

The Work Foundation transforms people's experience of work and the labour market through high quality applied research that empowers individuals and influences public policies and organisational practices. For further details, please visit www.theworkfoundation.com.

Acknowledgements

The work described in this report was carried out under contract as part of the Office for Manpower Economics (OME)'s research programme. However, the views and judgements expressed in this report are those of the Work Foundation and do not necessarily reflect those of OME.

We would like to thank OME staff, particularly Nicola Allison and Margaret McEvoy, who managed the project and provided helpful advice and comments on the report.

Special thanks are due to our advisory group of experts who provided extremely helpful and timely comments on report drafts and directed us to literature and sources:

- Professor Stephen Bevan, Director of the Centre for Workforce Effectiveness at the Work Foundation;
- Professor Colin Green, Professor of Economics at Lancaster University Management School;
- Professor John Heywood, Distinguished Visiting Scholar of Economics with Lancaster University and Director of the Masters in Human Resources and Labour Relations at the University of Wisconsin-Milwaukee;
- Professor Ian Walker, Professor of Economics at Lancaster University Management School;
- Professor Paul Sparrow, Director of the Centre for Performance Led HR and Professor of International Human Resource Management at Lancaster University.

We would also like to thank Andy Holgate, librarian in business and management resources at Lancaster University for conducting the literature searches; Nye Cominetti, Ceri Hughes, and Katy Jones for reviewing assistance and Anthony Hind for editorial assistance.

Table of Contents

Executive summary	4
1. Introduction	12
The context of the review	12
Aims and objectives	13
Use of PRP in the UK public sector.....	15
Our approach to the review	17
Review methods	18
Our short-list of evidence	20
Report outline	21
2. Conceptual Issues: PRP and the public sector	22
PRP and the public sector.....	23
PRP mechanisms	24
Challenges to PRP in the public sector	26
3. Previous findings and the evolving evidence base	30
Scope of the two reviews	30
Civil Service evidence.....	30
Healthcare Evidence.....	31
Education evidence.....	32
Remaining limitations of the evidence base	33
4. Findings: Recent evidence on the effectiveness of PRP in the public sector	36
The effects of PRP on incentivised outcomes	36
The effects of PRP on unincentivised outcomes	39
Unintended negative consequences of PRP	40
The effects of PRP on staff attitudes, motivation and behaviour.....	42
The effects of PRP on work organisation and team relationships.....	46

Cost effectiveness of PRP schemes	48
5. Performance-related pay scheme design	50
6. Conclusions	59
Unintended behavioural effects of PRP.....	59
Performance measurement problems	63
Sub-optimal or worse social outcomes	64
References.....	67
Appendix A: List of included studies.....	75
Reviews	75
Impact studies.....	78
Process studies.....	85
Contact details	90

Executive summary

The Office of Manpower Economics (OME)¹, commissioned The Work Foundation to conduct a review of the recent literature on the impact, effectiveness, and value for money of performance-related pay (PRP) systems in the public sector. Performance-related pay is defined as pay systems in which some component of remuneration is contingent on individual (or team/organisation) performance, measured by objective criteria and/or subjective assessments (Bregm, 2013). The review updates and builds on a previously commissioned review of PRP in 2007 (Prentice et al, 2007), providing a picture of the emerging evidence-base since then.

PRP is still relatively uncommon in the public sector in the UK (van Wanrooy et al, 2013), although its use is increasing. Currently, the only well-established, large-scale example of a national pay-for-performance system in the UK public sector is the Quality and Outcomes Framework (QOF) introduced in 2004 for general practitioners (GPs). The QOF provides payments to GPs for compliance with targets across a spectrum of clinical activities (Coleman, 2010). In the civil service, individual PRP schemes have been widespread for some time, while team bonuses have also been trialled among staff administering taxation and benefits. More recently the Government has also committed to ending automatic pay progression (where it still exists), and linking progression to performance in the civil service, schools, the NHS, prisons and the police, with these services at varying stages of implementing the reforms.

Our approach to the review recognises the complexity of PRP and the different perspectives that academic disciplines bring to understanding it. We therefore include both evidence grounded in the Economics literature – that explores PRP through quantitative analysis of the impact of financial incentives on a range of outcomes – and evidence from the Human Resources, Sociology of Management and Psychology literature – that places more emphasis on explaining the mechanisms by which performance-related pay may operate, with a focus on organisational context, non-financial motivations and group behaviour.

We used a transparent approach to searching for and selecting the relevant literature, appraising studies according to their relevance to the review questions and the methods employed. 59 studies were included in the review. Of these, 27

¹ The Office for Manpower Economics provides an independent secretariat to eight public sector Pay Review Bodies. Together these bodies make recommendations for the pay of the armed forces, doctors and dentists, the National Health Service (NHS), the prison service, school teachers, the senior Civil Service, the National Crime Agency and the police.

focused on health, 16 on education and 16 on the civil service, which reflects the make-up of the current evidence base.

PRP and the public sector

Under the most simple logic chain, PRP aims to strengthen the link between the employee input of effort and the incentives they receive in terms of pay. This should motivate the individual to expend more effort and increase their outputs, which should ultimately improve the organisation's outcomes. However, this model neglects the wider range of factors that may motivate employees, including peer effects, perceptions of fairness and intrinsic motivations. These issues may be particularly applicable to the public sector. For example, a number of studies highlight the existence of some form of 'public service motivation' among public sector employees, suggesting that they may derive much of their motivation from their belief in the intrinsic value of the service they perform, rather than its financial reward (Anderfuhren-Biget et al., 2010; Ashraf et al., 2014; Leigh, 2013; OECD, 2009; Taylor and Taylor, 2011). There may even be a risk that performance pay can detract from, or 'crowd out', intrinsic motivation, thus subverting the intentions of a PRP scheme.

The nature of the public sector's activities and objectives also creates a number of challenges for using performance measurement as a basis for pay:

- **The complexity of the public service 'good':** Public services generate a multitude of outcomes, some of which are more easily measured than others. In addition, the ultimate outcomes of many public sector activities may only be visible in the long-term, raising questions about the feasibility of accurate and meaningful performance measures within a PRP scheme.
- **Multiple principals:** The public sector involves a wide variety of potential 'owners' and stakeholders (service users, managers, unions, professional bodies, the Government, taxpayers). Any PRP scheme in the public sector must be capable of reconciling the variety of outcomes from these multiple stakeholders and interests.
- **Multi-task problems and collaborative activity:** The delivery of public services tends to be a complex and inherently collaborative activity. Attributing individual responsibility for performance and outcomes may therefore be challenging, and individual incentives could mitigate against team work.
- **Misallocation of effort:** PRP schemes may incentivise outcomes which are more easily and directly measurable (OECD, 2009), encouraging employees to focus on these outcomes at the expense of others, e.g. 'teaching to the test' in education.
- **Gaming:** Where performance indicators become 'high stakes' this may lead to attempts to game the system (Neal, 2011), where workers seek to

maximize their gains while minimizing effort or without increasing performance. This can lead to significant problems in the public sector, where outcomes can have a wide social impact.

Nature of the evidence base

The 2007 Review found evidence on PRP for the civil service, healthcare workers and teachers (Prentice et al, 2007). In 2014, while the literature in each of these areas had expanded, the occupational focus has remained similar. Since 2007, the literature for health, in particular, has burgeoned, with a number of systematic reviews of various aspects of PRP in healthcare, largely US and UK-based, with a significant literature on the impact of the QOF in the UK. Post-2007, the literature on PRP in education has also grown, although it is smaller than that in health. The geographical focus is predominantly the US, and most impact studies focus on the effect of PRP on pupils' scores in assessment. The recent evidence base in the civil service comprises reviews of earlier evidence and studies examining the interaction of PRP with public service motivation, and the effects of PRP implementation on staff behaviour. Despite the burgeoning literature, key limitations in the evidence base remain, the most important of which are:

- **Occupational coverage:** Gaps include the armed forces, prisons, the police and the judiciary. Moreover, the evidence is primarily focused on those in professional occupations, rather than the entirety of the public sector workforce.
- **Timescales of initiatives and evaluations:** Some incentive schemes are either relatively new or were implemented for only a short period of time; evidence on long-term effects is limited. This is a problem as a PRP scheme's impact on attitudes and outcomes may well change over time.
- **Robust evidence on scheme design:** Reviews of the evidence suggest that the effects of PRP initiatives vary according to several factors including scheme design; however, heterogeneity between studies have inhibited meta-analysis, making it difficult to draw firm conclusions about which types of schemes are most successful.
- **Cost effectiveness of performance-related pay:** There is very little robust cost-benefit analysis of PRP, or research assessing the relative merits of PRP vis-à-vis other types of incentives.

Effects of PRP on incentivised and non-incentivised outcomes

There is some evidence of positive effects from PRP schemes on directly incentivised outcomes across education, health and the civil service. However, the overall conclusion is that findings are mixed, and often context- or outcome-specific, making it difficult to draw overall conclusions about the effectiveness of PRP for a particular public service.

- In **education**, there is widespread evidence that the introduction of PRP results in positive impacts on the incentivised outcome, which is usually students' performance in standardised tests (Podgursky, 2007, OECD, 2009). However, the evidence is not completely consistent (e.g. Goodman and Turner, 2009; Springer et al, 2012) and much depends on the scheme design, teachers' support, and wider accountability and performance measures. The effects of PRP on other outcomes – such as students' levels of motivation and engagement – are under-researched.
- In **health**, there is also evidence of positive impacts of PRP, although again findings are mixed. In general, positive outcomes have been found most frequently for quality of care measures for chronic diseases, especially for diabetes and asthma, and for preventative health measures such as immunisations. However, while positive effects have been widely found, improvements are often small and sometimes short lived. In acute care (hospitals), there is a less extensive evidence base on PRP and results tend to be, overall, less positive than for primary care.
- In the **civil service**, there is some evidence that PRP can have a positive impact on incentivised outcomes but results are again mixed. There is a dearth of recent high quality impact evidence.

Unintended consequences

The 2007 review identified some evidence of strategic behaviour or 'gaming' as a result of PRP, and in more recent studies there is further evidence of this. In education, there have been cases of outright cheating, where school officials have conspired to alter students' test scores, as well as examples of low-level strategic behaviour, for example focusing on the performance of 'borderline' pupils in order to achieve higher pass rates (Lavy, 2009). In health, Gravelle et al (2010) found a non-trivial amount of 'gaming' of the QOF system by reducing the number of eligible patients in order to raise performance on a target measure. However, the extent of this behaviour is contested (Van Herck et al, 2010). There is also some evidence of 'effort displacement' where un-incentivised aspects of care have not improved at all or as much as those that are directly rewarded. However, conversely, there is also some evidence of 'positive spillovers' in health, where non-incentivised aspects of care have improved alongside those directly targeted (Eijkenaar et al 2013; Elovainio, 2010; Houle et al 2012; Herck et al 2010).

Effects on staff attitudes and motivation

There are a range of studies on the effects of PRP on staff attitudes and motivation, several of which consider whether introducing financial incentives might 'crowd out' public servants' *intrinsic* motivation. The answer to this question is inconclusive though, with variation across services and individual studies.

This question has been most extensively discussed in the **civil service** literature. Here the evidence suggests that intrinsic motivation is as, or more, important than financial incentives in motivating behaviour, and there is also some evidence that financial incentives can 'crowd out' this intrinsic motivation. Whether it does so depends both on the design of the scheme and its implementation. Some research suggests that more participative schemes enhance employee acceptance and motivation (Schmidt et al, 2011).

In **education**, several studies found that PRP had a detrimental effect on teachers' job satisfaction and motivation (Belfield and Heywood, 2008; Gius, 2013). Teachers are often unsupportive of PRP systems, particularly those based on standardised test scores, feeling they fail to capture important aspects of performance (Leigh, 2013; Yuan et al, 2013). Studies also suggest that the response of teachers to PRP may differ according to individual characteristics, such as gender and level of prior experience. For example, male teachers appear more likely to support PRP and to respond to PRP systems more positively than their female counterparts (Leigh, 2013; Jones, 2013). Experienced teachers also tended to display more negative reactions to PRP than early career teachers. These findings have implications for the acceptance and success of PRP in the public sector, since the majority of staff in the NHS and in education, for example, are female (Cribb et al, 2014).

In **health**, the evidence is limited on the effects of PRP on staff attitudes and motivations and findings are mixed. Some UK studies (mostly qualitative) have suggested that PRP has resulted in a perceived loss of autonomy for healthcare providers and undermined their sense of professionalism, while others have suggested that staff are largely supportive of the QOF since the incentivised targets are perceived to be closely aligned with their professional priorities (given that they are evidence-based) (Gillam et al, 2012; Lester et al, 2013).

Effects on individual and team behaviour

The evidence shows a range of effects of PRP on individual and team behaviour, which vary depending upon the structure of the target measures used.

In education, where targets are predominantly at the level of the individual teacher, some studies have shown an increase in teachers' discretionary effort (Lavy, 2009), but others have indicated a reduction in effort as a result of PRP (Jones, 2013). The effect on collegiality and co-operative activity is also variable – studies have shown both positive and negative effects from PRP (Belfield and Heywood, 2008; Jones, 2013; Yuan et al, 2013). The effects on individual behaviour also appear to vary by gender; for example, one study showed that women's (but not men's) participation in unpaid co-operative activities was reduced as a result of PRP. In health, there is some evidence of individual targets raising physicians' activity rates while the incentives were in place which dropped off thereafter (Petersen et al, 2013).

Team-based incentives offer the advantage of focusing on the broader structure and allocation of tasks, rather than purely individual effort. In the UK QOF system, which comprises practice-level targets, studies found performance improvements were maintained after targets were removed from the system, which suggested that performance improvements were obtained via changes to working practices rather than increasing individual effort per se (Kontopantelis et al, 2014). Similarly, in a pilot of team targets in the UK tax office, Burgess et al (2010) found that productivity was raised primarily via managers allocating tasks within the teams more efficiently.

Cost effectiveness

The cost-effectiveness of PRP schemes is an area that is under-researched, partly due to the difficulties in finding the necessary evidence on the costs and effects of interventions in order to make value-for-money assessments. In health, there have been some cost-effectiveness studies, but a review concluded that it was unable to come to robust conclusions due to 'variable methodological quality' and differences across programmes in context and design (Emmert et al, 2012). A cost-effectiveness study of the UK QOF (Walker et al, 2010) suggested that it would represent value for money at relatively small levels of impact. However the administrative costs of the scheme were not taken into account and only a small number of indicators were considered due to data limitations. A meta-review of a range of quality improvement strategies in health showed that clinician or patient-driven quality improvement strategies (such as audit and feedback cycles or clinical decision support systems) had stronger evidence of efficacy and larger effect sizes than manager or policy-driven strategies, including PRP, although relative cost-effectiveness data was not available (Scott, 2009).

Design of PRP schemes

Results from PRP schemes are often inconsistent and much of this variation can be linked to differences in scheme design. However, simple prescriptions for how PRP should be designed are not possible, since design must consider the context of the service involved. Key design features include:

Nature and range of goals/targets – A broader range of targets helps to prevent misallocation of effort, but simple, specific and easily measurable targets are clearer and easier to understand for staff and may enhance scheme effects. In health, process targets have yielded greater improvements than outcome targets, but for process targets to be effective, the link to desired outcomes needs to be robustly evidence-based.

Individual or team targets – Evidence in health and the civil service suggests that individual or small group targets are more effective than larger group or institutional level targets. This may be related to the free-riding problems associated with a

larger entity, as there is less opportunity for peer monitoring of behaviour, or to a lack of staff awareness of incentives in a larger grouping. Whether targets are at the individual or team level may also have differential effects on working relationships, with the risk that individual targets may detract from collaborative working relationships, especially important in the public sector.

Performance measures -

- **Absolute** measures of performance reduce the incentive for low performing providers, given their reduced likelihood of attaining the targets, but **relative** measures that reward improvement can end up rewarding the lowest performing providers the most because they are often able to improve the most. In health, absolute measures have been shown to be more effective.
- If the aim is general performance improvement, **linear** (or continuous) measures are preferable because they reward per unit improvement and thus encourage progress at all parts of the distribution. If there is one central performance goal (e.g. increasing the pass rate), a **threshold** measure may be preferable, since this rewards the achievement of a certain level. Threshold measures tend to result in a focus on the performance of those just below the threshold, rather than on those at the bottom of the distribution who are unlikely to meet the threshold.
- **Rank order** tournaments (which award bonuses to a set number of individuals) allow better control of expenditure, but depending on their perceived equity may disincentivise co-operation and demotivate individual effort through crowding-out effects.

Size and frequency of payments - Smaller bonuses may produce little effect, but larger payments may raise concerns about the cost-effectiveness of schemes. No robust relationship has been established between the size and frequency of payments and effect sizes, but larger rewards may increase participation levels in voluntary schemes.

Conclusions

The review has found evidence that PRP schemes can be effective in improving outcomes across the three public services for which evidence is available (health, education and the civil service), although the central conclusion is that the outcomes from PRP are mixed, which is much dependent upon organisational and occupational context and scheme design and implementation. Where positive effects have been found, effect sizes are sometimes small and may also be short-lived. As well as evidence gaps across much of the public services, the weight of evidence also varies, with the more robust evidence coming from education and health rather than the civil service. Cost-effectiveness data to assess the value for money of PRP interventions is also rare.

If PRP schemes are to be implemented in the public sector, the review highlights the following key challenges that should be taken into consideration:

- 1) Unintended behavioural consequences as a result of PRP incentives, such as:
 - gaming behaviour
 - crowding out effects
 - an absence of behavioural change
 - misallocation of effort and
 - detrimental consequences to teamwork and co-operation.
- 2) Difficulties in the measurement of outputs in public sector PRP, including
 - negative effects of particular performance measures (e.g. absolute/relative, linear/threshold) and
 - managerial subjectivity in assessment; and
- 3) A lack of fit between incentivised outputs and desired social outcomes, including:
 - Poor long-term outcomes; and
 - Poor cost-effectiveness

1. Introduction

The context of the review

The Office of Manpower Economics (OME) provides an independent secretariat to eight Pay Review Bodies. Together these bodies make recommendations for the pay of the armed forces, doctors and dentists, the National Health Service (NHS), the prison service, school teachers, the senior Civil Service, the National Crime Agency and the police, impacting on 2.5 million workers, or around 40% of public sector staff in the UK. The OME supports the review bodies in offering evidence-based, independent advice on public sector pay and conditions. In February 2014, the OME commissioned a review of the evidence on the impact, effectiveness, and value for money of performance-related pay (PRP) in public services. This report presents the results of that review.

The review has been undertaken by Lancaster University's Work Foundation, a leading provider of research-based analysis, knowledge exchange and policy advice that focuses on people's experience of work and the labour market. The research team at The Work Foundation drew on the expertise of a panel of advisers from the Lancaster University Management School, which includes specialists in labour market and personnel economics, together with leading thinkers in workforce management and human resources.

The OME had previously commissioned a review of performance-related pay in 2007: "*Performance pay in the public sector: A review of the issues and evidence*" (Prentice et al, 2007). This new review has taken account of those earlier findings, and built on them to provide a picture of the emerging evidence on PRP in public services since 2007.

This review must be seen within the context of the wider picture of both economic change and public sector reform. Ongoing economic changes, such as the lasting effects of the financial crisis and recession, pressures on public spending, lower overall productivity and real wages, and a changing labour market, all have implications for the potential of performance-related pay. The period since 2007 has also seen significant public sector reforms which are relevant across public sector occupations. In particular, since 2010 there have been reforms to public sector pay and conditions and major reviews and policy changes that affect staff in the NHS, the Civil Service, the armed forces, policing and justice, and in primary and

secondary education. In the context of fiscal constraint and reduced public spending, there is increasing emphasis on value-for-money, efficiency and responsiveness in public services.

Aims and objectives

The overall aim of the research set out by the Office of Manpower Economics was:

To review the recent literature on the impact, effectiveness, and value for money of performance-related pay systems in public services in the UK.

This is an ambitious aim; from the outset it is helpful to emphasise that:

- (i) The focus of this review is on the recent literature. Our reference point is the review that was undertaken in 2007 by the Centre for Market & Public Organisation and the University of Warwick (Prentice et al, 2007). This means we have not revisited all the references within that review. We have instead looked to articulate the messages in that earlier review and test the extent to which newer (post-2007) evidence builds on or challenges those messages.
- (ii) The focus of the review is on ‘performance-related pay’. This is defined as pay systems in which some component of remuneration is contingent on individual (or team/organisation) performance, measured by objective criteria and/or subjective assessments (Bregm, 2013). In practice, it is often difficult to disentangle performance-related pay from related issues of progression and pay, performance management, and aspects of Human Resources Management. This means the topic can easily expand into much wider concerns and debates. Throughout the review, we bring the focus back to the target issue of performance-related pay.
- (iii) The focus is ‘public services in the UK’. This does not prevent us from considering evidence from overseas, or from the private sector if relevant, but the focus is principally on the public services in the UK that are most relevant to the occupations and activities covered by the Pay Review Bodies and boards (see Figure 1.1).

Figure 1.1: Pay Review Bodies supported by the Office of Manpower

Economics

Body	Public Sector Occupations
Armed Forces' Pay Review Body	Members of Naval, Military and Air Forces
Review Body on Doctors' and Dentists' Remuneration	Doctors and Dentists
NHS Pay Review Body	NHS staff with the exception of doctors, dentists and some senior managers
Prison Service Pay Review Body	Governors, operational managers, prison officers
School Teachers' Review Body	School teachers (primary and secondary)
Senior Salaries Review Body	Senior civil servants, judiciary, senior officers of armed forces, senior managers in the NHS, Police and Crime Commissioners, senior police (from September 2014)
National Crime Agency Remuneration Review Body	Operational staff in the National Crime Agency
Police Remuneration Review Body	Police officers up to Chief Superintendent level (from September 2014)

The review is structured around six objectives, specified by the Office of Manpower Economics:

1. To provide an up-to-date (since 2007), comprehensive, independent and credible assessment of the relevant academic and other literature on performance-related pay in public services.
2. To identify the advantages and disadvantages of using performance-related pay schemes in public services.
3. To identify features of the public services that make the implementation of performance-related pay schemes different to the private sector.
4. To identify any environments or roles which are more conducive to effective performance-related pay than others.
5. To identify any types of performance-related pay schemes that are more or less effective than others – including any perverse effects or unintended consequences.
6. To highlight difficulties, weaknesses or gaps in the research and how these might be addressed.

These objectives are addressed in chapters 2 through 6 of the report. Underpinning these objectives is the need for the OME to have an evidence base that can clearly inform the arguments and recommendations put to the Pay Review Bodies.

Use of PRP in the UK public sector

Performance-related pay is still relatively uncommon in the public sector, both in the UK and elsewhere, although its use is increasing. According to the 2011 Workplace Employment Relations Study, only 7 per cent of public sector employees in Britain received payment by results², compared with 28 per cent of private sector employees (van Wanrooy et al, 2013). Where it was found, public sector PRP was mostly individually-based and concentrated in public administration. This reflects the fact that individual PRP schemes have been widespread across the civil service since the mid 1990s. More recently team bonuses have also been trialled among staff administering taxation and benefits. The Government has also recently committed to phasing out contractual ‘time-served’ pay progression (where it still exists) and replacing it with conditional progression dependent on performance, in the civil service, schools, the NHS, prisons and the police, with these services at varying stages of implementing the reforms.

The only large-scale example of a national pay-for-performance system in the UK public sector to date, however, is for general practitioners (GPs), who are not employees but self-employed contractors³. The Quality and Outcomes Framework (QOF) was introduced in 2004 and provides payments to GPs, representing up to 20 per cent of their income, for compliance with targets (called ‘indicators’) set across a spectrum of clinical activities (Coleman, 2010). The QOF has contributed to higher earnings for GPs and the General Medical Services contract (of which the QOF was a key element) has been costly⁴. Recent changes for the 2014-15 contract have significantly scaled back the number of QOF points available, diverting the savings to core funding of general practice instead. Figure 1.2 summarises the current state of play regarding pay-for-performance in UK public services.

² Defined as any method of payment determined by objective criteria, i.e. the amount done or its value, rather than just the number of hours worked. It includes commission, and bonuses that are determined by individual, workplace or organisation productivity or performance.

³ Most of the GPs working in the UK are independent contractors – self-employed people running their own practices as small businesses, usually in partnership with other GPs and sometimes others such as practice nurses or managers; some practices belong to sole practitioners and some to companies which employ salaried doctors to staff them (Review Body on Doctors’ and Dentists’ Remuneration, 2014).

⁴ The contract is estimated to have cost £8 billion in the first three years (Walker et al, 2010).

Figure 1.2 Elements of PRP in UK public services

Public service	Performance-related pay
Education	<p>In 2000 a form of PRP scheme was introduced for more experienced teachers in England. This comprised the introduction of an Upper Pay Scale with additional increments above the basic pay scale, with progression depending on performance (Atkinson et al, 2009).</p> <p>Since September 2013, schools in England and Wales have been given more freedom to link teachers' pay to performance, ending pay increases based on length of service. This is intended to "... support heads in attracting teachers in specific subjects based on their school's needs ... [and to] help schools across the country recruit and retain excellent teachers." (Department for Education, 2014)</p>
NHS	<p>All NHS employees, apart from doctors and dentists and very senior managers, are covered by the Agenda for Change pay and grading structure. This pay structure includes competency-based pay in the form of 'gateways' within each pay band, where staff have to demonstrate agreed competencies. From 2014, employers have greater flexibility to implement such gateways at any spine point (IDS, 2013).</p>
Doctors	<p>The Quality and Outcomes Framework (QOF) was introduced in 2004, rewarding GPs for achievement of quality indicators across a range of domains. Payments to practices are proportional to achievement, between a lower and an upper threshold. Financial incentives in the QOF are substantial. In 2011-12, a maximum of 1,000 points were available to practices per year, with an average payment of £130 for each point achieved (Gillam et al, 2012). The total number of QOF points was recently reduced to 559.</p> <p>Payment systems for hospital doctors are primarily salary-based, but there is a reward scheme, Clinical Excellence Awards, introduced in 2004, which rewards consultants for contributing to the delivery of safe and high quality care to patients, and for continuous improvement in the quality of services to patients and to the NHS (Review Body on Doctors' and Dentists' Remuneration, 2012).</p>
Civil service	<p>Some form of PRP has existed for most civil servants since the mid 1990s. The current system for the senior civil service links increases in base pay to performance via an appraisal system that ranks staff. Additional non-consolidated, performance-related payments have since 2010/11 been restricted to the top-performing 25% of staff.</p> <p>Pilot team-based PRP schemes have been trialled for frontline staff in Jobcentre Plus, the Child Support Agency and HM Customs and Excise (now HMRC).</p>
Police	<p>The Winsor Review (2011) recommended a stronger link between pay and skills and pay and performance for police officers, with annual pay increases limited to those who have performed satisfactorily or better. These changes are the subject of current negotiations.</p>

Our approach to the review

Our approach to the evidence review is built upon recognising the complexity in the topic of performance-related pay and the different perspectives that academic disciplines can bring to understanding it.

Previous reviews have highlighted limitations in the evidence base, and this is a conclusion that remains the case. Despite much discussion and case study examples of performance-related pay in public services, it is not a topic that is well-served in terms of robust impact evidence. The terminology of performance-related pay can also refer to a range of different things, meaning that the evidence is complex to navigate due to the many variations in scheme design. Different performance management design choices result in a range of variable features, including:

- *Whose performance is measured* - individuals, teams or organisations?
- *How is performance defined* – inputs (e.g. values, behaviours); outputs (e.g. targets, standards), or outcomes (e.g. objectives)?
- *Who measures performance* – employer / manager, peers or customers?
- *How is it measured* - subjective assessment or output figures?
- *What types of rewards are given* – bonuses, progression up a pay scale, or prizes & gifts?

It is also a topic area that engages interest and research across different academic disciplines including Economics, Psychology, Sociology and Management Studies. These disciplines may interrelate and complement one another, but they also bring contrasting perspectives and vocabularies. Our approach to the review has been to work across these disciplines and to consider a range of forms of evidence. This includes:

- Evidence grounded in the Labour Economics literature that explores performance-related pay through quantitative evaluation, testing the impact of financial incentives on a range of outcomes.
- Evidence from the Human Resources, Sociology of Management and Psychology literature, both quantitative and qualitative, that places more emphasis on explaining the mechanisms by which performance-related pay may operate, with a focus on organisational context, the accuracy and reliability of

performance measures and processes, non-financial motivations and group behaviour.

Review methods

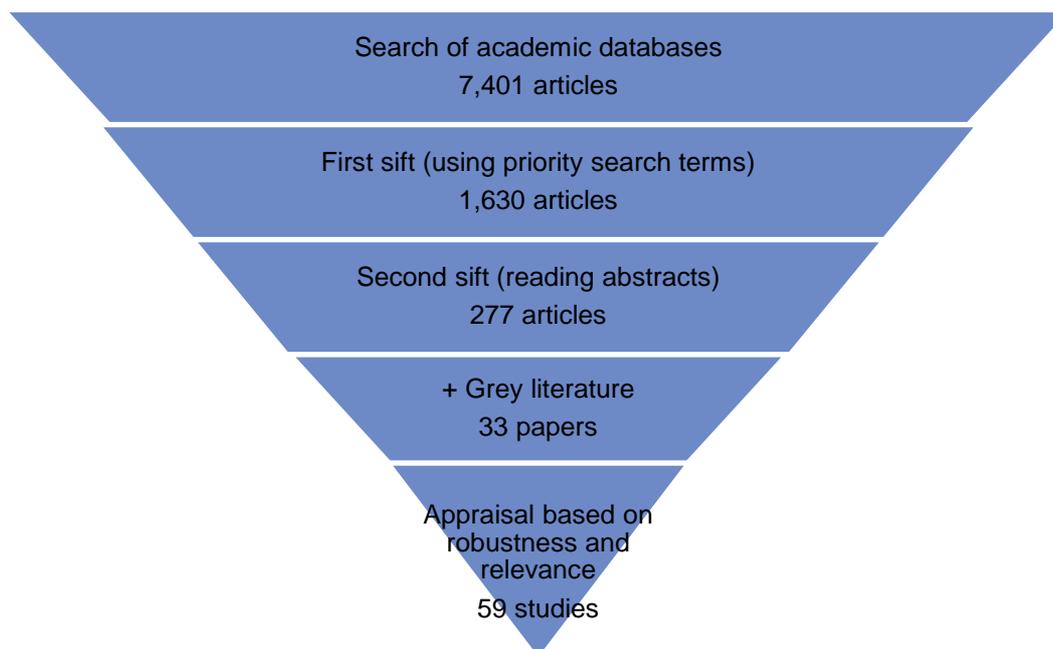
Our method has been to develop a rigorous and transparent approach to searching the relevant literature, using search criteria that balance the need to be comprehensive, with the practicality of ensuring the review was conducted within the time available. The key search terms and databases used to search the literature are presented in Figure 1.3. We also included 'grey literature' that is not published in academic media or is in the process of publication.

Figure 1.3: Our literature search terms and databases

Key search terms	Sources
<p>Set A: Populations</p> <p>public sector; public service*; public organisation*; government; non-profit; social services; teach*; civil serv*; doctor*; physician*; nurs*; health*; polic*; judicia*; prison*; military; armed forces; local authorit* NHS; dentist*; general practitioner*, GP*, consultant*, hospital*, school*</p>	<p>Academic databases</p> <ul style="list-style-type: none"> • Business Source Premier • ProQuest Business Databases • Academic Search Complete • Web of Science • EconLit <p>Grey literature</p> <ul style="list-style-type: none"> • Centre for Economic Performance (CEP) • CIPD • Institute for Employment Studies (IES) • National Bureau of Economic Research (NBER) • National Center for Performance Incentives (NCPI) • National Institute of Economic and Social Research (NIESR) • OECD • Policy Exchange • World Bank
<p>Set B: Interventions</p> <p>Performance related pay; pay for performance; performance based pay; performance based wages; performance pay; performance targets; performance management; performance measure*; merit pay; bonuses; wage*; pay; compensation; reward*; performance incentives; financial incentives; payment by results; contribution; team-based; variable pay; appraisal-related pay; incentive pay</p>	
<p>Set C: Study type</p> <p>Evidence; evaluat*; empirical; experiment*; impact; effect*; outcome*; assess*; case stud*; value for money; cost effectiveness; implementation; review; experience; study; survey</p>	
<p>Set D: Outcomes</p> <p>performance; improve*; motivat*; job satisfaction; quality; service delivery; effectiveness; achievement*; productivity; output; efficien*; behaviour; effort; morale; discretion*; input; outcome*; recruitment; retention; turnover; divisive; discriminat*</p>	

This initial search identified over 7,400 documents and articles. The next stage was to sift and appraise these references to form a manageable shortlist of studies to include in the review. This involved an extensive sifting process, as shown in Figure 1.4. The initial sift, reviewing titles and abstracts, reduced the scale of the review to under 300 articles⁵; a more detailed review of content reduced this further to 59 articles that were included in the review.

Figure 1.4: Our approach to filtering down the evidence



The appraisal process focused on two key factors:

- **Relevance to the review questions.** This involved judgements on the sectors and occupations covered by studies, the types of performance-related pay scheme in question, the geographical focus of the study, and its aims.
- **The methods employed.** Where studies were impact assessments, we applied the Maryland Scale for evaluations⁶, which scores studies according to a five-

⁵ This long-list is available from the authors on request.

⁶ This is a five-point scale designed by the University of Maryland to classify the strength of evidence. It is used as the basis of a toolkit published by the Government Social Research Service, which allows users to assess research evidence, see <http://www.civilservice.gov.uk/networks/gsr/resources-and-guidance/rapid-evidence-assessment/how-to-do-a-rea>

point scale, based on their use of comparison groups. Other types of studies (reviews and process studies) were appraised based on the rigour of their methods and their relevance to the review questions.

Our short-list of evidence

Our searching, screening, sifting and appraisal process reduced the volume of evidence from over 7,000 articles to a shortlist of 59. This in itself provides some insight into the current state of evidence on performance-related pay in public services:

- There is no shortage of published material, but much of it is not relevant for this review; for example of the initial range of articles much was general commentary about the issue rather than new evidence.
- It is often difficult to filter out what is fully relevant. Much research does not disentangle performance-related pay from wider issues of performance management, promotion and progression and organisational change.
- There is limited evidence from within UK, with a high share of the available research material from the United States.
- There are few experimental studies, which means there is little evidence that scores highly for robustness when using the Maryland Scale.
- There are some helpful existing evidence reviews, especially in health.

Among the 59 studies that we shortlisted, 27 were in health, 16 in education and 16 in the civil service (including both local and central government), as shown in Figure 1.5. This reflects the range and scope of the evidence available. Most recent studies are in health services, with fewer in education and public administration. Robust impact studies in the civil service were particularly sparse. The review found no new evidence of applications of performance-related pay that is relevant to the armed forces, policing, or prisons and justice. Full details of the included studies are provided in Appendix A.

Figure 1.5: Studies included in the review⁷

	Impact studies	Review studies	Process studies
Health	13	9	5
Education	10	5	1
Civil service	1	8	7

Report outline

Chapter 2 of the report outlines the conceptual underpinnings of PRP and highlights some of the challenges around its use in the public sector.

Chapter 3 provides an overview of findings from the earlier review of PRP in the public sector (Prentice et al, 2007) and discusses how the evidence base has evolved since then.

Chapter 4 presents the findings from this review, organised thematically.

Chapter 5 presents the findings on PRP scheme design.

Chapter 6 concludes by summarising the evidence on the effectiveness of PRP in the public sector and the key challenges to its design and implementation, as well as identifying potential ways of addressing these challenges.

⁷ *'Impact studies'* are defined as those employing robust quantitative methods to assess the impact of PRP on one or more outcomes. We included only studies scoring 3 or above on the Maryland scale. *'Review studies'* are defined as studies based on a comprehensive or systematic review of existing quantitative and qualitative evidence. *'Process studies'* are defined as qualitative or quantitative research exploring the mechanisms and implementation of PRP schemes in practice.

2. Conceptual Issues: PRP and the public sector

PRP has long been an important idea for economists. At its heart is the principal-agent problem, which defines the employment relationship and the role of pay in this. In short, it is argued that the employment relationship does not naturally align the interests of principal (employer) and agent (employee). The principal's interest is that the worker works as hard, efficiently and responsively as possible, in order to maximise profit and improve outcomes. The agent, on the other hand, will aim to minimise the effort they expend in order to reach the basic standards required to receive pay. From this perspective, PRP is a mechanism designed to better align the interests of principal and agent. The agent is incentivised to increase their outputs – and consequently organisational outcomes – by tying a proportion of their remuneration to their performance. In theory, PRP will therefore encourage effort and offset the 'shirking' behaviour which might otherwise be expected from employees (Goodman and Turner, 2009). It will also offer the employer cost flexibility, allowing them to reduce the pay bill in cases of poor performance.

It is important to outline the assumptions which underpin this logic, which are based on expectancy theory and reinforcement theory. The logic of expectancy theory is that expenditure of an individual's effort will be determined by their expectations that an outcome will be attained, and the degree of value placed on this outcome by the individual (Porter and Lawler, 1968). Reinforcement then operates when the intended effort elicits the desired outcome. In the case of PRP, employees will be expected to work harder if they value monetary rewards and believe that these will result from their increased efforts.

This idea was enshrined in Taylor's classic 'piece rate' system, which offered different wage rates to workers depending on their productivity (Reilly, 2003). If workers are primarily motivated by the money they earn, this is viewed as the best incentive to ensure they maximise their efforts. However, the system also has to consider issues of equity; that is rewards need to be perceived as reflective of employees' efforts. If the reward is either too small, or something the employee does not consider important, then it will fail to motivate the employee to increase their efforts. Moreover, if an employee does not feel they stand a chance of winning a reward through increased effort – either because they already feel they are working to capacity, because they do not know how to improve their performance, or because they realise they are unable to out-perform their colleagues – they may actually choose to *reduce* their efforts, 'cruising' rather than competing.

A further question is whether financial rewards should be offered to individuals or to teams. While offering incentives solely based on individual effort and performance may be simpler, it ignores the fact that work can be collaborative and the results dependent on the input of several employees. Where this applies, determining the effort attributable to individuals may be inaccurate and team-based incentives will be preferable for reasons of fairness. However the drawback is that team-based bonuses raise the possibility of free-riding, whereby certain individuals within a team might reduce their effort, hoping to gain from the input of colleagues.

Finally, in several applied studies, the underlying foundations of the standard economic theories of PRP have been called into question. For example, in a notable study on timber workers in North America, Locke and Latham (1990) found that the employees involved could also be motivated by the setting of clear goals and feedback on performance, without any financial reward being provided. It was argued that this was due to the fact that the workers both trusted their managers and viewed their goals as reasonable. In cases where PRP has not operated along the lines the theoretical logic would suggest, some studies have suggested that financial motivations may not be of particular importance to the employees, or that PRP has created divisive effects between co-workers (Perry et al, 2009). Conflicting evidence has thus created significant debate around PRP and its potential effectiveness.

PRP and the public sector

For some time the use of PRP was largely restricted to the private sector, where it was demonstrated to be effective in roles where individual effort and contribution are more easily measurable, such as sales or production. For example, the classic study by Lazear (2000) showed the effectiveness of piece rates – the simplest form of PRP – in raising individual worker output in the Safelight Glass Corporation during the 1990s. However, the rise of New Public Management theories during the 1980s opened a new discussion about its potential applicability in the public sector. These theories were based on the idea that performance in the public sector was being held back due to an absence of the incentives and management practices that drove organisations working for profit. The way to improve public sector performance was thus to simulate some of these features of private sector management practice.

The economic theories outlined above have been developed and applied to common concerns for public sector managers. For example, existing ways of setting public sector pay rates, based on experience or qualification levels, may not always identify and reward the best performers. In education, for example, while

teacher effectiveness plays a significant role in pupil performance, the characteristics frequently used to set teachers' pay rates, such as experience or acquisition of higher qualifications, have only a low power in predicting teacher effectiveness (Goodman and Turner, 2009; Leigh, 2013; Woessman, 2011). It is argued that this creates limited direct incentives for teachers to improve their performance. This is also compounded by the fact that the nature of many public sector contracts makes it difficult to remove poor performers (Figlio and Kenny, 2007). From this perspective, PRP is seen to be useful both for its motivation and selection effects. If effective teachers are hard to identify using recognisable characteristics, offering an element of PRP should better attract and retain those who perform well in post (Leigh, 2013). It may also induce a 'clearing effect', by which poor performers are encouraged to leave the profession.

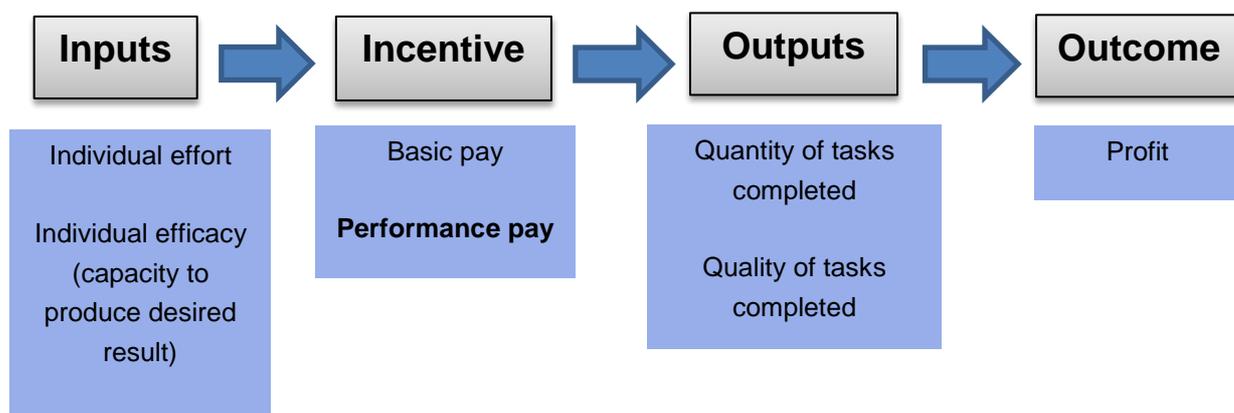
There are other issues affecting the public sector which PRP may also help to address. For example, the scale of public sector organisations and the nature of their work, means that many employees operate in a setting in which the majority of their actions and inputs are not directly visible to their managers (Neal, 2011). Its use can therefore lead to an increase in oversight and accountability. PRP schemes can also steer and direct the focus of employees, which may help to clarify to employees how best to fulfil their responsibilities where a job involves numerous tasks or its nature is broadly defined (Goodman and Turner, 2009).

PRP has therefore been seen as an answer to what New Public Management theory would view as one of the fundamental challenges facing the public sector: that it lacks the clear links between performance and pay that can be used to drive improvements in profit-seeking organisations. Below, we examine these links in more detail.

PRP mechanisms

Theories of PRP suggest a number of mechanisms by which it is expected to operate. Under the most simple logic chain based on economic foundations, PRP aims to strengthen the link between the employee input of effort and the incentives they receive in terms of pay. This should motivate the individual to expend more effort and increase their outputs, which should ultimately influence the organisation's outcomes and goals – in the private sector, their profits. This simple sequence is set out below in Figure 2.1.

Figure 2.1: PRP simple logic chain based on economic foundations



Clarifying these mechanisms allows us to better assess the assumptions behind them. For the theory to be effective in practice, we must be confident that these assumptions hold for the public sector. Applying the simple economic logic chain suggests its application in the public sector would be based on the mechanisms and assumptions shown in Figure 2.2, below.

Figure 2.2. PRP in the public sector: mechanisms and assumptions



As economic and organisational theories have developed over recent decades, however, many of these assumptions have been called into question. The development of other strands of economics, notably behavioural economics, as well as HR and management theories, have suggested that the logic behind PRP cannot be easily applied to all organisations – and that these challenges may be particularly marked in the public sector. Some of the most significant theoretical challenges posed to the simple economic logic chain and its underlying assumptions are set out below.

Challenges to PRP in the public sector

While the basic economic theory behind PRP has been applied to many public sector challenges around performance and value for money, this theory is by no means uncontentious. Challenges have emerged from both the economics and HR disciplines, relating to wider issues of organisational practice and behaviour, as well as specific features of the public sector.

A number of theoretical critiques have centred on the idea that the economic logic behind PRP cannot fully capture the range of factors which motivate an employee. Peer effects, perceptions of fairness, and intrinsic motivation will all affect an employee's performance – and pay schemes will necessarily interact with these factors. Disregarding them may therefore subvert desired outcomes or lead to unintended consequences. For example, for a PRP scheme to be effective, employees must accept that their performance is being measured in an accurate and fair way. If an outcome cannot be easily measured, cannot be attributed to a particular individual, or where a degree of subjectivity is required in assessment, perceptions of unfairness may interfere with the motivational effects intended by PRP schemes.

The idea and significance of 'intrinsic' motivation was highlighted by Richard Titmuss in the 1970s. His most famous example was that of blood donation. Comparing the voluntary system for donation in the UK with the financial incentives offered for donation in the US, Titmuss found not only a marked difference between those prepared to donate blood under each system, but also posited the idea that attaching financial reward to an action may 'crowd out' the intrinsic motivations (i.e. the positive feeling of contributing to a favoured cause) that would otherwise drive it (Titmuss, 1997). If employees draw intrinsic motivation from their jobs, then factors such as job design, work environment and measures conducive to increased satisfaction will induce motivation and may be more cost-effective than direct

financial reward (Herzberg et al, 1959). There may also be a risk that financial rewards detract from, or ‘crowd out’, such intrinsic motivation, again subverting the intentions of a PRP scheme.

It has also been argued that these theoretical challenges to the underlying logic of PRP may be particularly applicable in the public sector. For example, the idea of ‘intrinsic’ motivation may have more applicability in this area than most. In the UK, while public sector schools and hospitals now compete in a quasi-market, with league tables and user choice, public services generally lack a clear profit incentive and focus primarily on the social benefits of the services they provide. A number of studies highlight the existence of some form of ‘public service motivation’ among public sector employees. For example, employees in public services may derive much of their motivation from their belief in the intrinsic value of the service they perform, with less regard for financial reward; they may be motivated by commitment to a professional code of behaviour; and may be focused on the collaborative, rather than competitive, nature of their relationships with colleagues (Anderfuhren-Biget et al., 2010; Ashraf et al, forthcoming; Leigh, 2013; OECD, 2009; Taylor and Taylor, 2011).

Other challenges to PRP in the public sector have focused on the nature of the public sector’s activities and objectives, arguing that this may create challenges for using performance measurement as a basis for pay. These unique features of the public sector include:

- **The complexity of the public service ‘good’:** By their nature, public services generate a multitude of outcomes, and thus their objectives are not clear-cut. In education, for example, these may encompass examination results, attendance, engagement in learning, citizenship, creativity, vocational skills, etc. In health, this could include the number of procedures performed, mortality rates, the patient experience, equity of health outcomes or broader measures of wellbeing or quality of life. A PRP scheme in the public sector has to identify and prioritise these various objectives. In addition, the ultimate outcomes of many activities performed in the public sector – educational performance, trends in crime rates, or health outcomes – may only be visible in the long-term, raising questions about the feasibility of accurate and meaningful performance measures within a PRP scheme.
- **Multiple principals:** Unlike in the simple ‘firm’ model, in which the desired outcome is profit and the main principal is the employer, the public sector involves a much wider variety of potential ‘owners’ and stakeholders. These include individuals using the service, their advocates, public sector managers, unions, professional bodies, the Government, and taxpayers. All could claim an important stake in public services, but their interests and

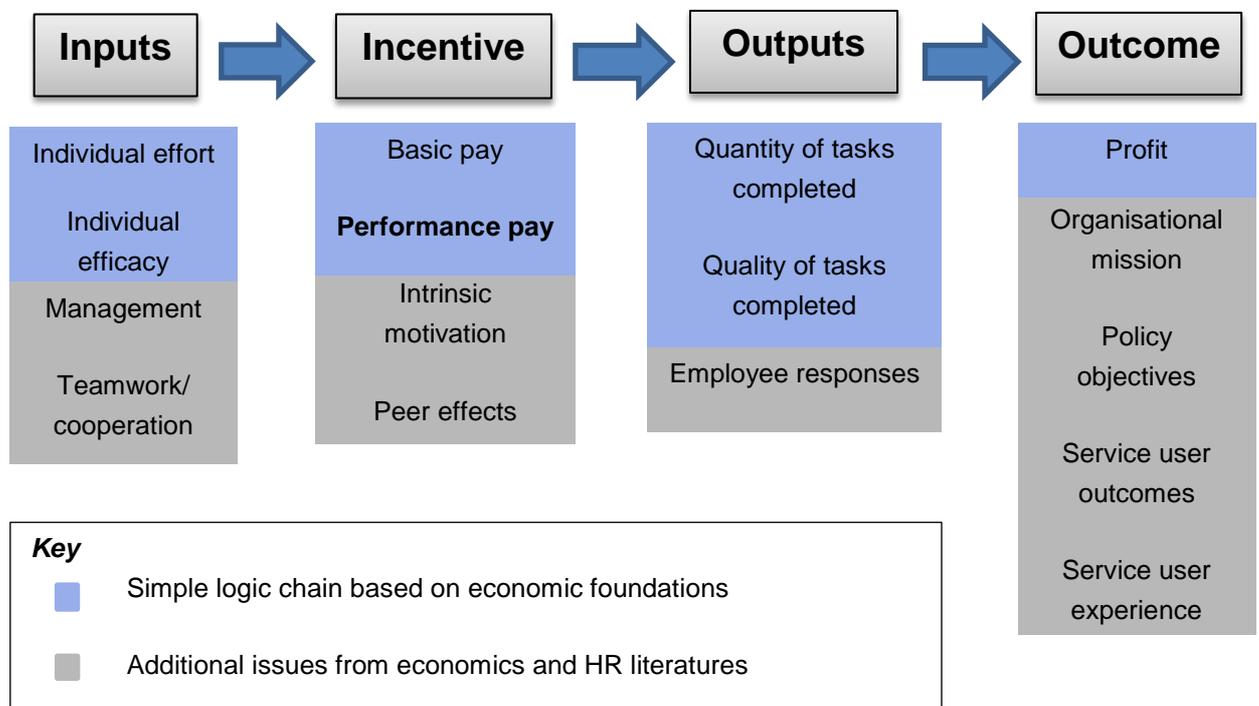
objectives may well differ. Moreover, the objectives of these stakeholders may also change over time – for example, with changes in government policy (OECD, 2005; Burgess et al, 2010). Any PRP scheme in the public sector must be capable of reconciling the variety of outcomes that emerge from these multiple stakeholders and interests.

- **Multi-task problems and collaborative activity:** The delivery of public services tends to be a complex and inherently collaborative activity, based on the combined contributions of a large number of individuals (e.g. in health this could include GPs, doctors within hospitals, nurses, community health workers, administrators, and many others). This means that inputs and outputs may be difficult to observe and attributing individual responsibility for performance may be challenging. There is also a risk that individual incentives may mitigate against this necessary teamwork. Furthermore, the outcome of public sector activity will depend heavily on factors and relationships outside of the workplace, e.g. teachers have raised concerns about the impact of factors outside their control on a child's education and achievement, including parental input, home life and family difficulties (Leigh, 2013). If employees do not feel they are capable of improving performance through their own efforts, PRP may in fact serve as a disincentive, with employees choosing to expend minimal effort, rather than to compete.
- **Misallocation of effort:** Another related challenge is the risk that PRP schemes will incentivise outcomes which are more easily and directly measurable (OECD, 2009). Prioritising these outcomes over those which are less clear-cut is likely to encourage employees to focus on these outcomes at the expense of others. This is likely to be a particular concern in the public sector given complex, multi-task environments where outputs are multiple and varied. In education, for example, this may fuel existing concerns about the rise of a narrow, 'teaching to the test' approach (Neal, 2011: 45), and in health, concerns about less patient-centred, 'protocol-driven care' (Gillam et al, 2012).
- **The risks of gaming or cheating:** Related to misallocation of effort, another widely-recognised potential pitfall of PRP is attempts to 'game the system' (Neal, 2011), where workers seek to maximize their incentive gains while minimizing effort or without increasing actual performance. This can lead to significant problems in the public sector, where outcomes can have a wide social impact and governments are expected to be accountable both for performance and for the use of taxpayers' money.
- **How to define measures that result in the desired outcomes:** Linked to the problem of competing outcomes, PRP schemes in the public sector must also seek to construct measures that direct staff towards prioritising the desired outcomes. For example, 'value-added' or relative measures may serve to boost the performance of the lowest attainers, while threshold

measures may encourage improvement amongst those at the median, and rank order tournaments can encourage improvement among top performers (Eijkenaar et al 2013; Elovainio, 2010; Lavy, 2009).

These issues, emerging from the economics and HR literature, ask significant questions of the simple economic logic underpinning PRP when applied to the public sector. This does not mean that the economic logic is discounted, rather the literature points to a number of additional issues that must be considered if PRP is to work effectively in the public sector. This more complex range of considerations is summarised in Figure 2.3, below.

Figure 2.3: PRP in the public sector: logic chain with additional considerations



This chapter has provided a brief overview of the theoretical concepts underlying the use of PRP in the public sector. It is now important to consider whether and to what extent these hypotheses are confirmed by the recent evidence in this area. In the following chapter, we summarise the evidence from the previous review in 2007 (Prentice et al, 2007) and consider how the evidence base has evolved since then. Then in Chapters 4 and 5 we present the findings from the more recent evidence collated for this review.

3. Previous findings and the evolving evidence base

In this chapter, we provide a summary of the 2007 findings and an overview of the more recent evidence base, including its limitations and gaps. The findings from these studies post-2007 will be discussed in detail in Chapters 4 and 5.

Scope of the two reviews

The 2007 Review (Prentice et al, 2007) presented evidence on PRP in three public services: the civil service, healthcare workers and teachers, and identified as gaps in the evidence base literature on the effects of PRP in the police, armed forces and prisons. In 2014, this characterisation of the literature remains valid. Practically all the literature retrieved relates to PRP in these three public services and we found no new evidence around performance-related pay in the armed forces, policing, or prisons and justice. However while the evidence included in the current review is in the same three public service areas, its scope is wider, since the 2007 Review focused exclusively on Economics-based impact studies of PRP, while the remit for this review is wider, covering relevant literature in Human Resources, Management Sociology and Psychology, in addition to that from Economics.

Civil Service evidence

The 2007 Review concluded that there is strong evidence that civil servants respond to financial incentives, even when there is no direct salary enhancement from benefits. All evaluation studies were on incentive schemes that rewarded staff at the group or team level.

- Among the strongest evidence of the power of financial incentives in the civil service was the study from Khan et al (2001), which examined the PRP scheme within the Brazilian Tax Collection Authority. The bonus scheme here was very generous, however, with payments reaching twice the mean annual salary, making this case study an outlier, as performance payments in most public sector PRP schemes are a small percentage of the base salary, especially among non-managerial employees (OECD, 2005).
- In a study of a UK pilot of team-based PRP in HM Customs and Excise, it was found that the pilot was effective in motivating additional employee effort, compared to teams not subject to PRP. However the highest achievement appeared to be the result of more efficient allocation of tasks by team managers in addition to the increase in individual effort (Burgess et al., 2007).

- Another UK-based evaluation of a PRP scheme in Jobcentre Plus found no overall effect but this was because positive effects in smaller teams were counter-balanced by negative effects in larger teams. The authors concluded that team incentives might be more effective in smaller groups where peer monitoring is present to reduce the extent of ‘free-riding’ (Burgess et al., 2004).
- The Jobcentre Plus study also found that the scheme resulted in improvements in the quantity but not the quality of services offered, leading the authors to conclude that measurement accuracy has an important impact on workers’ strategic behaviour. The latter may be more likely when performance is measured inaccurately, or when subjective assessment of performance is required (Burgess et al., 2004).
- There was evidence of strategic/gaming behaviour in studies of a US-based incentive programme, the US Job Training Partnership Act (JTPA), whereby workers improved measures of performance without increasing actual performance (Courty and Marschke, 2004). In this instance, bonuses were awarded based on annual data, so workers engaged in strategic reporting by spreading good and bad performance over multiple years.

Limitations and gaps in the literature identified included that the evaluated programmes were not in place for a long enough period of time to confirm the long-term effects of PRP, and that value-for-money was not considered in most of the studies reviewed, although one paper (Courty and Marschke, 2004) identified that strategic behaviour was imposing costs on society, thus questioning the wider welfare implications of financial incentives for civil servants.

The more recent evidence base in the civil service, post-2007, is relatively large but there are few additional robust impact evaluations. Rather the majority of the literature on PRP in the civil service in this period has been reviews of the earlier evidence or quantitative studies that relate not to the impact of PRP per se, but to the interaction of PRP with public service motivation, and the detrimental effects that can result from PRP implementation, such as perceived unfairness and diminished effort. This is partly a consequence of widening the scope of the current review to consider HR and management literature. Much of this literature questions the economic foundations of PRP by examining the salience of intrinsic motivations and organisational factors in motivating individual employee behaviour.

Healthcare Evidence

The 2007 Review found that improving the quality of health outcomes, rather than productivity per se (i.e. quantity of output) is the primary focus of performance pay schemes in the healthcare sector. Overall, the evidence on whether doctors respond to financial incentives was inconclusive, partly because the evidence was

from incentive schemes only in place for a limited period and because financial rewards were small.

- There was some evidence from the US that clinicians' responses to incentives depend upon the absolute and relative size of the financial incentives as well as the relative effort required to obtain them.
- Evidence from the UK suggested that GPs have been very successful in obtaining financial bonuses under the current performance pay scheme (the QOF), but as this is against a backdrop of previously high levels of performance, the additionality of the QOF scheme is unclear.
- UK evidence suggests that doctors may deliberately misreport performance measures to increase their financial rewards (gaming).
- However UK studies also suggested that GPs are motivated by concerns for patient health as well as financial incentives.
- Evidence on cost effectiveness of schemes was scarce.

Since 2007, the literature on the effects of PRP in healthcare has burgeoned. Especially helpful for this review have been a number of systematic reviews of various aspects of PRP in healthcare, examining, for example, the impact on patient outcomes, healthcare workers' behaviour and the cost-effectiveness of PRP systems. Several reviews set out to examine which aspects of the design of incentives or payment models work best. Most of the reviews focused on US and UK examples, and were primarily focused on primary care, with a large literature on the impact of the QOF in the UK. Studies of PRP in acute care (hospitals) and nursing homes were principally US-based.

Education evidence

The 2007 Review found strong evidence that teachers do respond to financial incentives, although most of the improvement appeared to come from previously weak students performing better under such schemes. All studies suggested that directly rewarded outcomes improved under school and teacher-level incentive schemes, but the evidence on unrewarded outcomes was inconclusive.

- A few studies were used to draw out pointers on the effects of different types of schemes. One found that, teacher-based merit pay schemes tended to be more effective if rewards were distributed selectively. Another compared the impact of group versus individual-level schemes, finding that individual-level schemes were more cost effective.
- Nearly all studies evaluated a sample of schools which were unrepresentative or specially selected to take part. This means that care must be taken in applying the results of studies more generally.

- There was little understanding of the processes of change within schools that have adopted a financial incentive scheme, and the subsequent effects on staff morale. Neither was there an understanding of the full costs and benefits of these schemes.

Post-2007, the literature on PRP in education has grown, although it is still smaller than that in health. The geographical focus of recent literature is predominantly the United States, but there are also studies from a range of other countries. Most of the impact studies focused on the effect of teacher pay-for-performance programmes on pupils' scores in assessment. However, several other angles are also explored. Some studies assessed the impact of voluntary versus mandatory performance pay programmes for teachers, or group-based versus individual incentives. Several also focused on their impact on teachers' job satisfaction and motivations. In the USA, this was often based on results from the Schools and Staffing Survey (SASS), which allowed analysis of teachers' working practices and self-reported levels of satisfaction. There were also several more general reviews of the empirical evidence base for PRP in education.

Remaining limitations of the evidence base

Occupational coverage: Similar to the 2007 Review, the bulk of recent evidence on the effects of PRP in the public sector centres around education, health and the civil service, leaving wide evidence gaps on PRP effects within the armed forces, prisons, the police and the judiciary. In addition, the evidence for health is primarily focused on doctors and in the UK centres primarily on the effectiveness of the QOF for GPs. The effects of PRP on nurses, consultants and dentists is limited, and there is no evidence of PRP's effects on low-earning employees in the NHS.

Timescales of initiatives and evaluations: Another key limitation that emerges from both the earlier and the more recent literature review relates to the timings and timescales of both PRP schemes and their evaluations. Some incentive schemes are either relatively new or were implemented for only a short period of time. Whether or not PRP system designs would retain any positive effects in the longer-run remains to be seen, as there could potentially be 'Hawthorne Effects', whereby modified behaviour is in response to the knowledge that one is being observed or studied, rather than in response to a treatment.

The time period of evaluations is also something important to consider. Often, studies can not comment on the long-term effectiveness of PRP schemes because the data does not cover a long enough period of time. For example, Lavy's (2009) study of the implementation of a PRP scheme in Israel surveyed teachers before the

first bonuses were issued. Two-thirds believed they would be among the winners and the scheme was strongly associated with increased teacher effort and positive teacher attitudes. However, a study on the implementation of PRP in the US, which included data over a longer timescale, found that the impacts of PRP were uneven, with teachers who had performed less well actually reducing their work hours and reporting decreased levels of enthusiasm over time (Jones, 2013). This suggests that employees who come to realise their chances of earning a bonus are low may begin to reduce their efforts accordingly – an effect not captured in results from early evaluations.

Robust evidence on scheme design: An important caveat discussed in many PRP impact studies is whether or not the findings of the particular study can be extrapolated to other contexts and/or organisations. Reviews of the evidence suggest that the effects of PRP initiatives vary by several factors including the organisational setting, the outcome measures used, and the level at which incentives are distributed. However, heterogeneity between studies (in terms of both inclusion criteria and focus) have inhibited meta-analysis, making it difficult to draw firm conclusions overall about which types of schemes/contexts are successful (Eikeaneer et al, 2013; Flodgren et al, 2011; Houle et al, 2012; Scott, 2009).

Cost effectiveness of performance-related pay: There are no robust cost-benefit analyses of PRP in education or the civil service and that available in health is scarce and inconclusive. This is likely to be due to the fact that accurately measuring and quantifying the various financial effects of public sector outcomes is notoriously difficult and contentious. There is a need for further robust cost-effectiveness research and for comparative cost-effectiveness research to assess the relative merits of PRP vis-à-vis other types of incentives (Eijkenaar et al, 2013).

Employee retention and types of employees: The relationship between employee retention, employee skills profiles and different incentive schemes (material and non-financial) is under-researched. Depending on how it affects staff turnover and the types of workers it attracts, PRP could have long-term implications for the overall success of a public sector objective, organisation or department.

Job satisfaction, PRP and productivity: There is evidence on the link between job satisfaction, public service motivation and PRP, but the impact of job satisfaction and PRP on employee effort and performance levels requires closer investigation. Some studies have shown that PRP does not cause crowding-out effects – i.e. negatively impacting on intrinsic motivation - but these utilise cross-sectional data which is less desirable than panel or time-series data, needed to determine causality (Stazyk, 2013). Job satisfaction and its relationship to PRP is likely to be

an important determinant of a successful organisation since past research demonstrates that job satisfaction is a reliable predictor of employee turnover (Stazyk, 2013).

Difficult to measure criteria, outputs and outcomes: There is a danger that both PRP schemes and their evaluations display a tendency to focus on the criteria or targets that are most easily measured, rather than on those deemed most significant or desirable. As noted previously, this is particularly problematic in the public sector given complex, multi-task environments where outputs are multiple and varied. For instance, there is little exploration of wider pupil outcomes beyond those captured and rewarded under PRP, such as test performance, in the education literature. None of the studies explored the reactions of pupils under PRP, in terms of their levels of motivation and engagement, for example. A 2012 World Bank review (Hasnain et al, 2012) also confirms that more rigorous studies linking PRP to positive performance tend to be for jobs with outputs, outcomes and effort that are more readily observable, thus limiting the conclusions that can be drawn from the data for other occupational contexts.

Fairness and underperformance: Potential issues with perceived fairness in PRP schemes in the public sector needs further investigation to determine the efficacy of incentive schemes. The link between fairness, reciprocity and performance was confirmed in a survey of the behaviour of American police officers (Bregm 2013: 31), whose arrest rates and average sentence length declined and crime reports rose, when they lost an arbitration case on wages. There is also evidence in psychology and from organisational studies that show the adverse effects of perceived unfairness on employee motivation. However, the actual interaction of PRP, equity and underperformance is an under-researched area.

4. Findings: Recent evidence on the effectiveness of PRP in the public sector

The effects of PRP on incentivised outcomes

In the recent literature (post-2007), there is some evidence of positive effects from PRP schemes across the three public services for which there is evidence available: education, health and the civil service. However, the overall conclusion is that findings remain mixed. Positive findings are often context- or outcome-specific, making it difficult to draw overall conclusions about the effectiveness of PRP for a particular service.

In education, there is widespread evidence that the introduction of PRP can result in positive impacts on students' performance in standardised tests - which in most cases are the outcome directly incentivised by the scheme. Reviews of PRP in education, such as Podgursky (2007) and OECD (2009), suggest that the weight of evidence is in favour of the effectiveness of PRP in improving students' test scores, although the evidence is not completely consistent and some individual programmes have reported no impact of a PRP scheme on these outcomes.

The findings of positive impacts on test scores are reflected in Woessman's (2011) analysis of the results of international PISA tests, which found that students in countries where teachers' salaries were adjusted for outstanding performance scored around 25 per cent of a standard deviation higher on maths tests, after controlling for student, school and country measures. Positive effects are also found in several individual impact studies of PRP schemes. For example, Figlio and Kenny (2007) found a positive association between the use of individual teacher incentives and pupil achievement⁸.

Several studies in education also show positive impacts of PRP on 'value-added', rather than simply absolute measures of performance. Atkinson et al (2009), for example, found that the introduction of a PRP scheme for teachers in England in 1999⁹ added on average nearly 90 per cent of a GCSE grade and 40 per cent of a

⁸ although given the scheme they evaluate was a voluntary one, selection bias affecting the results cannot be ruled out

⁹ The scheme included a bonus on passing a Performance Threshold, plus the introduction of an Upper Pay Scale in which progression depended on performance.

value added grade per child for eligible teachers. Springer et al (2010) also found that the 'District Awards for Teacher Excellence' (DATE) programme in Texas resulted in participating schools with higher levels of educational disadvantage catching up with their non-participating, but more advantaged counterparts.

However, it should be noted that several single-programme impact studies reported no significant effects of the use of PRP on pupil's examination performance (e.g. Goodman and Turner, 2009; Springer et al, 2012). It cannot therefore be concluded that PRP will consistently lead to improvements in the directly incentivised outcome – much depends on the design of the scheme, teachers' support, and how it fits with wider accountability and performance measures.

In the health field, there is also evidence of positive impacts of PRP, although again findings are mixed. In a synthesis of findings from systematic reviews, Eijkenaar et al (2013) report that findings vary but tend to be less positive for experimental (judged as higher quality) than for observational studies. A systematic review by Van Herck et al (2010) put the average effect of PRP on physician incentives to be of the order of a 5 per cent improvement in the incentivised performance measures, although there was a large amount of variation across studies.

In general, positive outcomes have been found most frequently for quality of care measures for chronic diseases, and for preventative health measures such as immunisations. For example, a systematic review of the effects of the Quality and Outcomes Framework (QOF) in the UK showed that it had resulted in modest improvements in the quality of care for chronic diseases, notably diabetes and asthma. Similarly, a large-scale study of the impact of performance-based incentives targeting a range of primary care services in Ontario, Canada, found that the incentives led to an increase in the provision of four of the five preventive care services targeted (Hurley et al, 2011). However the study also found that there was no impact on other additional services targeted by the scheme, which appeared to be related to the higher cost of responding to these incentives (Li et al, 2011)¹⁰.

However, while positive effects of PRP have been widely found on quality of care measures, the evidence also suggests that improvements as a result of PRP are often small and sometimes short lived (So and Wright, 2012; Houle et al 2012). This has been attributed both to the modest size of payments in many schemes (Chung et al, 2010; Greene et al, 2013) as well as to previous high levels of achievement resulting in 'ceiling effects', i.e. little room for further improvement

¹⁰ These services, based outside the practice, often incurred financial, as well as time and organisational costs

(Greene et al, 2013; Petersen et al, 2013; Van Herck et al, 2010). In the UK, for example, the QOF improved the quality of care for almost all incentivized conditions during the first year of the framework (2004-5), but improvements slowed for a number of conditions after the first year and subsequently returned to prior rates of improvement (Gillam et al, 2012). This may have been due to generally high levels of achievement and improving trends at the time the scheme was implemented (Kontopantelis et al, 2012).

Moreover, not all national performance-related pay schemes targeting primary care have been successful. In Australia, for example, a national PRP scheme for GPs introduced in 2001, the Practices Incentive Programme Performance, was associated with some short-term gains in diabetes testing and cervical cancer screening but this seems to have been a result of an improving national trend rather than specifically to the impact of the programme. This may have been because it was a voluntary scheme, and those GPs that participated had higher performance to begin with, thus there was limited room for improvement (Greene et al, 2013).

In acute care (hospitals), there is a less extensive evidence base on PRP and results tend to be, overall, less positive (Eijkenaar, 2013; Van Herck et al, 2010). However robust evidence from a large, nationwide performance-related pay scheme for hospitals in the US, the Premier Hospital Quality Incentive Demonstration, did show an improvement of 2-4 percentage points on quality measures in participating hospitals as a result of the programme (Eijkenaar et al, 2013). However, again these improvements were significantly reduced in the second phase of the scheme. It is unclear if this was due to ceiling effects, to enthusiasm for the scheme wearing off or to the effects of the more complex payment model in phase 2 (Ryan et al, 2012). There has been limited experience with PRP in UK hospitals, but one pilot scheme, Advancing Quality, implemented in hospitals in the northwest of England in 2008, showed a significant reduction in mortality rates as a result of the scheme¹¹. Mortality rates reduced by 1.3 percentage points more than non-participating hospitals (equating to a 6 per cent reduction or equivalent to 890 fewer deaths) (Sutton et al, 2012).

In the civil service, there is also some evidence that PRP can have a positive impact on incentivised outcomes (Ashraf et al, forthcoming; Burgess et al., 2010), although again findings are mixed, and much of the literature in this area tends to emphasise

¹¹ The authors note that other quality improvement strategies were also introduced at the same time as the PRP scheme (including specialist nurses, new or improved data collection systems linked to regular performance feedback, and face-to-face meetings with other hospitals in the programme to develop shared learning). Thus a combined effect of PRP plus other strategies cannot be ruled out.

the role of other factors in affecting outcomes, such as intrinsic motivation and team and management structures. In one of the few studies of PRP in the UK civil service, Burgess et al (2010) evaluated a pilot programme in the UK tax office¹², which was introduced in 2002 and ran for nine months. Using a triple difference design, the evaluation showed that the PRP scheme, which used team-based targets, had a positive effect on both individual and team performance. A similar evaluation of team-based PRP in the UK's public employment service (Jobcentre Plus), however, found no impact on overall performance – but this appeared to be because positive results for smaller teams were counterbalanced by negative results for larger teams (Burgess et al, 2004). There are limited studies of the effects of PRP at more senior levels in the civil service, but one study of PRP and executive pay in the Danish government found no evidence that incentive effects from PRP improved public sector management (Binderkrantz and Christensen, 2011).

The effects of PRP on unincentivised outcomes

The effects of PRP on non-incentivised outcomes is also of interest because there is a danger that PRP may lead to misallocation of effort, i.e. a focus on incentivised outcomes to the detriment of other, non-incentivised, but equally important outcomes. The evidence on this mainly comes from health, where a number of studies show evidence of 'effort displacement' as a result of PRP, where unincentivised aspects of care have not improved at all or as much as those that are directly rewarded (Elovainio, 2010; Houle et al 2012; van Herck et al 2010). In the UK, for example, performance improvements for conditions not included in the QOF were significantly lower at the outset and these differences increased over time (Gillam et al, 2012).

However, at the same time as 'effort displacement', there is also some evidence of the opposite effect, of 'positive spillovers' of the QOF on to non-incentivised aspects of care (Eijkenaar et al 2013; Elovainio, 2010; Van Herck et al 2010). For example in the UK, there is evidence of some process measures (e.g. recording disease risk factors) that did not directly attract QOF points improving as a result of the QOF (Sutton et al, 2010) and studies also show high performance levels being maintained after QOF incentives for some indicators had been removed (Kontopantelis et al, 2014).

Several reviews have also examined the impact of PRP on equity in patient outcomes. In the UK, there is some weak evidence that the use of financial

¹² Her Majesty's Customs and Excise (HMCE), now Her Majesty's Revenue and Customs (HMRC).

incentives reduced inequalities in chronic disease management between socio-economic groups in the second and third year of the QOF, but it was unclear from the evidence if this was directly attributable to the QOF or part of an underlying trend. There were also contradictory effects on gender, age and ethnic disparities according to condition, and some evidence that gender disparities in quality of care persisted or increased following the introduction of the QOF (Alshamsan et al, 2010; Gillam et al, 2012). However, concerns that pay for performance programmes might erode equity in the provision of health care (e.g. by encouraging providers to select healthier patients or reducing income to providers serving minority populations) do not seem to have been realised in the UK (Alshamsan et al, 2010).

In education, the impact of PRP on student outcomes other than test scores is an area that is under-researched. However there is suggestive evidence that test scores may not be a good measure of general improvements in students' educational attainment. Findings from the 'Texas Assessment of Academic Skills' (TAAS) tests in the 1990s, which resulted in a dramatic rate of improvement in test scores, leading the experiment to be dubbed the 'Texas miracle', revealed a failure of TAAS score increases to translate into increases on more general college readiness tests (see Haney, 2000). This suggests that while observable gains may be made in incentivised outcomes, their overall contribution to improving the quality of outcomes may be superficial.

Unintended negative consequences of PRP

One of the key concerns in the literature on PRP in the public sector is whether it will lead to unintended negative consequences, such as staff acting strategically to improve measures of performance without increasing actual performance, also referred to as 'gaming the system'. The previous review identified some evidence of this in earlier studies (e.g. Burgess et al, 2004; Courty and Marschke, 2004). In more recent studies, there is also some further evidence of strategic behaviour.

In education, there was one notable case in 2013 of outright cheating in response to PRP. Thirty-five school officials in the US state of Georgia were charged with conspiring to alter students' test responses under a system in which rising scores were rewarded with pay bonuses (BBC, 2013). While this type of outright 'cheating' is rare, there are other more widespread examples of strategic behaviour. In education, 'teaching to the test' is one example of this. Lavy's study of a PRP scheme for Israeli teachers revealed that, while the programme appeared effective in terms of increasing pupil attainment, much of this effect was driven by the performance of those just below the median within the ability distribution. The heavy weight ascribed by the scheme to the overall pass rate was likely to have

focused teachers' efforts on ensuring 'borderline' pupils increased their performance.

In health, there are also some studies which suggest elements of strategic or gaming behaviour, as well as risk selection (whereby healthier/more compliant patients are selected to increase performance measures) (Elovainio, 2010; Eijkenaar et al, 2013). Examining the effects of the QOF in the UK, Gravelle et al (2010) found that expressing the QOF outcome measures as ratios, i.e. the proportion of eligible patients treated, encouraged practices to reduce the number of patients eligible, through a practice known as 'exception reporting'. The authors estimate that without this incentive to reduce the number of eligible patients included, practices below the upper threshold in 2004-5¹³ would have had an exception rate of 7.25 per cent instead of the actual rate of 8.55 per cent, indicating, the authors argue, a non-trivial amount of 'gaming'. However, the extent of this behaviour within the QOF system in the UK is contested, with other studies suggesting that the level is minimal (Eijkenaar et al, 2013, Van Herck et al, 2010).

Studies in the UK also show that practices perform higher than the QOF targets require. For example, the Gravelle et al (2010) study, as well as finding evidence of gaming, also found that the majority of practices treated more patients than necessary to achieve the maximum points on QOF indicators. The percentage of cases where practices were above the upper threshold (i.e. achieved a higher score than necessary to gain the maximum payments) increased from 76 per cent in 2004-5 to 92 per cent in 2005-6. The authors thus conclude that clinicians act in a 'quasi-altruistic' way.

Some studies suggest that strategic behaviour may be more likely when performance is measured less accurately. For example, Coleman's (2010) analysis of the effect of the QOF target on delivering smoking cessation advice suggests that it encouraged strategic behaviour in the form of administrative changes (i.e. increased rates of *documenting* advice given in medical records) rather than any change in the rate of meaningful clinical interventions. The author concludes that outcome targets for less easily-defined interventions, such as health promotion advice, and which rely on clinicians documenting their own practice, are especially prone to gaming behaviour. Another example is Burgess et al's (2004) study of a pilot PRP scheme in the UK public employment service, which found improved outcomes on quantity but not quality indicators. The authors suggest that this resulted from strategic behaviour on the part of employees, since quality outcomes were measured with less precision and it was therefore more difficult to gain

¹³ i.e. those who still had room for improvement and thus higher QOF payments

improved scores. Individuals responded to this by focusing their effort on quantity rather than quality.

The effects of PRP on staff attitudes, motivation and behaviour

One issue which has attracted much attention in the research literature is the effect of PRP schemes on staff attitudes, motivation and behaviour. Understanding the link between PRP and job satisfaction is especially important, as there may be a significant link between job satisfaction and work motivation – and thus a link to productivity – while job satisfaction is also a predictor of job retention. The need to further explore these links becomes even more apparent in light of a 2013 report by CIPD showing that the majority (two out of three) of public sector employees do not think that their salary should reflect individual performance (CIPD, 2013). The evidence around PRP, job satisfaction, work motivation and productivity speaks directly to the question of how PRP schemes achieve their effects, since economic theory suggests that the principal mechanism is through increased motivation and effort among staff. One additional concern in the public sector literature is whether introducing financial incentives through a PRP scheme might negatively impact on public servants' *intrinsic* motivation to perform well at their job – a phenomenon known as 'crowding out' - since it is thought that staff in the public sector are driven by higher intrinsic motivation than those in the private sector (Frey et al, 2013).

In education, a number of studies have looked at the effect of PRP on teachers' attitudes and behaviours, such as the level of effort expended, collegiality and job satisfaction. The evidence for the effects of PRP on these factors is mixed, but overall is more negative than the findings for the impact on student test scores. One study by Lavy (2009) of a PRP system in Israel identified an increase in the provision of extra instruction after school hours amongst teachers involved in the scheme. This was identified as one factor driving the improved pass rates and test scores seen as a result of the scheme. Belfield and Heywood (2008), in a quantitative study, also found that co-operative working relationships were consistently associated with PRP. However, in a review of the effects of PRP on teacher attitudes, Yuan et al (2013) found no impact on the levels of collegiality between teachers. Another study (Jones, 2013) distinguished between paid and unpaid co-operative activities, and found a reduction in unpaid co-operative activities among teachers in performance-pay districts, while levels of paid co-operative activities remained the same. The study also reported that teachers in PRP districts worked 12 per cent fewer hours each week and were significantly more likely to take on an additional job outside of teaching, providing possible evidence that PRP increases teachers' focus on remuneration and negatively affects their intrinsic motivations for teaching.

Several studies have also found that PRP had a detrimental effect on teachers' job satisfaction and motivation. Belfield and Heywood (2008, 244), for example, found that, once earnings were controlled for, PRP emerged as a negative determinant of overall job satisfaction, teachers' belief that their effort was being rewarded, and their satisfaction with their salary. Gius (2013, 446) also found that teachers working within PRP districts were less likely to believe teaching was important, less likely to be enthusiastic about teaching, and were more likely to leave for better pay than were teachers who did not work in merit pay districts. However, in contrast, Jones (2013) found that teachers involved in a PRP scheme were more likely to pursue opportunities for professional development and less likely to express a desire to leave the profession.

Evidence also suggests that teachers have often been unsupportive of PRP systems, and the types of programmes that receive least support are those based on standardised test scores (Leigh, 2013). Several surveys of teacher attitudes to PRP reveal that most felt that standardised testing failed to capture important aspects of teaching performance (Yuan et al, 2013, 14). Teachers also felt unable to influence some factors that had a significant impact on achievement, such as pupils' family environments. Springer et al's analysis in Tennessee (2012, p.140) also found teachers unsupportive of the PRP scheme, with most feeling that they were already teaching to the best of their abilities.¹⁴

Studies also suggest that the response of teachers to PRP may differ according to individual characteristics, such as gender and level of prior experience. For example, some studies suggest that male teachers appear to respond to PRP systems more positively than their female counterparts and are also more likely to support PRP than female teachers (Leigh, 2013; Jones, 2013). Jones (2013) also found that there is a discrepancy between the effect of PRP on male and female teachers' work hours, with women more likely to reduce their hours under PRP than men. The study showed that women's participation in unpaid cooperative activities was reduced by 43 percentage points in PRP districts, while men's participation was unchanged. Experienced teachers also tend to display more negative reactions to PRP than early career teachers. For example, Leigh (2013) found lower support for PRP amongst experienced teachers and Jones (2013) reported that PRP was associated with lower work effort among experienced teachers. However, it is unclear whether this effect was driven by more experienced teachers' hostility to changes in the traditional running of the system, or to their previous negative experiences of PRP.

¹⁴ In this case, the initiative failed to register any significant improvement in students' test scores.

With regard to the civil service, the evidence suggests, firstly, that public service (or intrinsic) motivation is important – and more so than financial incentives - in motivating behaviour, and secondly, that the evidence is mixed regarding whether financial incentives ‘crowd out’ this intrinsic motivation.

Several studies examine the relationships between pay systems, public service motivation and work effort. One large study (Taylor and Taylor, 2010) analysing data from the 2005 *International Social Survey Programme (ISSP)* on public service employees from 15 countries found that although wages and public service motivation both possess motivational properties, the effort levels of many government workers, especially those at supervisory levels, were affected more by public service motivation than wages. This finding is mirrored in another study in Switzerland (Anderfuhren-Biget et al, 2010) looking at the relative influence of material incentives, public service motivation and team relations on individual work motivation. The study found that public service motivation and the perceived quality of team relations and support were more important determinants of work motivation than material incentives, which were only weakly related to work motivation. Boardman and Sundquist (2009) also introduce a new related variable, ‘perceived public service efficacy’, which quantifies public servants’ perception about the benefit their employing agencies provide the public. They provide empirical evidence that this variable is positively related to job satisfaction and organisational commitment and negatively related to role ambiguity, thus driving outcomes that are positive for both the worker and the employer (Boardman and Sundquist, 2009).

The evidence on whether the use of financial incentives, such as PRP, harms this intrinsic motivation is mixed for the civil service. The Swiss study referred to above (Anderfuhren-Biget et al, 2010) found a negative association between public service motivation and a disposition towards material incentives, which the authors interpret as evidence of the ‘crowding out’ effect of intrinsic factors by extrinsic motivators. This is also suggested in a meta-analysis by Weibel et al (2009) which found that while there is a significant positive effect of PRP on performance for simple tasks, there is a negative effect for complex tasks, which the authors argue is due to the crowding out of intrinsic motivation by PRP. A review of theoretical and empirical studies in the US, England and France by Forest (2009) also concluded that individualised remuneration practices can in the long-run produce negative effects on the intrinsic motivation that drives civil servants.

However other studies suggest that PRP systems do not harm intrinsic motivation. For example, a study by Stazyk (2013) of managers in US local government jurisdictions found that employees who worked under a variable pay system - rather than a standard compensation system - had higher levels of public service

motivation, enjoyed greater role clarity, and were more likely to report greater job satisfaction. The findings also suggested that variable pay systems were a more important determinant of job satisfaction among employees who had lower public service motivation, suggesting that PRP may be useful in offsetting lower levels of job satisfaction when employees do not possess strong altruistic motives. Finally, a randomised field experiment (Ashraf et al, forthcoming) comparing the effects of financial and non-financial incentives on the performance of agents recruited by a public health organisation to promote HIV prevention, also found little evidence of crowding out. The study found that while non-financial incentives ('stars' for good performance) were the more effective incentive, both types of incentives resulted in increased effort, and the effect of both types of incentives were stronger for pro-socially motivated agents.

There is less research in the health field examining the effect of PRP on health staff behaviour, attitudes and motivations, but the evidence that is available presents a mixed picture. One RCT study in the US (Petersen et al, 2013) that showed a positive impact of individual-level financial incentives on physicians' practices with respect to blood pressure control, found that the effects wore off rapidly after the incentives were withdrawn, providing strong evidence that physicians were responding directly to financial incentives through increased effort. However in the UK, studies have shown that high performance levels on outcome targets have generally been maintained after QOF incentives for some indicators were removed (Kontopantelis et al, 2014), suggesting that other mechanisms were at work in driving the improvements.

Several qualitative studies in health have attempted to assess the effects of PRP on healthcare providers' intrinsic motivation and professionalism. Two studies included in Eijkenaar et al's (2013) review found that PRP had no impact on this, but three UK studies suggested that PRP may have resulted in a perceived loss of autonomy for healthcare providers and undermined their sense of professionalism. In particular, some health professionals have been uneasy about the emphasis on protocol-driven care as a result of the QOF, and feel that care may be becoming less patient-centred (Gillam et al, 2012). A recent qualitative study (Lester et al, 2013) also pointed to staff dissatisfaction about 'micro-management' by the Department of Health in continual changes to indicators. Some staff wished to see a greater emphasis on involving front line practice teams in developing QOF indicators as a way of regaining some clinical autonomy.

The health literature also suggests that the motivational effects of financial incentives are contingent on the degree to which participants perceive incentives to be consistent with their own work-related values and needs. For example, a study

in the US (Young et al, 2012) found that physicians' responses to incentives (to promote clinical tests and screenings to diabetic patients) were stronger among those who were less concerned about whether the incentive programme represented a threat to their autonomy and among those who believed that the programme's goals were aligned with their own professional goals. In the UK, qualitative studies have found that general practice staff thought that the QOF acted as an incentive to provide what they themselves regarded as good clinical care. The fact that the indicators were evidence-based was a very important factor in staff support for the scheme (McDonald et al, 2007; Lester et al, 2013). This provides little evidence that the QOF is acting as a threat to, or 'crowding out', the internal motivation of healthcare staff.

The effects of PRP on work organisation and team relationships

While much of the theoretical underpinning for PRP derives from economic theory, which focuses attention on the effects of PRP on individual motivation and effort, HR theory places an emphasis on the role of workplace relationships and structures in influencing the effectiveness of PRP. Several studies in the civil service have examined the interrelationship between financial incentives and managerial or team relationships in determining PRP outcomes, finding that workplace relationships are key. In health, while this is not a key focus of the literature, evidence likewise suggests that changes to work organisation are a key mechanism through which improvements in quality may be achieved.

The UK tax office study referred to earlier (Burgess et al., 2010) highlighted the importance of effective managers in contributing to the outcomes from PRP. The evaluation showed that of the two teams that participated in the scheme, only one successfully achieved the bonus. While the individual performance of staff was heightened in both teams, the winning team engaged in task reallocation – with managers moving efficient workers towards the incentivised tasks - to a significantly greater degree than the other team, thus leading to the increase in performance. The authors suggest that this may have been because the office managers in the winning team were younger, and this may have resulted in more responsiveness to the incentives due to longer term career concerns.

In the health field, there is also some evidence that improvements in care quality as a result of performance-related pay may be achieved through changes to work organisation, rather than as a result of an increase in individual effort, particularly when incentives are targeted at teams or institutions. One study of the QOF in the UK (Kontopantelis et al, 2014), for example, found that following the withdrawal of incentives for several activities in the QOF, levels of performance were generally

maintained. The authors suggest that this was because the QOF supported the establishment of quality improvement infrastructures in practices that kept performance high even after incentives were withdrawn. This is supported by qualitative studies looking directly at changes to practice organisation following the QOF, which find that care for patients with long-term conditions is now more proactively planned and managed on an annual basis (Lester et al, 2013). This is reflected in increased use of computers, decision support, clinician prompts and patient reminders and recalls (Gillam et al, 2012). This may also help to explain the positive spillover effects of the QOF found for non-incentivised aspects of care (Eijkenaar et al 2013; Elovainio, 2010; Van Herck et al 2010).

Similarly, a study of hospital-level incentives in the US (Sautter et al, 2007) found that where the incentives were associated with improvements this was down to hospitals investing resources in quality improvement processes to support clinicians. These included redesigning clinical processes, establishing systems to provide feedback on physician performance, and better performance reporting. The authors conclude that whether institution-level incentives are effective largely depends on the organisational context; in particular whether hospitals are able to bring sufficient resources to bear in supporting clinical process improvements.

While the effect of the QOF in the UK was to support the establishment of quality improvement infrastructures, it also resulted in changes in the distribution of work between health teams within practices, with nurses in particular gaining greater responsibility (MacDonald et al, 2007). While this was broadly welcomed by nurses in that study, there is also some evidence that this has caused resentment amongst those whose work has increased as a result of the QOF but who are not in receipt of performance payments (e.g. nurses)¹⁵, with potentially adverse effects on team relations.

The issue of perceived unfairness of PRP resulting in detrimental consequences for team relations is also highlighted in the wider literature on PRP and New Public Management in the civil service. Bregn (2013), for example, reviews experimental and field studies on the relationship between perceived fairness in pay and performance. These studies show that differential payment for the same task is associated with reduced effort on the part of the employee who is paid less,

¹⁵ Only profit-sharing (independent contractor) GPs in England benefit from the QOF payments directly, although the activities of all practice staff, including salaried GPs, practice administrators and nurses, are essential to the achievement of the targets (Lester et al, 2013). General practices decide how payments are used and there is no requirement for these to be shared amongst the primary health care team; any diffusion of incentive payments to non-GP staff is controlled by those who run practices (Coleman, 2010).

although the effect is lessened if the wages are randomly generated; and that workers respond more negatively to wage cuts if the wages of co-workers are not similarly reduced (Bregm, 2013). The implication of this is that PRP systems may have detrimental effects on performance if they are perceived to be unfair in their design or implementation. While there is empirical evidence in the wider social psychology literature and from studies examining behaviour in the private sector of diminished effort from perceptions of unfairness, evidence of this happening as a result of a PRP scheme is limited.

Perceived unfairness of PRP systems may be one reason behind the finding in the education literature that female teachers respond more negatively to PRP schemes. There is wider evidence from the literature that women, on average, dislike highly competitive situations and that male employees tend to respond better to a competitive workplace environment (see e.g. Gneezy and Rustichini, 2004). Women may also respond less to direct incentives like PRP than men because if, on average, they have more non-work calls on their time (e.g. due to family responsibilities) or greater likelihood of absenteeism, then they will be less able to put in additional effort at the margin and gain returns from PRP (Green and Heywood, 2010).

Cost effectiveness of PRP schemes

The cost-effectiveness of PRP schemes is an area that is under-researched across all three public services considered in the review – and an area that is under-studied generally – although there have been some cost effectiveness studies of PRP schemes in health. A systematic review on this topic found several studies that showed improvements in quality from PRP schemes, alongside increased costs, which suggested the potential for PRP to be cost-effective. However the review concluded that it was unable to come to robust conclusions about the cost effectiveness of PRP due to ‘variable methodological quality’ and differences across programs in context and design (Emmert et al, 2012).

In the UK, Walker et al (2010) analysed the cost effectiveness of nine performance measures in the QOF for which there was cost-effectiveness data available (i.e. evidence on the costs and health effects of the interventions). The study was conducted using 2004/5 data, when robust evidence on the impact of the QOF was unavailable, and so the study presented the *potential* cost-effectiveness of the QOF subject to the level of its impact. This analysis suggested that for most of the indicators considered, QOF payments would be cost-effective even if the scheme

achieved only a small impact, although this varied considerably by indicator, ranging from 0.06 to 19.7%¹⁶. The assessment of cost-effectiveness was based on a value for money threshold of £20-30,000 per Quality Adjusted Life Year (QALY) achieved. The limitations of the study were that it only considered nine, out of a total of 149 QOF indicators (at that time), due to data limitations for other indicators, and it did not take into account the administrative costs of running the scheme.

Some studies in health have also suggested that other means of improving the quality of healthcare may be more effective than PRP. For example, in a review of the relative effectiveness of different quality improvement strategies in health, Scott (2009) concluded that clinician or patient-driven improvement strategies¹⁷ had stronger evidence of efficacy and larger effect sizes than manager/policy-maker driven strategies, including PRP – which was only found to be modestly effective¹⁸. The review did not have the evidence, however, to compare the cost effectiveness of these different methods.

¹⁶ i.e. some indicators would be cost effective if the proportion of eligible patients treated increased only by 0.06 per cent as a result of the QOF, whereas for others it would need to increase by more; the largest amount was 19.7 per cent. This diversity largely reflected the cost of each intervention and their expected health benefits.

¹⁷ Examples include clinician-directed audit and feedback cycles, clinical decision support systems, specialty outreach programmes, chronic disease management programmes, continuing professional education based on interactive small group case discussions, and patient-mediated clinician reminders

¹⁸ The review used evidence from controlled trials to determine the effect sizes for different strategies, and then these were compared across strategies in a narrative review.

5. Performance-related pay scheme design

As the evidence reviewed in the previous chapter suggested, results from PRP schemes are often inconsistent, with some programmes showing significant impact and others very little. Much of this variation can be linked to the design of PRP schemes, which plays a significant role in determining outcomes. There are a range of different permutations to consider regarding scheme design. These are discussed below, along with findings, where available, regarding their effect on PRP outcomes. However, there is a general conclusion in the literature reviewed that the evidence is not available to offer simple prescriptions on how PRP schemes should be designed. This is partly due to weaknesses in the evidence base – for example, there are very few individual studies that have made a robust assessment of the impact of design features (Eijkenaar et al, 2013) - and partly because of the sensitivity of scheme design to aspects of the service concerned, such as its aims and objectives, the organisational setting, and the composition of the workforce.

Range of goals and targets

The nature and range of the goals and targets included in PRP schemes is an important issue. Evidence suggests that PRP schemes will be more effective if the goals are specific and easy to track and measure (Eijkenaar et al, 2013). If the goals and targets are too complex, and staff are not easily able to understand the scheme, this will limit its effectiveness in altering behaviour (Leigh, 2013; Lundstrom, 2012). There is also some suggestive evidence that outcomes which are less easily measurable are more prone to gaming (Coleman, 2010). However, outcomes in public services are often complex and a simple model with a narrow range of targets can bring the risk of an overly narrow focus, potentially resulting in misallocation of effort and the neglect of less easily measurable aspects of performance.

Reviews of PRP scheme design in health also suggest that process indicators (e.g. treatment targets) have generally yielded higher improvement rates than outcome measures (e.g. hospital (re)admissions), with intermediate outcome measures (e.g. blood pressure control) yielding improvement rates somewhere in between (Eijkenaar, 2013; So and Wright, 2012; Van Herck et al, 2010). However if process indicators are used it is important for the link between the indicator and clinical outcomes to be soundly evidence-based to ensure that valued social outcomes are achieved (So and Wright, 2012). Reviews also find that selecting goals where there

is greater room for improvement results in a higher effect size than if there is less room for improvement (Eijkenaar, 2013; Van Herck et al, 2010).

Education and health vary considerably in the nature and range of goals and targets used in PRP schemes. In education, schemes are primarily based on a very simple measure of students' performance in standardised tests, which is easy to understand, measure and track. However, scepticism has been expressed by teachers about whether test scores adequately capture important elements of teaching performance (Yuan et al, 2013). It has been suggested that teachers could be incentivised on the basis of additional process or intermediate outcome measures (e.g. pupil attendance, completion of homework, additional reading, etc.) instead of the singular focus on test scores (Fryer, 2011). In the QOF system for primary care in the UK, in contrast, there are a complex range of measures, including process indicators and intermediate outcome measures, with the scheme described as the most elaborate and extensive PRP scheme in healthcare in the world (Gravelle et al, 2010; Lester et al, 2013). The wide range of targets included within the scheme, along with the continual updating of indicators¹⁹, is designed to guard against the problem of misallocation of effort. However the trade off is that the scheme is complex to administer and some of the targets have less accurate/reliable data.

Individual or team targets

Evidence from the health literature shows that positive effects of PRP are more likely and/or greater when schemes are directed at individuals or small teams than when directed at (large) groups (Eijkenaar et al, 2013; Van Herck et al, 2010). One example is a field experiment in the US which randomly allocated primary care physicians to receive individual-level, practice-level, both or no financial incentives for controlling patient blood pressure (Petersen et al, 2013). The study found that only individual-level incentives had an effect on the outcomes measured. An earlier study in education by Lavy (2004) also found that an individual level incentive scheme for teachers in Israel was more effective, at a significantly lower cost, than a team-based incentive scheme. A lack of impact from team targets for larger groups may be related to 'free-riding' problems or to a lack of staff awareness. Several studies have attributed a lack of impact from PRP to a lack of awareness of the incentives among staff where incentives were distributed at the organisational level (Eijkenaar, 2013). An earlier study in the UK employment service also found that

¹⁹ Indicators are removed if the evidence changes or if achievement rates have reached a ceiling, to allow new indicators, for which improvement is possible, to be introduced (Kontopantelis et al, 2014).

team-based incentives were more effective in smaller groups because there was more scope for peer monitoring of colleagues' effort which prevented 'free-riding' (Burgess et al, 2004).

Van Herck et al's (2010) review of the design impacts of PRP schemes in health concludes that programmes aimed at the institutional level *can* have a positive impact, but the effects are generally smaller, with additional effort required to generate support for the scheme or to transfer incentives internally. There is some evidence that institution-level incentives can have a positive effect if they support changes to quality improvement infrastructures within a setting. For example, a study of performance-related pay in US hospitals (Sautter et al, 2007) suggested that whether institution-level incentives have a positive effect depends on the extent of room for improvement and the resources available to support clinical process improvements. The study found that improved performance could be achieved through changes to working practices despite minimal direct knowledge of the incentives by individual physicians. However, the authors note that improvements might have been greater if individual physicians had been included in the rewards from the scheme more directly e.g. through gain-sharing arrangements.

Types of performance measure

As well as the types of goals rewarded, the way that measures of performance are constructed also has an important bearing on behavioural responses to PRP schemes. Whether absolute or relative (i.e. improvement-based) measures of performance are used is one key issue. Absolute measures of performance may reduce the incentive for low-performing providers, given their reduced likelihood of attaining the targets. However, relative measures tend to reward providers with the lowest baseline performance since they are often able to improve the most. To overcome these difficulties, it is suggested that combining absolute and relative payment systems which reward both target attainment and improvement are preferable (Eijkenaar et al 2013, Elovainio, 2010).

One study of a large pay-for-performance scheme in US hospitals (Ryan et al, 2012), however, which changed its design in order to reward improvement as well as attainment, found that this change had no impact on the lowest performers, who improved less than in phase one (where improvement was not rewarded). The authors suggest that this may have been because of poor incentive design, whereby hospitals had to both improve and be above the median, which excluded many of the poorest performers, or because the poorest performing hospitals lacked resources to respond (further) to the financial incentives. In general, reviews find that most PRP programmes in health have targeted absolute results and these

schemes have been more likely to find positive effects, possibly because absolute outcomes are easier to comprehend and track (Elovainio, 2010; Eijkenaar et al 2013).

Another factor is the use of threshold or linear (continuous) measures of performance. Linear measures reward per unit improvement and thus value progress at all parts of the distribution while threshold measures reward the achievement of a certain level. Which is the optimal measure depends on the ultimate goal of the scheme - whether overall performance improvement (at all levels) is desired or whether the attainment of one central performance goal (e.g. the pass rate) is the most important (OECD, 2009). The behavioural implications are illustrated in Lavy's (2009) study of a PRP scheme for Israeli teachers, which revealed that while the programme was effective in increasing pupil attainment, much of the effect was driven by the improved performance of pupils just below the median within the ability distribution. This was attributed to the heavy weight ascribed by the scheme to the overall pass rate. One way to address this is to combine linear and threshold measures – as with the QOF scheme in the UK, in which payments to practices for clinical indicators are proportional to achievement between a lower threshold (initially set at 25 per cent for all indicators) and an upper threshold (which varies across indicators from 50 to 90 per cent) (Kontopantelis et al, 2012). Van Herck et al's (2010) review in health found no consistent relationship between the two different types of performance measure and the presence of positive scheme effects, although it found that positive effects tended to be larger in many schemes for the lowest performing providers.

Rank order tournaments are another type of performance measure, which rank individuals according to performance and award bonuses to a set number. This can be beneficial from an administrative point of view, in that it allows better control of expenditure (because a set number of staff will receive a payment in any given round) (OECD, 2009). However, the possibility of “false distribution” or “false ranking” may also occur, e.g. where a high achiever could still be outside the top 10 per cent of performers, for example, who are rewarded, or a satisfactory performer in the bottom ten per cent who are penalised. This could disincentivise cooperation between staff and demotivate individual effort by causing crowding-out effects.

Related to this issue, the level of selectivity within a PRP scheme – i.e. the proportion of staff who are likely to receive bonuses – also has behavioural effects. Figlio and Kenny's (2007) review of PRP schemes in education found that PRP schemes which offered bonuses to a larger proportion of teachers produced fewer positive impacts in terms of student outcomes than those that were more selectively distributed. On the other hand, Schmidt et al's (2011) study of PRP schemes in

German public services found that highly selective PRP schemes reduced rather than enhanced employee motivation due to a dissatisfaction on the part of employees who perceived the quotas (that allowed only a limited number of employees to receive rewards) as unfair.

Finally, whether performance measures are based on objective or subjective criteria is another factor. Subjective measurements can result in perceived inequity, with negative implications for employee effort and motivation. There is also some evidence that inaccurate measures are less effective since employees act strategically and focus on more easily measurable (and hence attainable) outcomes (Burgess et al., 2004). However focusing on objective measures only carries the risk of misallocation of effort.

Size and volume of payments

The modest size of payments in many PRP schemes is often given as a reason for limited effectiveness (e.g. Chung et al, 2010; Greene et al, 2013; Schmidt et al, 2011). In the UK, the QOF scheme establishes a fairly large incentive – gaining the maximum points would account for around 20 per cent of a GP's salary - which is high by international comparison, and qualitative research with healthcare staff suggests that this is perceived to be an important factor in its success (Lester et al, 2013). However, reviews of PRP in health have failed to find a consistent relationship between the size of incentives and their impact (Frølich et al, 2007; Elovainio, 2010; Van Herck et al, 2010). This may be partly due to other variations in scheme design and outcomes as well as the size of payments (So and Wright, 2012). However, some studies suggest that larger incentive sizes may encourage participation where a scheme is voluntary (Van Herck et al, 2010).

The relationship between payment frequency and impact is not established either. Some studies in health suggested that timely payments following achievement were a factor contributing to scheme success (Eijkenaar, 2013). However, a randomised experiment in the US (Chung et al, 2010) found that altering the frequency of payments (quarterly or annually) had no effect on physician response to a pay for performance scheme. However, in this case, the modest overall response to the scheme may have limited any effect of payment frequency.

System and organisational characteristics

While not technically about scheme design, one review study in health (Van Herck et al, 2010) also examined the effect of system and organisational characteristics on the impact of PRP schemes. This showed that national-level PRP schemes (as in

the QOF in the UK) lead to more uniform results, while fragmented initiatives (as in the US) lead to more variable results. Some studies in the US also suggested that the presence of multiple payers with different incentive schemes dilutes the impact of any one PRP scheme. Limited evidence is available on the impact of existing payment systems on PRP impact, although it seems likely that the effect would be larger where incentive structures are aligned (Van Herck et al, 2010). In a Canadian study, Kantarevic & Kralj (2012) found that responses to a diabetes management bonus scheme differed according to the existing payment model in use. Specifically, physicians paid via a capitation model (in which they received a fixed payment for each enrolled patient) were more responsive to performance-related pay than those paid via a fee-for-service model (in which they received the full value of the services provided). The authors concluded that doctors are more responsive to pay for performance if they are operating in pay schemes that impose more cost-sharing.

Van Herck et al's (2010) review found mixed results regarding whether schemes were implemented on a voluntary or mandatory basis. One study that tested whether voluntary schemes resulted in an over-representation of high performers found this not to be the case. However a US study (of the Premier Quality Hospital Incentive Demonstration) found significant differences between participants and non-participants in a voluntary scheme. The national Practices Incentive Programme in Australia similarly found that higher performers were more likely to be participants in the scheme which resulted in limited room for improvement (Greene et al, 2013).

PRP schemes in health are often implemented in combination with other quality improvement measures such as feedback, education and public reporting. There is some evidence that combining PRP with other quality measures can enhance impact, particularly if PRP is combined with (public) performance reporting mechanisms (Eijkenaar et al, 2013; Van Herck et al, 2010). However some studies (primarily in the US) suggest that this has not been the case, so the evidence here is inconclusive.

A range of studies have concluded that participative PRP schemes in which employees are involved in the design of scheme outcome measures and targets are more effective than those in which employees have had no input. For example, in a study of a range of PRP schemes in German public services, Schmidt et al (2011) found that more participative PRP schemes achieved a higher level of acceptance among employees and were more effective in motivating employee effort than non-participative systems (although the study found that the motivational effect of PRP on employee effort was weak across all cases). Reviews in healthcare have also

concluded that programmes designed collaboratively, with providers involved in the selection and definition of performance measures and targets, and with effective communication of scheme targets and rewards among stakeholders can achieve better results (Eijkenaar et al 2013; So and Wright, 2012; Van Herck et al 2010).

Finally, several studies have related positive findings from PRP to aspects of organisational culture and structure, although findings are mixed and so robust relationships are lacking. For example, studies have found a positive relationship between positive PRP effects and:

- an organisational culture supporting the coordination of care,
- the willingness to try new projects,
- a history of engagement with quality improvement initiatives,
- a multidisciplinary team approach, and
- adequate human resources for quality improvement projects.

In conclusion, the findings from across the three public services suggest that the design of performance-related pay schemes has an important bearing on the results obtained, however it is not possible to state with certainty which types of scheme are most effective overall, partly due to methodological limitations of the evidence base - studies have rarely set out to test the impact of scheme design, and variable scheme and evaluation designs have tended to inhibit meta-analysis (Eikeaneer et al 2013; Flodgren et al 2011; Houle et al 2012; Scott, 2009) – and partly because effectiveness is contingent upon the public service context and the specific policy aims in these contexts. Nonetheless, further pilots and evaluations with robust, ideally experimental, designs would be beneficial to provide further guidance on aspects of scheme design within specific public policy contexts (Podgursky, 2007). The findings on scheme design are summarised in Table 5.1 below.

Table 5.1: Variations in PRP scheme design

Variations	Factors affecting decision
Nature and range of goals and targets	<ul style="list-style-type: none"> • Broader range of targets helps to prevent misallocation of effort, but simple, specific, easily measurable targets are clearer and easier to understand for staff and may enhance scheme effects. • Less well-defined and less easily measurable targets may allow room for gaming. • In health, process targets have yielded greater improvements than outcome targets, but the link between process measures and outcomes needs to be robustly evidence-based. • Selecting goals where there is greater room for improvement is likely to yield higher rates of improvement due to 'ceiling effects'.
Individual versus group-based targets	<ul style="list-style-type: none"> • The public sector is characterised by collaborative activity, which suggests team targets may be more effective; however if the target is set at the organisational level this may not be easily conveyed to individual staff. • Evidence across the three public services suggests that individual or small group targets are more effective than institutional level targets. Small groups may allow for more effective peer monitoring of behaviour • Process studies suggest that institutional targets can be effective if they support quality improvement infrastructures, but this requires an adequate level of resources.
Relative or absolute performance measures	<ul style="list-style-type: none"> • Absolute measures of performance reduce the incentive for low performing providers, but relative measures can reward the lowest providers the most. A combination of both types of measures may be optimal. • Studies in health suggest absolute measures are more likely to generate positive impacts, although studies using relative measures are more scarce.
Linear versus threshold payment structures	<ul style="list-style-type: none"> • No clear findings on which types of measure are likely to lead to performance improvements – this depends on the aims of the scheme and other design features. • If the aim is general performance improvement, linear measures are preferable; if there is one central performance goal, a threshold measure may be preferable. A combination of the two may be optimal. • Threshold measures may result in limited improvement at the bottom of the distribution.

Rank order tournaments	<ul style="list-style-type: none"> • Allow better control of expenditure because a set number of staff will receive a payment in any given round, but the possibility of “false distribution” or “false ranking” may disincentivise cooperation between staff and demotivate individual effort through crowding-out effects.
The extent of selectivity within the scheme	<ul style="list-style-type: none"> • In education, greater selectivity seems to simulate more improvement, however if quotas are perceived to be unfair this can also demotivate individual effort through crowding-out effects.
Objective versus subjective measurement criteria	<ul style="list-style-type: none"> • Objective results may raise fewer questions around fairness and favouritism, but risk focusing only on outcomes which can be easily measured.
The size and/or frequency of incentive payments	<ul style="list-style-type: none"> • Smaller bonuses may produce little effect, but larger payments may raise concerns about cost-effectiveness of schemes. • No robust relationship has been established between size and frequency of payments and effect sizes • Reward size may increase participation levels in voluntary schemes

6. Conclusions

The review has found evidence that PRP schemes can be effective across the three public services for which evidence is available (health, education and the civil service), although the central conclusion is still that findings are mixed, which much dependent upon organisational and occupational context and scheme design and implementation. The weight of evidence also varies across these services, with the more robust evidence coming mainly from education and health. In education, positive effects (and indeed scheme design) tend to be concentrated on students' performance in standardised tests, whereas in health, process measures (e.g. treatment targets) in primary care and preventive services have shown the most improvement as a result of PRP. Where positive effects have been found, effect sizes are often small and may be short-lived – emphasising the value of longer-term follow-up evaluations. Studies in health have also suggested that other improvement strategies may yield equal or greater results than PRP, but comparative – or indeed any - robust cost effectiveness data is rare.

Given limitations in the evidence base with regard to concluding when, where and how PRP is or is not (cost) effective, in the remainder of this chapter we summarise from the review findings the key challenges that face PRP schemes in the public sector and discuss the implications of these challenges for the future design and operation of PRP. Figure 6.1 visually depicts these key challenges, showing where the challenges arise within the context of the PRP logic chain.

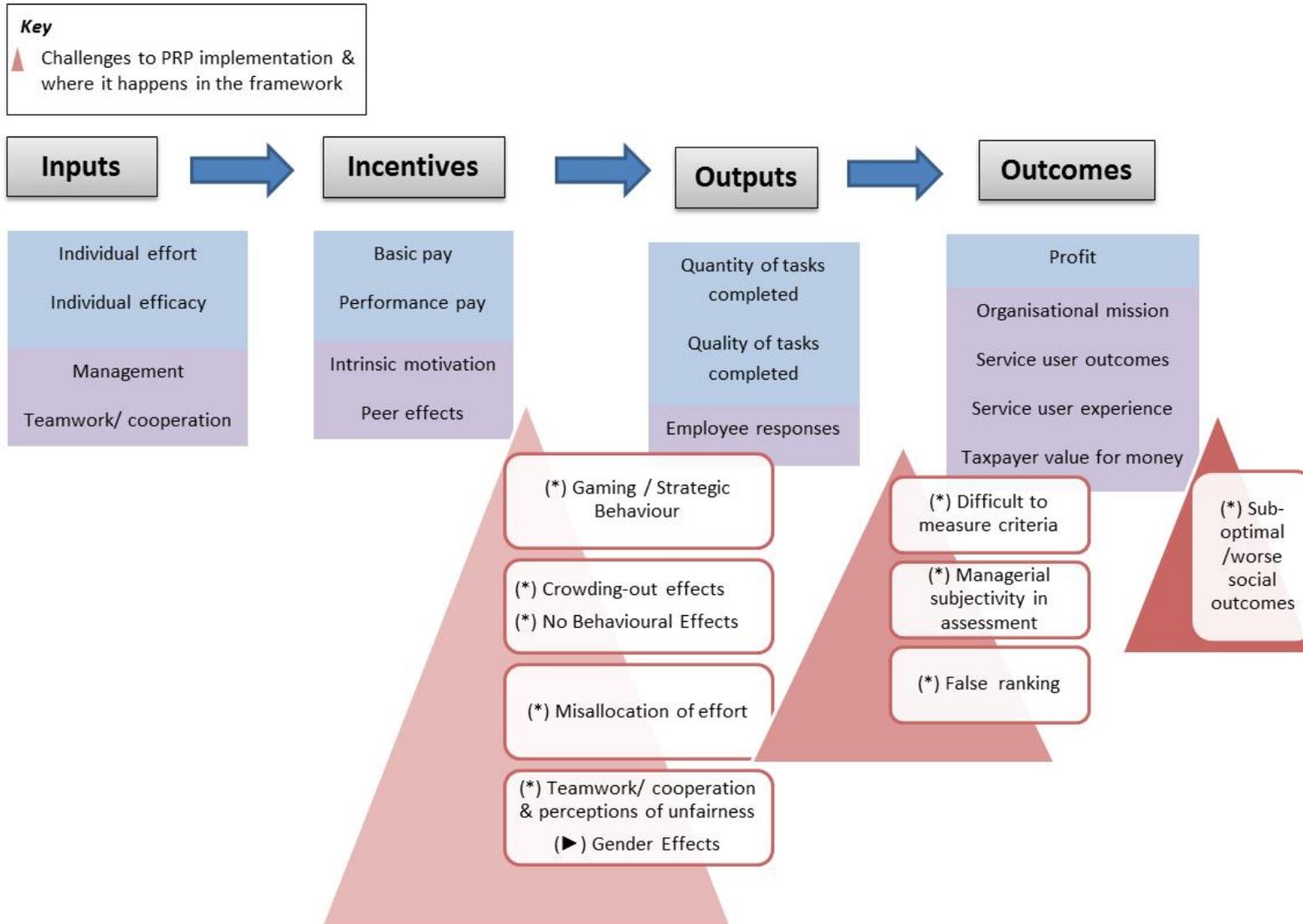
Unintended behavioural effects of PRP

The first set of challenges relate to unintended behavioural consequences as a result of the incentives enshrined in PRP. These include:

- gaming behaviour
- crowding out effects
- an absence of behavioural change
- misallocation of effort and
- detrimental consequences to teamwork and co-operation.

Some of these effects may suggest a challenge to the principles underlying PRP logic (for example, that financial incentives induce additional effort), whereas others suggest design challenges to PRP in the public sector. However, the empirical evidence is not necessarily able to distinguish satisfactorily between these two explanations for the unintended consequences of PRP.

Figure 6.1: Key challenges facing PRP in the public sector



Misallocation of effort and strategic/gaming behaviour

Misallocation of effort can occur through the tendency of performance incentives to focus on the tasks and outputs most easily measurable. This encourages employees to place more emphasis on these tasks and outputs to the detriment of others which – although they may not be as easily measured – may be equally significant to outcomes. There is some overlap here with strategic or ‘gaming’ behaviour, where workers seek to maximize their incentive gains while minimizing effort or without increasing actual performance. This may be seen as an extreme form of misallocation of effort that is intentionally strategic – or reward-maximising - on the part of participants.

To the extent that material rewards have an incentive effect upon individual behaviour, *reallocation* of effort will be an inevitable consequence of PRP, and indeed is often its intended effect. Whether this becomes significant *misallocation* of effort depends on how the outputs incentivised relate to the desired outcomes of the service. As noted previously, this is more difficult to determine in public services due to the complexity of the public ‘good’ and multiple stakeholders with an interest in the outcomes.

A contrast between PRP schemes in education and health is illuminating here. In education, the vast majority of PRP schemes incentivise outcomes on student test scores. That many of these schemes have been effective in raising test scores, while also having negative or contradictory effects on co-operative and collegial behaviour among teachers, suggests that these type of schemes encourage teachers to ‘narrow’ their approach, focusing largely on raising student attainment in the particular tests incentivised and, where necessary, focusing on ‘borderline’ pupils to achieve a threshold target. This has led to concerns about the rise of a narrow, ‘teaching to the test’ approach in education (Neal, 2011), which many educationalists view as a misallocation of effort. The extent to which this is the case depends upon whether test results are the only, or the most valued, outcome of the education system. As noted earlier, findings from the ‘Texas Assessment of Academic Skills’ (TAAS) tests in the 1990s suggested that observable gains in student test scores did not necessarily translate into wider educational outcomes.

In the Quality and Outcomes Framework, the PRP scheme for GPs in the UK, in contrast, there is less danger of misallocation of effort because the scheme is much more complex, with a range of targets across different domains. This multitude of indicators is designed to incentivise a broad range of valued outcomes from healthcare rather than a narrow focus on particular interventions. The evidence suggests that this approach has been relatively effective. While there is some evidence of reallocation of effort towards incentivised tasks (Gillam et al, 2012), and of low-level gaming of the system through ‘exception reporting’ practices (Gravelle et al, 2010), there is also evidence of positive spillovers onto non-incentivised aspects of care, which make a significant contribution to the overall value of the scheme (Sutton et al, 2010). The evidence overall suggests that the QOF has achieved its outcomes through changes to team working and workplace organisation – and this indeed was the intention of the QOF, which was designed to promote structured and team-based care with the aim of achieving evidence-based quality targets (Gillam et al, 2012). However

the downside to the complexity of the QOF is the costs of administering the scheme. While a full cost-benefit analysis has not been carried out, concerns have been raised about the costs of the scheme, given the modest improvements achieved as its result, and in recent contract negotiations, the scheme has been significantly scaled back.

Ways of preventing misallocation of effort and gaming centre around devising schemes with a broader suite of targets that reflect the full range of valued outcomes. The trade offs here are around PRP schemes becoming too complex and costly to administer and monitor. Single, high-stakes indicators can be combined with 'softer' forms of performance management, such as appraisals, in order to capture less easily measurable goals. However, strategic behaviour may also occur when performance is measured inaccurately, or less accurately in relation to other performance measures (Burgess et al., 2004; Coleman, 2010), and subjective measurements of performance can also result in perceived inequity, with negative implications for employee effort and motivation.

Effects on team work

One of the characteristics of public sector activity is its reliance on collaboration and team work to achieve outcomes, meaning that it is difficult to attribute outcomes to individual effort. One unintended consequence of PRP schemes may be to damage team relations with detrimental implications for productivity and performance. There is substantial evidence on the relationship between perceived fairness in pay and performance outcomes, which shows that workers respond negatively to perceived unfairness in pay. The implication of this is that PRP systems may have detrimental effects on employee effort if they are perceived to be unfair in their design or implementation, which has been found in some studies of PRP in public services (e.g. Schmidt et al, 2011).

Such outcomes can also depend on employee characteristics. Studies in education, for example, suggest that male teachers appear to respond to PRP systems more positively than their female counterparts and are also more likely to support PRP than female teachers (Jones, 2013; Leigh, 2013). There is wider evidence that male employees tend to respond better to a competitive workplace environment and that women may be less able to put in additional effort at the margin to gain returns from PRP if they have more outside calls on their time (e.g. due to family responsibilities). This may mean that perceived unfairness of PRP systems is a particular issue for women, which has wider implications for the acceptance and success of PRP in the public sector, since the majority of staff in the NHS and in education are female (Cribb et al, 2014).

On the other hand, PRP schemes can result in improvements in team relations and organisation, as seems to have been largely the case in UK general practice as a result of the QOF. Trials of team performance targets in the UK civil service have also seen productivity improvements as a result of better management practices, with managers redistributing tasks among team members more efficiently (Burgess et al, 2010). This has led to the suggestion that PRP schemes may operate via the introduction of better management practice. Marsden (2009) suggests that PRP directly links pay with appraisal,

forcing managers to set clearer goals, which may be in some instances act as a mechanism in which employees can be involved directly in their own goal-setting. This may indicate that employee involvement in workload and development are important underlying mechanisms for agency efficiency.

Ways of addressing the issue of perceived unfairness in PRP systems include making targets as clear, objective and transparent as possible (Marsden, 2009), albeit there are trade offs here with the need to also ensure that PRP schemes consider aspects of public servants' activity that are less easily measurable but nonetheless result in valued social outcomes, as discussed above. Designing PRP systems in consultation with staff, allowing input on the goals and targets that they consider to be most important is another possible way of overcoming perceived unfairness in PRP systems (Marsden, 2009; Schmidt et al, 2011). In health, there is evidence that collaborative PRP schemes are more effective than others.

'Crowding out' of intrinsic motivations

Some of the behavioural effects discussed above may occur as a result of the 'crowding out' of (intrinsic) public service motivations due to the introduction of financial incentives through PRP. The theory is that PRP may 'crowd-out' or impair intrinsic motivation (Frey, 1997), causing a reduction of individual effort. Evidence on the occurrence of this is suggestive but inconclusive in the civil service, partly due to the difficulties of defining and measuring intrinsic motivation (e.g. Anderfuhren-Biget et al, 2010; Stazyk, 2013). In education and health there is some evidence of unease among staff about the potentially negative consequences of 'narrow' performance measures to the detriment of patient-centred care or student-centred learning. Education studies also show mixed effects of PRP on levels of collegiality and co-operative activity, which varies by scheme as well as for individuals within schemes. In (UK) health, the evidence tends to suggest positive implications of PRP for structured and team-based care and broad staff support for performance targets that are felt to align with professionals' own goals.

Designing PRP systems in consultation with staff, allowing input on the goals and outcomes they consider most important has been suggested as one means of avoiding crowding out effects. It should also be noted that other ways of motivating performance may utilise intrinsic motivations more effectively than financial rewards (Ashraf et al, forthcoming; Frey et al, 2013). In healthcare, it is suggested that clinician- or patient-directed improvement strategies that are not reliant on financial rewards can be more effective in improving quality than PRP. Some of these strategies are based around strengthening professionalism, which can induce intrinsic motivation without direct financial reward.

Performance measurement problems

The second set of challenges depicted in Figure 6.1 are related to difficulties in the measurement of outputs in public sector PRP. These include:

- negative effects of particular performance measures (e.g. absolute/relative, linear/threshold) and
- managerial subjectivity in assessment

These issues suggest challenges in the design and implementation of PRP rather than a challenge to the underlying logic. The evidence shows that different ways of measuring and rewarding performance can have implications for participant behaviour. For example, relative measures that reward improvement are more likely to stimulate improvement among poorer performers; threshold measures are more likely to stimulate improvement among median performers to raise them above the threshold than either high or low achievers. Matching the measures to the overall performance improvement goal sought is therefore very important.

Measures that rely on manager subjectivity in performance assessment and awarding rewards can cause perceptions of unfairness, leading potentially to crowding-out effects, diminished effort and motivation, and diminished teamwork or cooperation. Similar effects may result from false ranking, a form of performance management where a manager or supervisor is forced under a PRP scheme to rank their subordinates, which can result in anomalies whereby even high performing workers are ranked low and do not receive a reward.

Ways of designing schemes to address these problems include:

- Ensuring targets align with overall improvement goals and combining measures where appropriate (e.g. linear above a minimum threshold, or combination of attainment and improvement)
- Making targets as clear, objective and transparent as possible (Marsden, 2009);
- Employing a balance between objective and subjective performance measures;
- Ensuring managers work collaboratively to make decisions on PRP;
- Using complementary performance management systems which assess less easily measurable goals and outcomes;
- Involving employees in discussions about the most appropriate metrics and performance measures;
- Clarifying objectives and engaging employees more directly with the goals of the organisation (Marsden 2009: 8).

Sub-optimal or worse social outcomes

A final challenge is around the outcomes of PRP, and occurs when the incentivised outputs do not satisfactorily represent valued or desired social outcomes from the service. As discussed earlier, this is a specific challenge in public sector PRP because of the complexity of the public service 'good' and the variety of stakeholders with an interest in service outcomes.

Some of the solutions discussed above are relevant here, such as including a wide range of outcome measures that adequately capture the desired outcomes of the service. However,

this potentially negative outcome of PRP may not arise within the timescale of an evaluation since these wider social outcomes (eg mortality rates, wellbeing, and labour market outcomes) are inevitably longer term. This suggests a role for longer-term monitoring and evaluation of PRP schemes in the public sector to evaluate their impact on longer-term outcomes and to ensure alignment of PRP outcomes with other policy goals.

Alongside this, improved cost-benefit analysis of PRP schemes is also important to assess not just whether schemes are effective in meeting desired social outcomes but also whether they do this *cost-effectively*, given that governments are accountable both for performance and for the use of taxpayers' money, especially in a context of fiscal constraint where there is increasing emphasis on value-for-money in public services. Some reviews in health suggest that other (non-financial) means of improving care quality may be more effective than performance-related pay and robust comparative cost-effectiveness analysis is needed to assess the relative merits of different ways of driving quality improvements.

Table 6.1: Potential solutions to PRP challenges in the public sector

Challenges	Potential ways of addressing
Misallocation of effort and strategic/gaming behaviour	<ul style="list-style-type: none"> • Devise schemes with a broad suite of targets that reflect the full range of valued outcomes. • Combine single 'hard' measures of performance with 'softer' forms of performance management, such as appraisals, to capture less easily measurable goals.
Perceived unfairness resulting in negative effects on team relationships	<ul style="list-style-type: none"> • Use team targets • Make targets as clear, objective and transparent as possible • Design PRP systems in consultation with staff, allowing input on the goals, targets and means of measuring performance
Crowding out effects	<ul style="list-style-type: none"> • Design collaborative PRP systems to ensure that targets align with professional goals • Consider other, non-financial, forms of quality improvement and ways of motivating performance e.g. strengthening professionalism
Performance measurement problems	<ul style="list-style-type: none"> • Ensuring targets align with overall improvement goals and combining measures where appropriate (e.g. linear above a minimum threshold, or combination of attainment and improvement) • Clear, objective and transparent targets • Employ a balance between objective and subjective performance measures • Ensure managers work collaboratively to make decisions on performance • Involving employees in discussions about the most appropriate metrics and performance measures; • Clarify objectives and engaging employees more directly with the

goals of the organisation

Sub-optimal/worse social outcomes

- Include wide range of outcome measures that adequately capture the desired outcomes of the service.
- More long-term monitoring and evaluation of public sector PRP schemes to evaluate their impact on longer-term outcomes and to ensure alignment with other policy goals.
- Improved cost-effectiveness analysis to assess whether PRP schemes are cost-effective in meeting desired social outcomes.
- Comparative cost-effectiveness analysis of different improvement strategies and means of motivating performance to ensure value for money

References

- Alshamsan, R., Majeed, A., Ashworth, M., Car, J., and Millett, C. (2010). "Impact of pay for performance on inequalities in health care: systematic review" *Journal of Health Services Research & Policy*, 15 (3), 178-184
- Anderfuhren-Biget, S., Varone, F., Giauque, D. and Ritz, A. (2010). "Motivating Employees of the Public Sector: Does Public Service Motivation Matter?" *International Public Management Journal*, 13 (3), 213-246
- Ashraf, N., Bandiera, O. and Jack, B (forthcoming). "No margin, no mission? A Field Experiment on Incentives for public service delivery." *Journal of Public Economics* http://www.people.hbs.edu/nashraf/NoMarginNoMission_JPubE.pdf
- Atkinson, A., Burgess, S., Croxson, B., Gregg, P., Propper, C., Slater, H. and Wilson, D. (2009) "Evaluating the Impact of Performance-Related Pay for Teachers in England" *Labour Economics*, 16 (3), 251-261
- BBC "Atlanta Educators in 'cheating for bonuses' scandal", *BBC News*, 2nd April 2013 <http://www.bbc.co.uk/news/world-us-canada-22005837>
- Belfield, C. and Heywood, J. (2008) "Performance Pay for Teachers: Determinants and Consequences" *Economics of Education Review*, 27 (3), 243-252
- Binderkrantz, A. and Christensen, J. (2012) "Agency Performance and Executive Pay in Government: An Empirical Test" *Journal of Public Administration Research & Theory*, 22 (1), 31-54
- Boardman, C. and Sundquist, E. (2009) "Toward Understanding Work Motivation: Worker Attitudes and the Perception of Effective Public Service" *American Review of Public Administration*, 39 (5), 519-535
- Bregm, K. (2013) "Detrimental Effects of Performance-Related Pay in the Public Sector? On the Need for a Broader Theoretical Perspective" *Public Organization Review*, 13 (1), 21-35
- Burgess, S., Propper, C., Ratto, M. and Tominey, E. (2004) *Incentives in the Public Sector: Evidence from a Government Agency* The Centre for Market and Public Organisation, Working Paper Series No. 04/103 <http://www.bristol.ac.uk/cmppo/publications/papers/2004/wp103.pdf>
- Burgess, S., Propper, C., Ratto, M., von Hinke Kessler Scholder, S. and Tominey, E. (2007) *Smarter task assignment or greater effort: what makes a difference in team performance?* The Centre for Market and Public Organisation

<http://www.psi.org.uk/pdf/2007/MarisaRatto.pdf>

- Burgess, S., Propper, C., Ratto M., von Hinke Kessler Scholder, S. and Tominey, E. (2010) "Smarter Task Assignment or Greater Effort: The Impact of Incentives on Team Performance" *Economic Journal*, 120, 968-989
- Cardona, F. (2007) *Performance Related Pay in the Public Service: in OECD and EU Member States* Paris: OECD
- Chung, S., Palaniappan, L., Wong, E., Rubin, H., and Luft, H. (2010) "Does the Frequency of Pay-for-Performance Payment Matter? - Experience from a Randomized Trial" *Health Services Research*, 45 (2), 553-564
- CIPD (2013) *Annual Survey Report 2012: Employee Attitudes to Pay*
http://www.cipd.co.uk/hr-resources/survey-reports/employee-attitudes-pay-2012.aspx?utm_medium=email&utm_campaign=&utm_content=research
- Cohn, A., Fehr, E., Herrmann, B., Schneider, F. (2011) *Social comparison in the workplace: Evidence from a field experiment*, Discussion paper series, Forschungsinstitut zur Zukunft der Arbeit, No. 5550, <http://nbn-resolving.de/urn:nbn:de:101:1-201104133729>
- Coleman, T. (2010) "Do financial incentives for delivering health promotion counselling work? Analysis of smoking cessation activities stimulated by the quality and outcomes framework" *BMC Public Health*, 10, 167-172
- Courty, P. and Marschke, G. (2004) "An empirical Investigation of Gaming Responses to Explicit Performance Incentives" *Journal of Labour Economics*, 22 (1), 23-56
- Cribb, J., Disney, R. and Sibieta, L. (2014) *The Public Sector Workforce: Past, Present and Future*, IFS Briefing Note BN145, London: Institute for Fiscal Studies
- Department for Education (2014) "New advice to help schools set performance-related pay"
<https://www.gov.uk/government/news/new-advice-to-help-schools-set-performance-related-pay>
- Eijkenaar, F., Emmert, M., Scheppach, M., & Schöffski, O. (2013) "Effects of pay for performance in health care: A systematic review of systematic reviews" *Health Policy*, 110 (2/3), 115-130
- Elovainio, R. (2010) *Performance incentives for health in high-income countries: key issues and lessons learned*, World Health Report, Background Paper 32
- Emmert, M., Eijkenaar, F., Kemter, H., Esslinger, A. and Schoffski, O. (2012) "Economic evaluation of pay-for-performance in health care: a systematic review" *European Journal of Health Economics*, 13, 755-767

- Figlio, D., and Kenny, L. (2007) "Individual Teacher Incentives and Student Performance" *Journal of Public Economics*, 91 (5/6), 901-914
- Flodgren, G., Eccles, M., Shepperd, S., Scott, A., Parmelli, E., and Beyer, F. (2011) "An overview of reviews evaluating the effectiveness of financial incentives in changing healthcare professional behaviours and patient outcomes" *Cochrane Database of Systematic Reviews* (7), doi: 10.1002/14651858.cd009255
- Forest, V. (2008). "Performance-related pay and work motivation: theoretical and empirical perspectives for the French civil service" *International Review of Administrative Sciences*, 74 (2), 325-339
- Frey, B. (1997) *Not just for the money: An economic theory of personal motivation* Cheltenham, Brookfield: Edward Elgar
- Frey, B., Homberg, S. and Osterloh, M. (2013) "Organizational Control Systems and Pay-for-Performance in the Public Service" *Organization Studies*, 34 (7), 949-972
- Fryer, R. (2011) "Financial Incentives and Student Achievement: Evidence from randomised trials" *The Quarterly Journal of Economics*, 126 (4), 1755-1798
- Frølich, A., Talavera, J., Broadhead, P. and Dudley, R. (2007) "A behavioral model of clinician responses to incentives to improve quality" *Health Policy*, 80 (1), 179-193
- Gächter, S. and Thöni, C. (2010) "Social comparison and performance: experimental evidence on the fair wage-effort hypothesis" *Journal of Economic Behavior and Organization*, 76 (3), 531-543
- Gillam, S., Siriwardena, A. and Steel, N. (2012) "Pay-for-Performance in the United Kingdom: Impact of the Quality and Outcomes Framework - A Systematic Review" *Annals of Family Medicine*, 10 (5), 461-468
- Gius, M. (2013) "The Effects of Merit Pay on Teacher Job Satisfaction" *Applied Economics*, 45 (31), 4443-4451
- Gneezy, U. and Rustichini, A. (2004) "Gender and Competition at a Young Age" *American Economic Review*, 94 (2), 377-381
- Goodman, S., and Lesley, T. (2009) "Group Incentives for Teachers: The Impact of the NYC School-Wide Bonus Program on Educational Outcomes" Columbia University, Department of Economics, Discussion Papers, 0910-05
- Gravelle, H., Sutton, M. and Ma, A. (2010) "Doctor behaviour under a pay for performance contract: treating, cheating and case finding?" *Economic Journal*, 120 (542), F129-F156

- Green, C. and Heywood, J. (2010) "Profit Sharing and the Quality of Relations with the Boss" *Labour Economics*, 17 (5), 859-867
- Greene, J. (2013) "An Examination of Pay-for-Performance in General Practice in Australia" *Health Services Research*, 48 (4), 1415-1432
- Haney, W. (2000) "The Myth of the Texas Miracle in Education" *Education Policy Analysis Archives*, 8 (41), <http://epaa.asu.edu/ojs/article/view/432/828>
- Hasnain, Z., Manning, N. and Pierskalla, J. (2012) *Performance-related Pay in the Public Sector: A Review of Theory and Evidence* Policy Research Working Paper, 6043, April 2012, The World Bank
- Herzberg, F., Mausner, B., Snyderman, B. (1959) *The Motivation to Work* (2nd Ed.) New York: John Wiley
- Houle, S., McAlister, F., Jackevicius, C., Chuck, A. and Tsuyuki, R. (2012) "Does Performance-Based Remuneration for Individual Health Care Practitioners Affect Patient Care?" *Annals of Internal Medicine*, 157 (12), 889-899
- Hurley, J., DeCicca, P., Li, J. and Buckley, G. (2011) *The Response of Ontario Primary Care Physicians to Pay-for-Performance Incentives* Centre for Health Economics and Policy Analysis (CHEPA), McMaster University, Hamilton, Canada, Centre for Health Economics and Policy Analysis Working Paper Series, 2011-02
<http://cheпа.org/docs/working-papers/11-02.pdf>
- Incomes Data Services (2013) *Pay progression in the public sector: a research report for UNISON*
<https://www.unison.org.uk/upload/sharepoint/Briefings%20and%20Circulars/IDS%20report%20on%20progression.pdf>
- Jones, M. (2013) "Teacher Behavior under Performance Pay Incentives" *Economics of Education Review*, 37, 148-164
- Kantarevic, J. and Kralj, B. (2013) "Link between Pay for Performance Incentives and Physician Payment Mechanisms: Evidence from the Diabetes Management Incentive in Ontario" *Health Economics*, 22 (12), 1417-1439
- Kahn, C., Silva, E. and Ziliak, J. (2001) "Performance-based Wages in Tax Collection: The Brazilian Tax Collection Reform and its Effects" *The Economic Journal*, 111 (468), 188 - 205
- Kontopantelis, E., Doran, T., Gravelle, H., Goudie, R., Siciliani, L. and Sutton, M. (2012) "Family Doctor Responses to Changes in Incentives for Influenza Immunization under the U.K. Quality and Outcomes Framework Pay-for-Performance Scheme" *Health Services Research*, 47 (3, pt1), 1117-1136

- Kontopantelis, E., Springate, D., Reeves, D., Ashcroft, D., Valderas, J. and Doran, T. (2014) "Withdrawing performance indicators: retrospective analysis of general practice performance under UK Quality and Outcomes Framework" *British Medical Journal*, 348 (7945), (doi: 10.1136/bmj.g330)
- Lavy, V. (2004) *Performance Pay and Teachers' Effort, Productivity and Grading Ethics* NBER Working Paper No. 10622, <http://www.nber.org/papers/10622>
- Lavy, V. (2009) "Performance Pay and Teachers' Effort, Productivity, and Grading Ethics" *American Economic Review*, 99 (5), 1979-2011
- Lazear, E. (2000) "Performance, Pay and Productivity" *American Economic Review*, 90 (5) 1346-1361
- Leigh, A. (2013) "The Economics and Politics of Teacher Merit Pay" *CESifo Economic Studies*, 59 (1), 1-33
- Lester, H., Matharu, T., Mohammed, M., Lester, D. and Foskett-Tharby, R. (2013) "Implementation of pay for performance in primary care: a qualitative study 8 years after introduction" *British Journal of General Practice*, 63 (611), E408-E415
- Li, J., Hurley, J., DeCicca, P. and Buckley, G. (2011) *Physician Response to Pay-for-Performance: Evidence from a Natural Experiment*. National Bureau of Economic Research Inc, NBER Working Papers: 16909
<http://search.ebscohost.com/login.aspx?direct=true&db=ecn&AN=1235935&site=ehost-live><http://www.nber.org/papers/w16909.pdf>
- Locke, E. and Latham, G. (1990) *A Theory of Goal-Setting and Task Performance*, Englewood Cliffs, NJ: Prentice Hall
- Lundstrom, U. (2012) "Teachers' Perceptions of Individual Performance-Related Pay in Practice: A Picture of a Counterproductive Pay System" *Educational Management Administration & Leadership*, 40 (3) 376-391
- Marsden, D. (2009) *The Paradox of Performance Related Pay Systems: Why Do We Keep Adopting Them in the Face of Evidence that they Fail to Motivate?* Centre for Economic Performance (CEP), Discussion Paper No 946, <http://cep.lse.ac.uk/pubs/download/dp0946.pdf>
- McDonald, R., Harrison, S., Checkland, K., Campbell, S. and Roland, M. (2007) "Impact of financial incentives on clinical autonomy and internal motivation in primary care: ethnographic study" *British Medical Journal (International Edition)*, 334 (7608), 1357-1359

- NHS Employers (2014) *2014/15 General Medical Services (GMS) Contract Quality and Outcomes Framework (QOF): Guidance for GMS Contract 2014/15*, bma.org.uk/-/media/files/pdfs/.../contracts/gpqofguidance2014-15.pdf
- Neal, D. (2011) *The Design of Performance Pay in Education*, NBER Working Paper, 16710
- OECD (2005) *Performance-related pay policies for government employees*, Paris: OECD
http://www.oecd-ilibrary.org/governance/performance-related-pay-policies-for-government-employees_9789264007550-en
- OECD (2009) *Evaluating and Rewarding the Quality of Teachers: International Practices*, Paris: OECD
- Perry, J., Engbers, T. and Yun Jun, S. (2009) "Back to the future? Performance-Related Pay, empirical research and the perils of persistence" *Public Administration Review*, 69, 39-51
- Petersen, L., Simpson, K., Pietz, K., Urech, T., Hysong, S., Profit, J., Conrad, D., Dudley, R. and Woodard, L. (2013) "Effects of Individual Physician-Level and Practice-Level Financial Incentives on Hypertension Care: A Randomized Trial" *Journal of the American Medical Association*, 310 (10), 1042-1050
- Podgursky, M. and Springer M. (2007) "Teacher Performance Pay: A Review" *Journal of Policy Analysis and Management*, 26 (4), 909-949
- Prentice, G., Burgess, S. and Propper, C. (2007) *Performance-pay in the Public Sector: a review of the issues and evidence*, Office of Manpower Economics,
<http://webarchive.nationalarchives.gov.uk/20130106083031/http://www.ome.uk.com/downloads/Performance%20pay%20in%20the%20Public%20Sector.%20A%20review%20of%20the%20issues%20and%20evidence.pdf>
- Reilly, P. (2003) *The Link between Pay and Performance* Brighton: Institute for Employment Studies
- Review Body on Doctors' and Dentists' Remuneration (2012) *Review of compensation levels, incentives and the Clinical Excellence and Distinction Award schemes for NHS consultants*,
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/229710/DDRB_CEA_Cm_8518__3_.pdf
- Review Body on Doctors' and Dentists' Remuneration (2014) *Forty-Second Report 2014*,
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/288697/Doctors___Dentists_42nd_report.pdf
- Ryan, A. M., Blustein, J. and Casalino, L. (2012) "Medicare's Flagship Test Of Pay-For-Performance Did Not Spur More Rapid Quality Improvement Among Low-Performing Hospitals" *Health Affairs*, 31 (4), 797-805

- Sautter, K., Bokhour, B., White, B., Young, G. and Burgess, J. (2007) "The Early Experience of a Hospital-Based Pay-for-Performance Program" *Journal of Healthcare Management*, 52 (2), 95-107
- Schmidt, W., Trittel, N. and Muller, A. (2011) "Performance-related pay in German public services" *Employee Relations*, 33 (2), 140-158
- Scott, I. (2009) "What are the most effective strategies for improving quality and safety of health care?" *Internal Medicine Journal*, 39 (6), 389-400
- So, J., and Wright, J. (2012) "The Use of Three Strategies to Improve Quality of Care at a National Level" *Clinical Orthopaedics & Related Research*, 470 (4), 1006-1016
- Springer, M., Lewis, J., Ehlert, M., Podgursky, M., Crader, G., Taylor, L., Gronberg, T., Jansen, D., Lopez, O. and Stuit, D. (2010) *District Awards for Teacher Excellence (DATE) Programme: Final Evaluation Report*, Nashville, TN: National Center on Performance Incentives, Vanderbilt University
- Springer, M., Hamilton L. and McCaffrey, D. (2012) *Final Report: Experimental Evidence from the Project on Incentives in Teaching (POINT)*, Nashville, TN: National Center on Performance Incentives, Vanderbilt University
- Stazyk, E. (2013) "Crowding Out Public Service Motivation? Comparing Theoretical Expectations with Empirical Findings on the Influence of Performance-Related Pay" *Review of Public Personnel Administration*, 33 (3), 252-274
- Sutton, M., Elder, R., Guthrie, B. and Watt, G. (2010) "Record Rewards: The Effects of Targeted Quality Incentives on the Recording of Risk Factors by Primary Care Providers" *Health Economics*, 19 (1), 1-13
- Sutton, M., Nikolova, S., Boaden, R., Lester, H., McDonald, R. and Roland, M. (2012) "Reduced Mortality with Hospital Pay for Performance in England" *New England Journal of Medicine*, 367 (19), 1821-1828
- Taylor, J. and Taylor, R. (2011) "Working Hard for More Money or Working Hard to Make a Difference? Efficiency Wages, Public Service Motivation and Effort" *Review of Public Personnel Administration*, 31 (1), 67-86
- Titmuss, R. (1997) *The Gift Relationship: From human blood to social policy*, New York: The New Press
- Van Herck, P., De Smedt, D., Annemans, L., Remmen, R., Rosenthal, M. and Sermeus, W. (2010) "Systematic review: Effects, design choices, and context of pay-for-performance in health care" *BMC Health Services Research*, 10, doi: 10.1186/1472-6963-10-247
- Van Wanrooy, B., Bewley, H., Bryson, A., Forth, J., Freeth, S., Stokes, L. and Wood, S.

(2013) *The 2011 Workplace Employment Relations Study: First Findings*,
Department for Business, Innovation and Skills

Walker, S., Mason, A., Claxton, K., Cookson, R., Fenwick, E., Fleetcroft, R. and Sculpher, M. (2010) "Value for money and the Quality and Outcomes Framework in primary care in the UK NHS" *British Journal of General Practice*, 60 (574), e213-e220

Weibel, A., Rost, K. and Osterloh, M. (2009) "Pay for Performance in the Public Sector - Benefits and (Hidden) Costs" *Journal of Public Administration Research and Theory*, 20, 387-412

Woessman, L. (2011) "Cross-Country Evidence on Teacher Performance Pay" *Economics of Education Review*, 30 (3), 404-418

Young, G., Beckman, H. and Baker, E. (2012) "Financial incentives, professional values and performance: A study of pay-for-performance in a professional organization" *Journal of Organizational Behavior*, 33 (7), 964-983

Yuan, K., Le, V., McCaffrey, D., Marsh, J., Hamilton, L., Stecher, B. and Springer, M. (2013) "Incentive Pay Programs Do Not Affect Teacher Motivation or Reported Practices: Results from Three Randomized Studies" *Educational Evaluation and Policy Analysis*, 35 (1), 3-22

Appendix A: List of included studies

Reviews

Author(s) and date	Geographical context	Public service and occupations	Time period	Type of PRP scheme	Aims of study
Alshamsan et al, 2010	UK	Health, GPs	1980-2008	Mixed but mainly QOF	Impact of PRP on health inequalities
Bregm, 2012	International (OECD)	Civil Service (public sector employees)	1960s-2011	All PRP schemes in the public sector (overall review of PRP) considered. Both theoretical and empirical evidence-based discussion.	Reviews theoretical and empirical findings from experimental economics and field studies to explore whether perceived unfairness may have important negative effects on performance where PRP is implemented in the public sector.
Cardona, 2007	OECD and EU member states	Civil Service (civil service staff)	1980s-2005	All PRP schemes in the OECD and EU civil service (overall review of PRP)	Reviews whether performance measurement and PRP in the civil service are working in practice and how OECD countries have applied the PRP technique.
Eijkenaar et al, 2013	Mainly UK and US	Health	2000-2011	Mixed	Review of evidence on effectiveness of PRP; cost-effectiveness; extent of unintended consequences; effect on equity of care; effect of combining financial and non-financial incentives; and impact of different design features
Flodgren et al, 2011	international	Health	Up to 2010	mixed	the effects of financial incentives on professional practice and patient outcomes
Forest, 2008	France, USA, England	Civil Service (civil service employees)	1963-2007	All PRP schemes in the public sector (overall review of PRP)	Review of theoretical and empirical perspectives on the relationship between PRP and work motivation

				considered. Both theoretical and empirical evidence-based discussion.	and its implications for PRP in the French civil service.
Frey, et al., 2013	International	Civil Service (civil service employees)	1940-2012	All PRP schemes in the public sector (overall review of PRP) considered. Both theoretical and empirical evidence-based discussion.	Systematic review of evidence and perspectives from Behavioural Economics and Management Control Theory (as well as theoretical contributions) to examine whether negative effects from PRP are worse in the public sector than in the private sector.
Frolich et al, 2007	international	Health	1980-2005	mixed	impact of incentives on quality of care
Gillam et al, 2012	UK	Health, GPs	2004-2011	QOF –practice-level payments for a range of outcomes and recommended care across a spectrum of clinical activities	Impact of QOF
Hasnain et al., 2012	International	Civil Service (public sector employees)	1960s-2012	All PRP schemes in the public sector (overall review of PRP) considered. Both theoretical and empirical evidence-based discussion.	Reviews the theoretical and empirical literature on PRP in the public sector. It covers work from the following disciplines: public administration, psychology, economics, education, and health. The review aims to distil “useful lessons for policy-makers” in developing countries.
Houle et al, 2012	international	Health	Up to 2012	Individual incentive schemes only	the effect of PRP schemes targeting individual health care providers on patient outcomes
Leigh, 2013	International	Education, teachers	1983-2012	All PRP schemes in education	Survey of three sets of data on PRP: impact studies; surveys of teacher attitudes to PRP; surveys of public perceptions of PRP
Marsden, 2009	UK	Civil Service (civil	1955-	All PRP schemes in the civil	Reviews the paradox of using PRP in British public

		service staff)	2007	service (overall review of PRP)	services when there has been evidence that it does not motivate employees.
Neal, 2011	International	Education, teachers	1985-2011	All PRP schemes in education	Review of empirical studies which evaluate PRP programmes for teachers
OECD, 2005	OECD	Civil Service (government employees; ministries/department/agencies of the central/federal government level)	1980s-2005	All types of PRP schemes in the OECD applied to ministries/departments/agencies of the central/federal government level.	Book provides a comprehensive overview of the different trends in performance pay policies across the OECD (including why and how PRP policies are being implemented, as well as apparent impacts of PRP policy), with an aim of “lessons learned” from the different successes and failures.
OECD, 2009	International - OECD	Education, teachers	1968-2009	All PRP schemes in education	Review of design, implementation and politics and outcomes of PRP in education
Podgursky and Springer, 2007	US	Education, teachers		All PRP schemes in education	Examine the case for PRP in the US K-12 public education system. Includes history of policy, theoretical arguments and empirical evaluation evidence
Scott, 2009	international	Health	1985-2008	mixed	Review evidence assessing the relative effectiveness of different quality improvement strategies
So and Wright, 2012	Primarily US	Health	2000 onwards	mixed	Impact of pay for performance on the quality of care, relative to impact of surgical safety checklists and practice guidelines.
Van Herck et al, 2010	international	Health	2004-2009	mixed	Impact of PRP on clinical effectiveness, access and equity, coordination and continuity, patient-centeredness and cost-effectiveness; effects of PRP design and implementation; effects of PRP context.

Weibel, et al., 2013	International	Civil Service (public sector workers)	1968-2007	All PRP schemes in the public sector (overall review of PRP) considered. Both theoretical and empirical evidence-based discussion. The effects of PRP on both simple and complex tasks examined, including those with quantitative and qualitative outcome measurements.	The review employs a meta-analytic review of previous experimental studies (46 high quality empirical published studies from economics and psychology) on the effects of pay for performance to determine which conditions, if any and if present, would PRP negative effect personal efforts.
Woessman, 2011	International - OECD	Education, teachers	2003	Scheme in which a teacher's base salary can be adjusted on the basis of performance	Impact of use of PRP in education on pupil performance in PISA tests

Impact studies

Author(s) and date	Geographical context	Public service and occupations	Year(s) of data	Type of PRP scheme	Aims of study	Methods and data
<u>Experimental studies – rated 5 on the Maryland scale</u>						
Ashraf et al., forthcoming	Zambia	Civil service, Agents of a public health organisation	2009-2010	Performance-related rewards (financial and non-monetary) to individual “agents” in a public health organisation working to get condoms into the community	The experiment is designed to compare the effects of monetary and non-monetary incentives on public service delivery.	Randomly assigns 205 distinct geographical clusters containing 1,222 agents to one of four groups that receive different rewards based on condom sales. Three treatment groups (randomised at neighborhood level) and one control group (no incentives, financial or otherwise).
Burgess, et al., 2010	UK	Civil service, UK tax office –	Apr 2002-	Team-based incentive PRP scheme	Experiment studies the impact of team-based	The main identification strategy employed was a triple difference design. Three teams were

		HMCE (now HMRC)	Dec 2002		incentives on productivity. The authors examine whether this effect was due to increased effort or strategic task reallocation	examined: two teams subject to the treatment, each of which had a different incentive structure, and one control team. Used data from personnel records and management information systems. Teams comprised around 110 people.
Chung et al, 2010	California, US	Health, primary care physicians	2007	Bonus payments either quarterly or annually for achieving threshold measures of performance on 15 indicators related to care quality for chronic diseases	Whether frequency of payments affected doctors' response to the pay-for-performance scheme	Primary care physicians (sample size = 179) were randomized into two study arms differing by the frequency of incentive payment - either four quarterly bonus payments or a single year-end bonus. Quality measure scores between the two arms over four quarters were compared.
Goodman and Turner, 2009	US (New York City)	Education, secondary school teachers	2007-2009	School-based incentive payments: individual teachers in participating schools offered a share in a lump sum bonus if the school met goals based on pupil achievement	Impact on student performance and teacher absences and attitudes	185 schools were randomly selected from a pool of high-poverty schools to participate in the scheme. Analysing the control and treatment groups of schools, the authors examined the programme on maths and reading exams in the first and second years following programme implementation, teacher effort, and outcomes from surveys of teachers and students.
Lavy, 2009	Israel	Education, secondary school teachers		A trial bonus scheme, paying individual teachers on the basis of their students' performance in matriculation examinations.	Impact of PRP on teachers' pedagogy and effort, on teacher's productivity as measured by students' achievements, and on teachers' grading ethics.	Schools were included in the trial programme if their matriculation rate was equal to or lower than a critical value (45 per cent). However, measurement error in this variable – with some schools erroneously allocated to the trial - enabled a quasi-randomised trial to take place. Panel data was used to estimate a difference in differences estimate within the natural experiment. A regression discontinuity design in

						a difference in differences model, comparing all treated schools to non-treated schools was also conducted.
Petersen et al, 2013	US	Health, primary care physicians	2008-9	Individual and practice-level payments for outcomes and recommended care	Impact of individual and/or team financial incentives on translation of recommended care for hypertension into clinical practice.	Four-arm cluster RCT of individual incentives, group-level incentives, both types of incentives, and performance audit and feedback. The trial operated in 12 Veterans Affairs outpatient clinics over 20 months. The trial enrolled 83 primary care physicians and 42 non-physician personnel (eg, nurses, pharmacists).
Springer et al, 2012	US (Tennessee)	Education, secondary school teachers	2006-2009	Project on Incentives in Teaching (POINT). Teachers in the treatment group were eligible for bonuses of up to \$15,000 per year on the basis of student test-score gains on the Tennessee Comprehensive Assessment Program (TCAP).	Impact on pupil performance in standardised tests	296 teachers based in Nashville volunteered for the experiment. Approximately half were randomly assigned to a treatment group and the other half to a control group. The adjusted group mean differences in scores were estimated by a linear regression (or logistic regression model for dichotomous outcomes) that controlled for randomization block.
<u>Observational studies – rated 3 or 4 on the Maryland scale</u>						
Atkinson et al, 2009	England	Education, secondary school teachers	1999-2002	Introduction of a Performance Threshold, plus an Upper Pay Scale, offering an annual bonus of £2000 to teachers who could demonstrate they had reached acceptable standards in five key areas of performance.	Impact of scheme on pupil attainment (both absolute and value-added scores)	Authors observed the change in each teacher's average test score gain between two complete teaching cycles, before and after policy introduction, and compared the results for eligible and ineligible teachers. The sample used data from 18 schools, covering 181 teachers and almost 5000 pupils. Evaluation design controlled for pupil, school and teacher effects, and adopted a difference-in-difference methodology.

Belfield and Heywood, 2008	US	Education, secondary school teachers	2000	Performance-related bonuses to individual teachers	Impact of PRP on teachers' overall pay and job satisfaction; impact of cooperative working arrangements on the likelihood of receiving PRP	Using the Schools and Staffing Survey 2000 (a sample of 41,000 regular full-time or part-time public school teachers), the authors used probit and OLS estimations to examine the links between teachers' pay, measures of satisfaction and working arrangements for teachers in districts that did and did not offer PRP. The effects of gender and union status were also assessed.
Figlio and Kenny, 2007	US	Education, secondary school teachers	1993-2000	Individual incentive payments for teachers	Impact on pupil achievement	Combines a survey of schools on their performance pay policies with data from the National Education Longitudinal Survey and the Schools and Staffing Survey. This is used to analyse the impact of PRP schemes on pupil performance, controlling for a variety of pupil and school-level data. Uses sample of 500 schools and 4,500 students.
Gius, 2013	US	Education, secondary school teachers	2007	Effect of individual and group-based PRP offered to teachers in certain US school districts	Impact of performance pay on teachers' job satisfaction	Uses data from one wave of the Schools and Staffing Survey, including over 32,000 teachers. Responses for teachers in districts which did and did not employ PRP were analysed, controlling for a variety of demographic factors.
Gravelle et al, 2010	Scotland	Health, GPs	2004-2006	QOF	Impact of QOF on gaming	Relationship between practice performance in 2004/5 and extent of disease prevalence and exceptions reported in 2005/6. Practice-level data from 932 practices in Scotland
Greene et al, 2013	Australia	Health, GPs	1995-2010	Practices Incentive Programme Performance – voluntary scheme with individual and practice-level	Impact of programme on incentivised treatments	Analysis of pre/post trends in incentivised outcomes; modelling of impact of GP participation on outcomes (panel of 1,131 GPs); and qualitative interviews with GPs about perceived impact.

				payments for providing recommended care cycles		
Hurley et al, 2011	Ontario, Canada	Health, primary care physicians	1996-2008	Performance-based incentives to physicians targeting a small number of preventive care services, plus special payments for the delivery of specified office-based and hospital services.	Whether the scheme increased the provision of the incentivized services in the target populations	Difference in difference model comparing before and after results for a treatment group of family practitioners (eligible for the bonuses) and a control group who were ineligible. Total sample of 2,185 physicians.
Jones, 2013	US	Education, secondary school teachers	2003-2007	Effect of individual and group-based PRP offered to teachers in certain US school districts	Impact of performance pay on teacher effort (work hours – total and composition), cooperation and retention	Employs data from two waves of the Schools and Staffing Survey, comparing results for teachers that operate in PRP and non-PRP districts. Total sample size of 58,000 teachers. Controls for teacher, school and district characteristics are included.
Kantarevic and Kralj, 2012	Ontario, Canada	Health, primary care physicians	2006-2010	Diabetes Management Incentive (DMI) – a voluntary scheme with individual-level payments for providing recommended diabetes management	Testing primary care physician response to the Diabetes Management Incentive, and the relationship between the programme and different compensation models	Difference in difference model to measure effect of DMI on number of patients receiving recommended treatments, comparing physicians operating under 2 distinct compensation models. Uses administrative dataset of physician payments and services, data from 1 year pre- and 3 years post-intervention. Total sample of 4,455 physicians.
Kontopantelis et al, 2012	England	Health, GPs	2005-2010	QOF	Impact of setting higher QOF targets on rates of influenza immunization and exception reporting	Compared changes in practice immunization and exception rates for four chronic conditions before and after the increase in the upper threshold immunization rate for CHD patients in 2006/07. Data from electronic practice records of all practices in England (sample size = 8,200).

Kontopantelis et al, 2014	England	Health, GPs	2004-2012	QOF	Impact of withdrawing some targets from QOF on recorded quality of care	Compared outcomes after withdrawal of indicators with pre-intervention trends for range of indicators. Data from Clinical Practice Research Datalink (CPRD) (which holds electronic patient records used in approx a fifth of all English practices). Total sample size = over 5,000,000 patients.
Li et al, 2011	Ontario, Canada	Health, primary care physicians	1996-2008	Performance-based incentives to physicians targeting a small number of preventive care services, plus special payments for the delivery of specified office-based and hospital services.	Whether physicians' responses to the scheme differed by age, practice size and baseline compliance level.	Assesses the impact of P4P incentives within a difference-in-differences framework, by comparing the responses of the GPs exposed to the P4P incentives against those not exposed, controlling for selection on observables and unobservables. Uses an administrative data source covering nearly all the services provided by practicing primary care physicians in Ontario. Core sample of 2,185 GPs.
Ryan et al, 2012	US	Health, hospitals	2004-2009	Bonuses paid to top performing hospitals on a composite quality measure for a range of clinical diagnoses and procedures. In the 2 nd phase, bonuses were paid for improvement as well as attainment.	Whether the introduction of 'value-added' measures of performance incentivized the lowest-performing hospitals to improve	Examined changes in quality performance from phase 1 to phase 2 of the demonstration using a difference in differences approach, with a matched sample of non-participating hospitals for the comparison group. Sample size = 500 hospitals (250 in each group).
Springer et al, 2010	US (Texas)	Education, middle school teachers	2008 - 2010	District Awards for Teacher Excellence (DATE) Programme – individual incentive payments to teachers	Impact on pupil achievement, teacher turnover and teacher attitudes	Results were analysed using data from participating, formerly-participating and control schools (selected based on similarities with DATE schools). During the first year of the programme, 203 districts participated, during Year 2, 191 of the original districts participated. For the student

						outcomes, around 600,000 observations per year (2005/6 to 2009/10) were analysed.
Sutton et al, 2010	Scotland	Health, GPs	2000-2006	QOF	Impact of introduction of QOF and aspects of its design	Compared recording of range of risk factors, both incentivised and unincentivised, before and after introduction of QOF. Data from the Scottish Programme for Improving Clinical Effectiveness in Primary Care (SPICE-PC) which holds individual patient records. Data from 315 practices, which comprises over 9,000,000 observations for 391,323 individuals in each of up to 6 years.
Sutton et al, 2012	NW England	Health, UK hospitals	2007-2010	Advancing Quality pilot programme. Top performing hospital trusts were paid bonuses. In subsequent phases, improvement was rewarded as well as attainment.	The impact of the Advancing Quality programme on patient mortality	Difference-in-difference model comparing mortality rates for 3 conditions in participating hospitals (total of 134,435 patients in 24 hospitals) with mortality rates for the same 3 conditions in non-participating hospitals (722,139 patients in 132 hospitals) and with mortality rates for 6 other (unincentivised) conditions at both sets of hospitals (241,009 patients in 156 hospitals).
Young et al, 2012	New York, US	Health, primary care physicians	1999-2004	Individual payments to physicians on basis relative ranking on a composite performance measure for diabetes-related tests and screens	Effect of financial incentives on motivation of physicians and interaction with professional values.	Used practice data (3 years pre- and 3 years post-intervention) and survey of participants to test effect of financial incentives on performance and the interaction of financial incentives with work autonomy and goal importance measures. Sample of 337 physicians.
Yuan et al, 2013	US	Education, secondary school teachers	2006-2010	Compares results from three randomised control trials, offering i) individual incentives on the basis of improved pupil performance (POINT); ii)	Impact on teachers' attitudes and behaviours (including hours works, stress and collegial relationships)	Analysis of survey results for teachers in 3 schemes, comparing programme and control group teachers' reported changes in practice. A two-level mixed-effects hierarchical linear model used to analyse differences within each

group-based incentives on the basis of improved value-added (PPTI); and iii) school-based incentives on the basis of multi-measure scores (including student test scores, graduation rates, student attendance, and school environment) (SPBP).

programme. POINT included 283 teachers randomly assigned to programme and control groups; PPTI included 78 teaching teams that were randomly assigned; and SPBP included 402 schools randomly assigned.

Process studies

Author(s) and date	Public service and occupations	Year(s) of data	Geographical context	Type of PRP scheme	Aims of study	Methods and data
Anderfuhren-Biget, et al., 2010	Civil service, public servants in Swiss municipalities	Year of survey not specified	Switzerland	N/A	Authors seek to understand what the added value is of Public Service Motivation (PSM) compared to other competing factors, such as material incentives, in explaining work motivation.	Primary survey data collected. 3,754 public servants from 279 municipalities took part (38.1% response rate). Secondary analysis was undertaken using Structural Equation Modelling.
Binderkrantz and Christensen, 2011	Civil service, agency heads in Danish central government	2000-2008	Denmark	Agency-wide PRP schemes that give financial rewards to agency heads.	The study tests the hypothesis that incentive effects (PRP) will improve public sector management.	The study tests whether there is a positive relationship between agency performance and executive pay for the heads of Danish central government agencies. The data consists of the full set of performance contracts and chief executive contracts for Danish agency heads between 2000 and 2008. The analysis uses

						annual reports of agency performance and the salaries of individual agency heads, with multivariate analyses, using a GLS model.
Boardman and Sundquist, 2009	Civil Service, managers in state health and human services agencies	Fall 2002 and Winter 2003	USA	N/A	The study introduces a new explanatory variable, Perceived Public Service Efficacy (PPSE), quantifying public servants' perception about the contribution their employing agency makes for public benefit. This variable is used to explore the relationship between PPSE, role ambiguity, job satisfaction and organisational commitment.	Data is from a national survey of managers in state health and human services agencies. Sample includes managers in information management in these agencies. 274 managers responded (53% response rate). A series of ordinary least squares regressions were ran to determine the impact of PPSE on organisational commitment, role ambiguity and job satisfaction.
CIPD, 2013	Civil Service, UK public sector workers	2012	UK	N/A	The annual survey explores employee attitudes to, and expectations of, pay and bonuses, including public sector employee attitudes to PRP.	The survey is of 3,016 working adults across all sectors (private and public) and is representative of the UK working population in relation to sector, size and industry type.
Coleman, 2010	Health, GPs	2003-2005	England	QOF	the effect of QOF smoking indicators on the delivery of primary care smoking cessation interventions	Review of observational data (simple before/after design, based on data from over 4 million sets of primary care medical records contained in The Health Improvement Network research database) on effect of QOF on GPs' smoking cessation practices, plus review of literature.
Elovainio, 2010	Health		Mainly UK and US	mixed	Descriptive overview of PRP schemes in different	Literature review

international contexts						
Lester et al, 2013	Health, GPs	2012	England	QOF	To ascertain GPs' and healthcare teams' views on QOF 8 years after introduction	Qualitative interviews with 47 staff (GPs, managers, nurses and admin staff) across 23 practices randomly selected from among those participating in a pilot of potential new QOF indicators in 2012
Lundstrom, 2012	Education, Upper-secondary teachers	2003-2004	Sweden	Individual PRP for teachers, as part of a scheme for school development agreed between municipalities and teaching unions	To explore teachers' perceptions of the practical implications of a PRP system once it has been implemented	Qualitative semi-structured interviews with 23 teachers at four upper-secondary schools in three different municipalities. Teachers were interviewed twice, with a year between interviews.
MacDonald et al, 2007	Health, GPs	2005-2006	NW England	QOF	Effect of the QOF on practice organisation, clinical autonomy, and internal motivation of doctors and nurses working in primary care	Observation and interviews in 2 practices over 5 months during 2 nd year of QOF
Sautter et al, 2007	Health, hospitals	2005	US	Hospital-level annual incentive payment depending on performance in composite measure of patient safety, community outreach and selected quality indicators	To understand how improvements were achieved and how the organizational context affected this	Semi-structured interviews with senior staff (managers and cardiologists) in ten hospitals involved in the scheme. Sampled hospitals across four performance categories ('always optimal', 'improved to optimal', 'improved' and 'not improved').

Schmidt et al., 2011	Civil Service, employees in German municipalities	Post-2005	Germany	A typology of different PRP schemes were created and then examined	The study examines the recent introduction of PRP in German public services. It looks at how different designs of PRP schemes, and the circumstances under which PRP is implemented, influence PRP functionality and acceptance by employees, including PRP's effects on employee motivation.	The study provides an analysis of 215 works and establishment agreements; 17 case studies in municipalities of the federal state of North Rhine-Westphalia, including employee attitude surveys; and interviews with experts from the employers' federation and trade unions.
Stazyk, 2013	Civil Service, city managers, assistant city managers, and department heads in U.S. local government jurisdictions	2007	USA	N/A: PRP is captured as a general variable by asking whether the following variable pay systems exist in the organisation: skill-based pay, spot awards, competency-based pay, team-based pay, gain-sharing, and specifically, pay-for-performance.	The study aims to explore the relationship between performance-related pay, Public Service Motivation (PSM), and employee job satisfaction.	Sample taken from Phase IV of the National Administrative Studies Project (NASP-IV) and the International Public Management Association for Human Resources 2007 Total Compensation Benchmarking Survey. NASP-IV is a multi-method study, which includes a nationwide survey. Sample consisted of 3,316 potential respondents. Departments surveyed included Finance/Budgeting, Public Works, Personnel/HR, Economic Development, Parks and Recreation, Planning, and Community Development. Ordered logistic regression and a series of Monte Carlo simulations used to evaluate the relationship between PRP, PSM and employee job satisfaction.
Taylor and Taylor, 2011	Civil Service, full-time public employees	2005	15 countries, including US, Great Britain	N/A	The study explores the relationship between wages, PSM and the effort of government workers. The study investigates whether wages or	Secondary data analysis of (9,961) full-time public employees from 15 countries in the 2005 International Social Survey Programme (ISSP) Work Orientations dataset, samples d from full-time government employees. The

and Australia

PSM matter more to effort.

study uses an elasticity measurement: the responsiveness of a change in one variable (effort) on another variable (wages). It also tests the elasticity of effort with respect to PSM.

Contact details

The Work Foundation
21 Palmer Street
London SW1H 0AD

info@theworkfoundation.com
www.theworkfoundation.com

All rights reserved © The Work Foundation (Lancaster University). No part of this publication may be reproduced, stored in a retrieval system or transmitted, in any form without prior written permission of the publishers. For more information contact info@theworkfoundation.com. The Work Foundation Alliance Limited, London, SW1H 0AD, UK. Registered Charity No. 1146813.
No. 7746776. www.theworkfoundation.com

21 P e
Registe

University of Lancaster, Bailrigg, Lancaster LA1 4YW, UK.

