

Davidson *et al* (1992) evaluated two ways of compensating physicians to serve children under a demonstration programme in a county of New York. Utilisation on the part of the children in the two intervention groups was compared with traditional Medicaid. Regression analysis was used. Capitation did not affect primary care visits but it did reduce referrals to specialists. The authors said that there could be concern about under-service, but they did not attempt to address quality in the paper. The results were limited in that children were not randomised, the demonstration was in one area only and the follow-up period was brief.

Type of study: Randomised controlled trial

Dusheiko *et al* (2003) used difference-in-difference modelling, with data before and after a change in GP reimbursement, to see if a move away from fundholding by GPs had an effect on the rate of elective admissions for patients. A four-year panel of physicians was constructed, which included data from two years before the switch over from fundholding to two years after. A comparison group of GPs who were not fundholders was used in the analysis. The authors found that GPs switching out of fundholding reduced elective admission rates by 3.3 per cent, with 57 per cent of that reduction due to the changeover (with the remainder due to selection.).

Type of study: Quasi-experimental study

Grootendorst and Stewart (2005) analysed the impact of introducing reference pricing on expenditures for ACE inhibitors and Calcium Channel Blockers in a public programme in British Columbia, Canada. Data were from a public plan with reference pricing and a public plan (Ontario) that did not introduce reference pricing. Sophisticated multivariate statistical methods were used to estimate the impact of reference pricing. The authors reported significant savings attributable to reference pricing, but the magnitude of those savings was much less than had been previously estimated using a pre–post research design.

Type of study: Quasi-experimental

Gruber and Owings (1994) analysed the response of obstetrics and gynaecology physicians to a drop in demand, with the response being a shift from vaginal deliveries to Caesarean sections which have higher reimbursement rates. Data for the study were taken from the National Hospital Discharge Survey from 1970 to 1982, with the final sample consisting of 250,000 live births. Fertility changes during this period varied across states, providing variation in the independent variable. Fertility changes accounted for a fall in average income of about 6.75 per cent. Regression analysis was used to measure the effect of the income reduction on use of Caesarean sections. The authors found a strong correlation between declines in fertility rates (and, implicitly, income declines) and the use of Caesarean sections.

Type of study: Observational study without a control group

Hadley *et al* (1997) conducted a study of the impact of enrolment in an HMO on treatment choice for breast cancer among non-elderly patients. Hospital discharge data were collected in five states for two years: 1988 and 1991. Logistic regression and ordinary least squares regression were used to estimate the relationships. HMO enrollees were less likely to receive breast conserving therapy (odds – 0.93). Lengths of stay were shorter for HMO patients receiving a mastectomy. The authors stated that research on quality differences was needed.

Type of study: Controlled observational study

Hadley *et al* (2003) analysed the impact of variation in Medicare fees on the type of treatment received for breast cancer. Medicare claims and physician survey data for a national sample of older (67 years and older) beneficiaries with localised breast cancer and treated in 1994 were used in the study. Multiple logistic regression analysis was employed to estimate the treatment choice model. The principal finding was that a 10 per cent increase in the fee for breast conserving therapy increased the odds of receiving this therapy to 1.34, while a 10 per cent decrease in the fee increased the odds to 1.86. The authors concluded that physicians were responsive to financial incentives when there were alternative therapies with similar clinical outcomes. These results confirm the results of an earlier study done at the small area level. The authors suggested that fee responsiveness may be greatest for procedures that have similar expected outcomes.

Type of study: Observational study without control group

Hillman *et al* (1989) conducted an empirical analysis at HMO level relating physician payment methods to utilisation and finances to determine if HMO use of financial incentives affected physician behaviour towards individual patients. Data were collected from a survey (57 percent response rate) of HMOs. Stepwise regression and stepwise logistic regression were used to analyse the data. Use of capitation or salaries was associated with a lower rate of hospitalisation. Placing physicians at financial risk as individuals and imposing penalties for deficits in the HMO's hospital fund beyond the loss of withhold funds was associated with fewer outpatient visits. More HMO patients in a physician's caseload were associated with more frequent visits. HMOs were more likely to break even if physicians were at risk for outpatient tests. The authors did not assess how these differences were related to quality of care.

Type of study: Observational study without control group

Hutchison (1996) conducted an assessment of whether payment of primary care physician organisations using capitation reduces hospital utilisation in Canada. The authors tracked 39 physicians who converted to capitated payments and a matched sample of 77 physicians who remained in fee-for-service. Statistical tests of means were conducted. There was no statistically

significant difference between the two groups of physicians. The author concluded that the payments did not reduce hospital use. The small number of physicians and control variables limit the usefulness of the findings.

Type of study: Quasi-experimental study

Iversen (1998) conducted an empirical examination of how financial incentives affected the decision-making of salaried physicians in the public sector in Norway. The study explored whether salaried physicians in a hospital setting responded to a change in financial incentives in the payment for outpatient hospital care. Thirty-four hospitals (half of the hospitals in Norway) were asked to provide data; only 15 did so and the data were usable for only 12 of these hospitals. The author acknowledged that the hospitals in the study may not have been representative. Weighted least squares was used to estimate relationships. Physicians responded to incentives in the choice of inpatient versus outpatient therapy for patients with the same diagnosis but not for patients with different diagnoses. The incentive structure facing physicians was not clear. The author observed that it was not obvious why physicians should care about relative fees in these circumstances.

Type of study: Controlled observational study

Kristiansen and Holtedahl (1993) analysed the effect of financial incentives on a GP's choice of surgical consultation versus home visits. Data were gathered by recording patient contacts for one week for 116 GPs in rural areas of northern Norway during 1982. The data were analysed using multivariate analysis, but the exact method was not stated. Fee-for-service physicians used home visits more often, but the result was statistically significant only for home visits. The authors suggest that home visits may be associated with better care but did not include any measures of clinical care or patient outcomes in their study.

Type of study: Controlled observational study

Liu and Mills (2004) assessed how payment to hospital-based doctors affects hospital performance. The purpose of the assessment was to determine influence of bonus payments to physicians on hospital revenues. The study took place in Shandong Province, China. Survey data were collected for 108 hospitals (85 per cent response rate). The data covered the period from 1975 to 1997. Cross-sectional survey data from hospital administrators also were collected for 1997. A multivariate regression analysis was used to estimate the relationship between bonus payment and hospital revenues in 1997. When bonus arrangements switched from weaker to stronger incentive schemes there was a sudden increase in the rate of growth in hospital revenues. Bonus type was associated with size of hospital revenues in the cross-sectional analysis.

Type of study: Observational study without control group

Nassiri and Rochaix (2006) conducted an empirical study of a natural experiment that changed the financial incentives for primary care physicians in Quebec, Canada. The purpose of the study was to determine if primary care physicians responded strategically to financial incentives and, if so, the nature of that response. A panel specification with random individual effects and fixed temporal effects was utilised. Data were for the 50 most frequently prescribed procedures by each of 677 primary care physicians in the area of metropolitan Montreal from 1977 to 1983. Physicians adjusted their consultations, both quantitatively and qualitatively, to defend their incomes. The results were parallel to the phenomenon of DRG creep for hospitals.

Type of study: Quasi-experimental study

O'Malley *et al* (2006) evaluated the effectiveness of four different interventions directed at increasing dispensing rates for generic drugs, with one of the interventions being financial incentives for physicians. Reward payments to physician groups were made every six months, based on estimates of ingredient costs saved, with total payments ranging from $250 to $500 per physician per six months. Claims data from BCBS of Michigan were used to evaluate the impact of four interventions (written communication with members, statewide advertising, physician financial incentives and pharmacist visits with physicians). Control groups were constructed for each intervention and logistic regression models were estimated with the dependent variable indicating if the prescription was filled with a generic. None of the interventions, including physician financial incentives, increased use of generics.

Type of study: Quasi-experimental

Pockney *et al* (2004) completed an empirical review of the evidence regarding a response by primary care physicians in the UK to the institution of a fee-for-service payment system for certain procedures. Data were collected from 17 health authorities but the researchers used information only from the six authorities which provided a full eight years of data. Rates were calculated for each type of procedure for each year. Total claims rose throughout the period from 1993 to 2000, with the exception of the year 2000. Most of the increase was due to a rise in cautery using cryotherapy (mostly for wart removal). The authors stated that this procedure was no better than cheap, readily available over-the-counter treatments for warts. The authors referred to unpublished data that suggested that there was no decline in procedures done by specialists. Issues relating to poorer quality care are raised in the discussion section.

Type of study: Observational study without control group

Ransom *et al* (1996) compared surgical services in a gynaecology clinic under fee-for-service with capitated payment incentives to evaluate the variation in physician behaviour leading to performance of gynaecologic surgical procedures. The researchers examined 3,780 consecutive gynaecologic visits in a university practice during 1994. Visits during the first six months were reimbursed fee-for-service; during the second six months the same physicians received capitated payments and were paid on a salary basis. Chi-squared analysis was used to assess the differences. There was a 15 per cent overall decrease in the number of procedures performed during the capitated period. The reductions occurred primarily in elective procedures. Practice patterns were unchanged for severe problems.

Type of study: Controlled observational study

Stanton and Shortt (2003) analysed the impact of a change from fee-for-service to block funding payment system on surgical practices in a Canadian academic medical centre. Before the change, clinical staff billed the province and were paid fee-for-service. After the change, the medical centre received a fixed yearly sum, from which clinical staff were paid. Each physician's payment was based primarily on historical billings. Income was decoupled from volume of services provided and was static for five years. The idea was to remove incentives for 'marginally necessary' care. Four sentinel procedures known to have a large discretionary component were followed. The performance of the four sentinel procedures at the medical centre before and after the change was compared with the performance at four other academic centres in Ontario, Canada. The analysis was carried out over five years (split evenly before and after the change) using administrative data. The period covered was from 1992 to 1996. Multivariate repeated measures analysis was used. Changing the payment mechanism did not alter practice patterns. A limitation was that the change in payment may have had very little impact on physician incomes because of the way it was implemented.

Type of study: Quasi-experimental study

Tai-Seale, Rice and Stearns (1998) conducted an empirical investigation of how physicians adjusted volume of care they provided in the face of Medicare fee reductions. A panel dataset covering discharges from about 200 hospitals over 45 months was used in the analysis. A fixed effects model with generalised least squares and instrumental variable specifications was estimated for eight specialties experiencing varying degrees of payment reductions. The findings from the analysis were not clear cut. Physicians performed more surgeries both for their Medicare and private patients. The results did not consistently support theoretical predictions.

Type of study: Controlled observational study

System-level comparisons

We identified only two studies that analysed the impact of financial incentives on providers that were multinational in nature or analysed impacts at the system level. Wodchis, Fries and Pollack (2004) compared the effect of payment incentives on provision of physical and occupational therapy to non-elderly nursing home residents in the USA (where Medicare payment changed from a cost-based approach to per patient payment) and Canada (which retained a budget approach to financing care). They concluded that there was a greater likelihood of Medicare beneficiaries receiving occupational therapy after the change, but fewer minutes of occupational and physical therapy were provided. Shorten and Shorten (2004) examined the impact of public and private insurance coverage in New South Wales, Australia on maternity treatments. They found a higher incidence of birth intervention and higher operative birth rates under private insurance, although the authors did not describe any conceptual link between private insurance and physician incentives.

Shorten and Shorten (2004) conducted an empirical analysis of the impact of private versus public health insurance coverage in New South Wales, Australia on maternity treatments and outcomes. Data were provided by the New South Wales Midwives Data Collection Unit during the period from 1997 to 2001. Logistic regression methods were used to analyse the data. Private health insurance appeared to have a negative impact on childbirth, in terms of higher birth intervention and higher operative birth rates. The authors speculated that economic incentives for hospitals in treating patients with private insurance may explain this. The conceptual link between private insurance and physician behaviour was not discussed.

Type of study: Controlled observational study

Wodchis, Fries and Pollack (2004) conducted a retrospective US/Canada cross-sectional empirical analysis of the effect of payment incentives on provision of occupational therapy to non-elderly nursing home residents. Logistic and linear regression methods were used to analyse the data. Data for 1998 and 1999 were extracted from the University of Michigan Assessment Archive Project. These data were gathered using the Minimum Data Set resident assessment instrument for nursing homes. The randomly-selected final sample included 2807 Ontario residents and 3803 Michigan residents. The Medicare payment change was associated with a greater likelihood of receiving occupational therapy but reduced minutes of both kinds of therapy.

Type of study: Controlled observational study

Conclusions

We have reviewed the literature regarding the impact of provider financial incentives on medical care utilisation and costs. The large number of studies on this general topic find, with very few exceptions, that providers will take steps to reduce utilisation or costs when rewarded to do so. However, the usefulness of this literature for decision-makers is limited in many respects. For instance, the exact nature of the financial incentives often was not explained and, as a consequence, possible causal linkages were not clear. Many of the studies used datasets with very small numbers of providers or practice settings. In these cases it was difficult to adequately control for the context in which the financial arrangements were implemented. In general, more information about context would have been helpful in drawing conclusions regarding the applicability of study findings to other contexts. Finally, except in a small number of studies, the empirical analysis was not carefully linked to an explicit theory of provider decision-making. This left the results open to a variety of ad hoc interpretations, especially where information on context was lacking. The strength of this body of research lies in the consistency of its finding that providers do respond to financial incentives, rather than in the insight it provides for the design of effective incentive arrangements.

APPENDIX 2.
Information on database searches

Database of Abstracts of Reviews of Effects (DARE) (CRD administration database)
Searched on 7 December 2005 (supplementary search August 2006)
Publication dates: 1995 onwards
No language or country limits were used.
Retrieved 110 hits

Search strategy:

1 s incentiv$
2 s competition$
3 s contest$
4 s lotter$
5 s reward$
6 s prize$
7 s bonus$
8 s token(w)economy
9 s salary(w)based (w)compensation
10 s compensatory(w)plan$
11 s target(w)pay$
12 s pay$(3w)result$
13 s reimbursement
14 s pay$(3w)performance
15 s contingent(w)pay$
16 s deposit(w)contract$
17 s clinical(w)productivity (w)compensation
18 s capitation
19 s fundhold$ or fund(w)hold$
20 s merit(w)award$
21 s reputation
22 s profession$(3w)standing
23 s profession$(3w)status
24 s profession$(3w)recog$
25 s career(w)mobility
26 s career$(3w)develop$
27 s staff(3w)develop$
28 s job(3w)develop$
29 s job(3w)satisfaction
30 s rating(w)system$
31 s report$(3w)physician (w)performance
32 s earned(w)autonomy
33 s market(w)share
34 s clinical(w)volume
35 s case(w)load
36 s work(w)relative(w)value (w)unit$
37 s clinical(w)productivity
38 s patient(w)volume
39 s fringe(w)benefit$
40 s performance(w)report$
41 s revenue(3w)enhance$
42 s revenue(3w)maximi$
43 s revenue(3w)increas$
44 s income(3w)enhance$
45 s income(3w)maximi$
46 s income(3w)increas$
47 s recognition(w)award$
48 s gp$(w)contract$
49 s consultant$(w)contract$
50 s general(w)practitioner$ (w)contract$
51 s salary
52 s salaries
53 s inducement$ andnot (pregnan$ or labor or labour or deliver$)
54 s quit(3w)win
55 s voucher$
56 s social(w)market$
57 s social(w)franchis$
58 s s1 or s2 or s3 or s4 or s5 or s6 or s7 or s8 or s9 or s10 or s11 or s12 or s13 or s14 or s15 or s16 or s17 or s18 or s19 or s20 or s21 or s22 or s23 or s24 or s25 or s26 or s27 or s28 or s29 or s30 or s31 or s32 or s33 or s34 or s35 or s36 or s37 or s38 or s39 or s40 or s41 or s42 or s43 or s44 or s45 or s46 or s47 or s48 or s49 or s50 or s51 or s52 or s53 or s54 or s55 or s56 or s57
59 s 1995:2005/xyr
60 s s59 and s58

Cochrane Database of Systematic Reviews (CDSR)
(www.thecochranelibrary.com)
Issue 4 2005
Searched on 6 December 2005 (supplementary search August 2006)
No language or country limits were used.
Publication dates: no limits
Retrieved 31 hits

Search strategy:

#1 MeSH descriptor Employee Incentive Plans, this term only in MeSH products

#2 MeSH descriptor Salaries and Fringe Benefits explode all trees in MeSH products

#3 MeSH descriptor Reimbursement, Incentive, this term only in MeSH products

#4 MeSH descriptor Reward explode all trees in MeSH products

#5 MeSH descriptor Physician Incentive Plans, this term only in MeSH products

#6 incentiv* in Record Title or incentiv* in Abstract

#7 competition* in Record Title or competition* in Abstract

#8 contest* in Record Title or contest* in Abstract

#9 lotter* in Record Title or lotter* in Abstract

#10 reward* in Record Title or reward* in Abstract

#11 prize* in Record Title or prize* in Abstract

#12 bonus* in Record Title or bonus* in Abstract

#13 MeSH descriptor Capitation Fee, this term only in MeSH products

#14 “token economy” in Record Title or “token economy” in Abstract

#15 “salary based compensation” in Record Title or “salary based compensation” in Abstract

#16 “compensatory plan*” in Record Title or “compensatory plan*” in Abstract

#17 “target pay*” in Record Title or “target pay*” in Abstract

#18 pay* near/3 result* in Record Title or pay* near/3 result* in Abstract

#19 reimbursement in Record Title or reimbursement in Abstract

#20 pay* near/3 performance in Record Title or pay* near3 performance in Abstract

#21 “contingent pay*” in Record Title or “contingent pay*” in Abstract

#22 “deposit contract*” in Record Title or “deposit contract*” in Abstract

#23 “clinical productivity compensation” in Record Title or “clinical productivity compensation” in Abstract

#24 capitation in Record Title or capitation in Abstract

#25 fundhold* or “fund hold*” in Record Title or fundhold* or “fund hold*” in Abstract

#26 “merit award*” in Record Title or “merit award*” in Abstract

#27 reputation in Record Title or reputation in Abstract

#28 profession* near/3 standing in Record Title or profession* near/3 standing in Abstract

#29 profession* near/3 status in Record Title or profession* near/3 status in Abstract

#30 profession* near/3 recog* in Record Title or profession* near/3 recog* in Abstract

#31 “career mobility” in Record Title or “career mobility” in Abstract

#32 career* near/3 develop* in Record Title or career* near/3 develop* in Abstract

#33 staff near/3 develop* in Record Title or staff near/3 develop* in Abstract

#34 job near/3 develop* in Record Title or job near/3 develop* in Abstract

#35 job near/3 satisfaction in Record Title or job near/3 satisfaction in Abstract

#36 “rating system*” in Record Title or “rating system*” in Abstract

#37 report* near/3 “physician performance” in Record Title or report* near/3 “physician performance” in Abstract

#38 “earned autonomy” in Record Title or “earned autonomy” in Abstract

#39 “market share” in Record Title or “market share” in Abstract

#40 “clinical volume” in Record Title or “clinical volume” in Abstract

#41 “case load” in Record Title or “case load” in Abstract

#42 “work relative value unit*” in Record Title or “work relative value unit*” in Abstract

#43 “clinical productivity” in Record Title or “clinical productivity” in Abstract

#44 “patient volume” in Record Title or “patient volume” in Abstract

#45 “fringe benefit*” in Record Title or “fringe benefit*” in Abstract

#46 “performance report*” in Record Title or “performance report*” in Abstract

#47 revenue near/3 enhance* in Record Title or revenue near/3 enhance* in Abstract

#48 revenue near/3 maximi* in Record Title or revenue near/3 maximi* in Abstract

#49 “revenue near/3 increas*” in Record Title or “revenue near/3 increas*” in Abstract

#50 income near/3 enhance* in Record Title or income near/3 enhance* in Abstract

#51 income near/3 maximi* in Record Title or income near/3 maximi* in Abstract

#52 income near/3 increas* in Record Title or income near/3 increas* in Abstract

#53 “recognition award*” in Record Title or “recognition award*” in Abstract

#54 MeSH descriptor Awards and Prizes explode all trees in MeSH products

#55 “gp* contract*” in Record Title or “gp* contract*” in Abstract in all products

#56 “consultant* contract*” in Record Title or “consultant* contract*” in Abstract

#57 “general practitioner* contract*” in Record Title or “general practitioner* contract*” in Abstract

#58 salary in Record Title or salary in Abstract

#59 salaries in Record Title or salaries in Abstract

#60 inducement* not (pregnan* or labor or labour or deliver*) in Record Title or inducement* not (pregnan* or labor or labour or deliver*) in Abstract

#61 quit near/3 win in Record Title or quit near/3 win in Abstract

#62 voucher* in Record Title or voucher* in Abstract

#63 “social market*” in Record Title or “social market*” in Abstract

#64 “social franchis*” in Record Title or “social franchis*” in Abstract

#65 MeSH descriptor Social Marketing, this term only in MeSH products

#66 (#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29 OR #30)

#67 (#31 OR #32 OR #33 OR #34 OR #35 OR #36 OR #37 OR #38 OR #39 OR #40 OR #41 OR #42 OR #43 OR #44 OR #45 OR #46 OR #47 OR #48 OR #49 OR #50 OR #51 OR #52 OR #53 OR #54 OR #55 OR #56 OR #57 OR #58 OR #59 OR #60 OR #61 OR #62 OR #63 OR #64 OR #65)

#68 (#66 OR #67)

MEDLINE – Winspears
1995–2005 week 10
Searched on 6 December 2005 (supplementary search August 2006)
No language or country limits were used.
Publication dates: 1995 onwards
Retrieved 6558 hits

Search strategy:

1. "Employee-Incentive-Plans"/ all subheadings
2. explode "Salaries-and-Fringe-Benefits"/ all subheadings
3. "Reimbursement-Incentive"/ all subheadings
4. explode "Reward"/ all subheadings
5. "Physician-Incentive-Plans"/ all subheadings
6. incentiv* in ab,ti
7. competition* in ab,ti
8. contest* in ab,ti
9. lotter* in ab,ti
10. reward* in ab,ti
11. prize* in ab,ti
12. bonus* in ab,ti
13. "Capitation-Fee"/ all subheadings
14. (token adj economy) in ab,ti
15. (salary adj based adj compensation) in ab,ti
16. (compensatory adj plan*) in ab,ti
17. (target adj pay*) in ab,ti
18. (pay* near3 result*) in ab,ti
19. reimbursement in ab,ti
20. (pay* near3 performance) in ab,ti
21. (contingent adj pay*) in ab,ti
22. (deposit adj contract*) in ab,ti
23. (clinical adj productivity adj compensation) in ab,ti
24. capitation in ab,ti
25. fundhold* or (fund adj hold*) in ab,ti
26. (merit adj award*) in ab,ti
27. reputation in ab,ti
28. (profession* near3 standing) in ab,ti
29. (profession* near3 status) in ab,ti
30. (profession* near3 recog*) in ab,ti
31. (career adj mobility) in ab,ti
32. (career* near3 develop*) in ab,ti
33. (staff near3 develop*) in ab,ti
34. (job near3 develop*) in ab,ti
35. (job near3 satisfaction) in ab,ti
36. (rating adj system*) in ab,ti
37. (report* near3 (physician adj performance)) in ab,ti
38. (earned adj autonomy) in ab,ti
39. (market adj share) in ab,ti
40. (clinical adj volume) in ab,ti
41. (case adj load) in ab,ti
42 (work adj relative adj value adj unit*) in ab,ti
43. (clinical adj productivity) in ab,ti
44. (patient adj volume) in ab,ti
45. (fringe adj benefit*) in ab,ti
46. (performance adj report*) in ab,ti
47. (revenue near3 enhance*) in ab,ti
48. (revenue near3 maximi*) in ab,ti
49. (revenue near3 increas*) in ab,ti
50. (income near3 enhance*) in ab,ti
51. (income near3 maximi*) in ab,ti
52. (income near3 increas*) in ab,ti
53. (recognition adj award*) in ab,ti
54. explode "Awards-and-Prizes"/ all subheadings
55. (gp* adj contract*) in ab,ti
56. (consultant* adj contract*) in ab,ti
57. (general adj practitioner* adj contract*) in ab,ti
58. salary in ab,ti
59. salaries in ab,ti
60. (inducement* not (pregnan* or labor or labour or deliver*)) in ab,ti
61. (quit near3 win) in ab,ti
62. voucher* in ab,ti
63. social market* in ab,ti
64. social franchis* in ab,ti

65. “Social-Marketing”/ all subheadings
66. #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24 or #25 or #26 or #27 or #28 or #29 or #30
67. #31 or #32 or #33 or #34 or #35 or #36 or #37 or #38 or #39 or #40 or #41 or #42 or #43 or #44 or #45 or #46 or #47 or #48 or #49 or #50 or #51 or #52 or #53 or #54 or #55 or #56 or #57 or #58 or #59 or #60 or #61 or #62 or #63 or #64 or #65
68. #66 or #67
69. review in ab
70. review in pt
71. meta-analysis in ab
72. meta-analysis in pt
73. meta-analysis in ti
74. #69 or #70 or #71 or #72 or #73
75. letter in pt
76. editorial in pt
77. comment in pt
78. #75 or #76 or #77
79. #74 not #78
80. (retrospective* near2 review*) in ab,ti,mesh
81. (case* near2 review*) in ab,ti,mesh
82. (record* near2 review*) in ab,ti,mesh
83. (patient* near2 review*) in ab,ti,mesh
84. (patient* near2 chart*) in ab,ti,mesh
85. (peer* near2 review*) in ab,ti,mesh
86. (rat or rats or mouse or mice or hamster or hamsters or animal or animals or dog or dogs or cat or cats or bovine or sheep) in ab,ti,mesh
87. #80 or #81 or #82 or #83 or #84 or #85 or #86
88. #79 not #87
89. animals in tg
90. humans in tg
91. #89 not (#89 and #90)
92. #88 not #91
93. #92 and #68

EMBASE
(www.ovid.com/)
1988–2005 week 49
Searched on 8 December 2005 (Supplementary search August 2006)
No language or country limits were used.
Publication dates: 1995 onwards
Retrieved 1256 hits

Search strategy:

1 exp Income/
2 Reward/
3 incentiv$.ab,ti.
4 competition$.ab,ti.
5 contest$.ab,ti.
6 lotter$.ab,ti.
7 reward$.ab,ti.
8 prize$.ab,ti.
9 bonus$.ab,ti.
10 Capitation-Fee/
11 token economy.ab,ti.
12 salary based compensation.ab,ti.
13 compensatory plan$.ab,ti.
14 target pay$.ab,ti.
15 pay$ adj3 result$.ab,ti.
16 reimbursement.ab,ti.
17 pay$ adj3 performance.ab,ti.
18 contingent pay$.ab,ti.
19 deposit contract$.ab,ti.
20 clinical productivity compensation.ab,ti.
21 capitation.ab,ti.
22 fundhold$.ab,ti. or fund hold$.ab,ti.
23 merit award$.ab,ti.
24 reputation.ab,ti.
25 profession$ adj3 standing.ab,ti.
26 profession$ adj3 status.ab,ti.
27 profession$ adj3 recog$.ab,ti.
28 career adj mobility.ab,ti.
29 career$ adj3 develop$.ab,ti.
30 staff adj3 develop$.ab,ti.
31 job adj3 develop$.ab,ti.
32 job adj3 satisfaction.ab,ti.
33 rating adj system$.ab,ti.
34 report$ adj3 physician performance.ab,ti.
35 earned autonomy.ab,ti.
36 market share.ab,ti.
37 clinical volume.ab,ti.
38 case adj load.ab,ti.
39 work relative value unit$.ab,ti.
40 clinical productivity.ab,ti.
41 patient volume.ab,ti.
42 fringe benefit$.ab,ti.
43 performance report$.ab,ti.
44 revenue adj3 enhance$.ab,ti.
45 revenue adj3 maximi$.ab,ti.
46 revenue adj3 increas$.ab,ti.
47 income adj3 enhance$.ab,ti.
48 income adj3 maximi$.ab,ti.
49 income adj3 increas$.ab,ti.
50 recognition award$.ab,ti.
51 gp$ contract$.ab,ti.
52 consultant$ contract$.ab,ti.
53 general practitioner$ contract$.ab,ti.
54 salary.ab,ti.
55 salaries.ab,ti.
56 inducement$.ab,ti. not (pregnan$ or labor or labour or deliver$).ab,ti.
57 quit adj3 win.ab,ti.
58 voucher$.ab,ti.
59 social market$.ab,ti.
60 social franchis$.ab,ti.
61 Social-Marketing/
62 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30
63 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51 or 52 or 53 or 54 or 55 or 56 or 57 or 58 or 59 or 60 or 61
64 62 or 63
65 exp meta analysis/
66 meta-analys$.ti,ab.
67 metaanalys$.ti,ab.
68 meta analys$.ti,ab.
69 review$.ti.
70 overview$.ti.

71 (synthes$ adj3 (literature$ or research$ or studies or data)).ti,ab.

72 pooled analys$.ti,ab.

73 ((data adj2 pool$) and studies).mp.

74 (medline or medlars or embase or cinahl or scisearch or psychinfo or psycinfo or psychlit or psyclit).ti,ab.

75 ((hand or manual or database$ or computer$) adj2 search$).ti,ab.

76 ((electronic or bibliographic$) adj2 (database$ or data base$)). ti,ab.

77 ((review$ or overview$) adj10 (systematic$ or methodologic$ or quantitativ$ or research$ or literature$ or studies or trial$ or effective$)).ab.

78 65 or 66 or 67 or 68 or 69 or 70 or 71 or 72 or 73 or 74 or 75 or 76 or 77

79 (retrospective$ adj2 review$).ti,ab,sh.

80 (case$ adj2 review$). ti,ab,sh.

81 (record$ adj2 review$). ti,ab,sh.

82 (patient$ adj2 review$). ti,ab,sh.

83 (patient$ adj2 chart$). ti,ab,sh.

84 (peer adj2 review$).ti,ab,sh.

85 (chart$ adj2 review$). ti,ab,sh.

86 (case$ adj2 report$). ti,ab,sh.

87 (rat or rats or mouse or mice or hamster or hamsters or animal or animals or dog or dogs or cat or cats or bovine or sheep).ti,ab,sh.

88 79 or 80 or 81 or 82 or 83 or 84 or 85 or 86 or 87
89 78 not 88
90 editorial.pt.
91 letter.pt.
92 90 or 91
93 89 not 92
94 animal/
95 human/
96 94 or 95
97 human/
98 96 not (96 and 97)
99 93 not 98

100 64 and 99

101 Limit 1995-2005

APPENDIX 3:
Information on database searches (Phase 2)

Phase 2 searches focused on fewer search terms but were more inclusive in terms of research design. However, it was considered desirable to try and filter out as many general articles as possible, and to this end a study filter was used.

The advantage of using a filter is that it can help to reduce the hits retrieved to a manageable number, by excluding irrelevant material. This was particularly desirable for this topic since initial searches (without the study filter) indicated that there was a vast amount of literature available, and that much of it would not be relevant.

There has been much research done on the effectiveness of filters to identify randomised controlled trials on MEDLINE, but very little on the identification of other types of primary research. One problem with searching for other types of studies is that the type of study used is often poorly reported in articles. A second issue is that study types are poorly indexed in most databases. As a result, any attempt to limit a search by study type runs the risk of accidentally excluding relevant material. For this reason, filters should be used with care.

The following databases were searched in July 2006 (with a supplemental search in August 2006 prompted by new publications coming to light).

These databases were searched from inception to the date of the search.

No language or country limits were used.

Cochrane Central Register of Controlled Trials (CENTRAL)
(www.thecochranelibrary.com)
Issue 2 2006
Searched on 5 July 2006
Retrieved 906 hits
(CENTRAL consists solely of reports of trials, so it was not necessary to use a study filter.)

Search strategy:

#1 (incentiv*):ti or (incentiv*):ab in Clinical Trials

#2 (disincentiv*):ti or (disincentiv*):ab in Clinical Trials

#3 (lottery or lotteries or reward* or bonus* or prize*): ti or (lottery or lotteries or reward* or bonus* or prize*): ab in Clinical Trials

#4 (pay near/3 performance): ti or (pay near/3 performance):ab in Clinical Trials

#5 (merit award*):ti or (merit award*):ab in Clinical Trials

#6 (earned near/3 autonomy): ti or (earned near/3 autonomy):ab in Clinical Trials

#7 (#1 OR #2 OR #3 OR #4 OR #5 OR #6)

#8 (letter or editorial or comment):pt in Clinical Trials

#9 (#7 AND NOT #8)

MEDLINE
(www.ovid.com/)
1966–2006 (June), week 3
Searched on 5 July 2006
Retrieved 6982 hits

Search strategy:

1 incentiv$.ti,ab.

2 disincentiv$.ti,ab.

3 (lottery or lotteries or reward$ or bonus$ or prize$).ti,ab.

4 (pay adj3 performance).ti,ab.

5 merit award$.ti,ab.

6 (earned adj3 autonomy).ti,ab.

7 or/1-6

8 clinical trial.pt.

9 randomised.ab.

10 placebo.ab.

11 randomly.ab.

12 trial.ab.

13 groups.ab.

14 (case reports or clinical trial phase i or clinical trial phase ii or multicenter study or clinical trial phase iii or clinical trial phase iv or controlled clinical trial or randomised controlled trial or twin study or validation studies).pt.

15 Case control.ti,ab.

16 Cohort$.ti,ab.

17 Prospective.ti,ab.

18 Quantitative.ti,ab.

19 Longitudinal.ti,ab.

20 Comparator.ti,ab.

21 (Evaluation stud$ or Evaluation trial$).ti,ab.

22 comparators.ti,ab.

23 Follow-up.ti,ab.

24 (Intervention stud$ or Intervention trial$).ti,ab.

25 (Multicentre or multicenter). ti,ab.

26 (Open stud$ or open trial$). ti,ab.

27 Observational.ti,ab.

28 Feasibility stud$.ti,ab.

29 Pilot project$.ti,ab.

30 Sampling stud$.ti,ab.

31 Empirical.ti,ab.

32 Cross-over.ti,ab.

33 Matched pair.ti,ab.

34 retrospective.ti,ab.

35 Multivariate analysis.ti,ab.

36 Regression.ti,ab.

37 Correlat$.ti,ab.

38 Quality.ti,ab.

39 outcome$.ti,ab.

40 or/8-39

41 40 and 7

42 animals/ not (animals/ and humans/)

43 (letter or editorial or comment).pt.

44 42 or 43

45 41 not 44

EMBASE
(www.ovid.com/)
1980–2006 week 26
Searched on 5 July 2006
Retrieved 6742 hits

Search strategy:

1 incentiv$.ti,ab.
2 disincentiv$.ti,ab.
3 (lottery or lotteries or reward$ or bonus$ or prize$).ti,ab.
4 (pay adj3 performance).ti,ab.
5 merit award$.ti,ab.
6 (earned adj3 autonomy).ti,ab.
7 or/1-6
8 clinical trial.pt.
9 randomised.ab.
10 placebo.ab.
11 randomly.ab.
12 trial.ab.
13 groups.ab.
14 Case control.ti,ab.
15 Cohort$.ti,ab.
16 Prospective.ti,ab.
17 Quantitative.ti,ab.
18 Longitudinal.ti,ab.
19 Comparator.ti,ab.
20 (Evaluation stud$ or Evaluation trial$).ti,ab.
21 comparators.ti,ab.
22 Follow-up.ti,ab.
23 (Intervention stud$ or Intervention trial$).ti,ab.
24 (Multicentre or multicenter).ti,ab.
25 (Open stud$ or open trial$).ti,ab.
26 Observational.ti,ab.
27 Feasibility stud$.ti,ab.
28 Pilot project$.ti,ab.
29 Sampling stud$.ti,ab.
30 Empirical.ti,ab.
31 Cross-over.ti,ab.
32 Matched pair.ti,ab.
33 retrospective.ti,ab.
34 Multivariate analysis.ti,ab.
35 Regression.ti,ab.
36 Correlat$.ti,ab.
37 Quality.ti,ab.
38 outcome$.ti,ab.
39 or/8-39
40 39 and 7
41 animal/ not (animal/ and human/)
42 (letter or editorial or note).pt.
43 41 or 42
44 40 not 43

EconLit
(www.ovid.com/)
1969–2006
Searched on 3 July 2006
Retrieved 874 hits

1 incentiv$.mp

2 disincentiv$.mp

3 lottery or lotteries or reward$ or prize$ or voucher$.mp

4 pay and performance.mp

5 health or healthcare or NHS.mp

6 outcome$ or data or performance or quality or ratio$ or rate$ or empirical or experiment$ or randomi$. Mp

7 1 or 2 or 3 or 4

8 7 and 5 and 6

EconLit
(www.ovid.com/)
1969–2006
Searched on 30 August 2006
Retrieved 631 hits

1 capitat$.mp

2 fee for service.mp

3 reimburs$.mp

4 health or healthcare or NHS

5 1 or 2 or 3

6 5 and 4