

Incentives in the Public Sector: Evidence from a Government Agency

Simon Burgess*, Carol Propper*,
Marisa Ratto[^] and Emma Tominey[^]

*University of Bristol, CMPO and CEPR
[^]CMPO

March 2004

Abstract

This paper addresses a lack of evidence on the impact of performance pay in the public sector by evaluating a pilot scheme of incentives in a major government agency. The incentive scheme was based on teams and covered five different targets, measured with varying degrees of precision. We use data from the agency's performance management system and personnel records, plus matched labour market data. We focus on three main issues: whether performance pay matters for public service workers, what the team basis of the scheme implies, and the impact of the differential measurement precision. We show that the use of performance pay did have a significant effect on the main quality measure (job placements), but that there was significant heterogeneity of response. This heterogeneity was patterned as one would expect from a free rider *versus* per monitoring perspective. We found that the incentive had a substantial positive effect in small teams, and a negative response in large teams. We found little impact of the scheme on quality measures, and we interpret this as due to the differential measurement technology. We show that the scheme in small teams had non-trivial effects on output, and our estimates suggest that the use of incentive pay is much more cost effective than a general pay rise.

Keywords: Incentives, Public Sector, Teams, Performance, Personnel Economics

JEL Classification: J33, J45, D23

Acknowledgements

This work was funded by the Department for Work and Pensions (DWP), the Public Sector Productivity Panel, the Evidence-based Policy Fund and the Leverhulme Trust through CMPO. The views in this paper do not necessarily reflect those of these organisations. Thanks to individuals in the DWP for helping to secure the data for us, particularly Storm Janeway, Stavros Flouris and Phil Parramore. Thanks for comments sent to seminar participants at Bristol, The Public Economics Working Group Conference at Warwick, the IIES at Stockholm, University of Melbourne, HM Treasury, CPB in The Hague, Tinbergen Institute and Department for Work and Pensions.

Address for Correspondence

Simon Burgess
Department of Economics
University of Bristol
12 Priory Road
Bristol
BS8 1TN
Simon.burgess@bristol.ac.uk

1. Introduction

Governments employ a lot of people – in the US around 20m people work for the government, and in Britain about 3.5m do so¹. The productivity of these workers, forming such a substantial fraction of the labour force (15% in the US, more than are employed in manufacturing at 13%) is therefore a major issue for these economies. Beyond their simple numerical importance, many governments have an explicit agenda of improving the efficiency of public service delivery². Examples include Osborne and Gaebler’s influential (1993) book “Reinventing Government”, promoted by Vice-President Gore in the US, and a UK Government White Paper, “Modernising Government” (1999). One way often proposed to boost productivity in the public sector is to use explicit financial incentives. But while theorists have addressed the role of such incentives in the public sector, a number of surveys have noted that the advance of theory has outstripped the available evidence³.

This paper begins to fill the gap. We evaluate a pilot programme of the use of team-based financial incentives in a large UK public agency. The agency, Jobcentre Plus, is one of the main government agencies dealing with the public; its role is to place the unemployed into jobs and administer benefits. The nature of the incentive scheme allows us to investigate a number of important issues in the theoretical literature. First, what is the impact of an explicit financial incentive scheme on public sector workers? Dixit’s (2002) review of theoretical contributions suggests that such incentives may be counter-productive. Second, what is the impact of a team-based incentive scheme? Economists have typically been skeptical of a team basis for obvious free-rider problems⁴, and such effects have been estimated (Gaynor and Pauly (1990); see also Gaynor et al, 2001). Knez and Simester (2001) argue that peer monitoring outweighed such effects in a scheme in Continental Airlines based on large teams. The incentive scheme we analyze was introduced across teams of very different sizes and structures,

¹ Total public employment for the US comes from OECD (2001) and relates to all levels of government, in 2000. For the UK (not reported in the OECD report), the data are from Black, Herbert and Richardson (2003) and again relate to 2000.

² Concern is not limited to developed countries: see the World Development Report (2003) also highlights the problem

³ See Dixit (2002), Prendergast (1999) and Burgess and Ratto (2003) for recent surveys.

and this allows us to quantify the impact of these factors. Third, how do workers respond to relative task measurement precision in an explicitly multi-tasking environment? The implications of multi-tasking for scheme design are a major part of the literature on incentives. The incentive scheme in Jobcentre Plus incorporated five targets, covering most of the tasks of the agency. These were measured with very different degrees of precision, and we can evaluate the response to these different targets. There are a number of other issues that we plan to investigate using this data, including the scope for gaming the system, but these are saved for future work.

Our results suggest that the scheme was effective in raising job placements in certain contexts, though the overall average effect was close to zero. We find significant heterogeneity of response that fits with important free rider effects in production. The impact of the incentive scheme is greatest in small offices and in districts with fewer offices. Thus while some mechanism such as peer monitoring does overcome the free-riding problem in small teams, it appears not to do so in large teams. We find that quantity increased, but that the scheme had little effect on the quality of service. The results suggest that relative measurement precision in a multi-tasking context is important. It is inherent in the scheme design that quantity is measured much more precisely than quality, and we find very different impact of the incentives on these outcomes. The scheme design was not optimal in a number of ways and these are briefly discussed in the Conclusion.

The paper is organised as follows. The next section briefly reviews the evidence on public sector incentive schemes. Section 3 describes the nature of the organisation and the incentive scheme introduced; it also sets out our modelling framework and identification strategy. Section 4 introduces the data. Section 5 presents estimation results and in section 6 we use these to evaluate the scheme. Section 7 concludes.

2. Literature

Recent surveys of the large literature on incentives in organisations are available in Prendergast (1999), Malcomson (1999), Murphy (1999), Dixit (2002) and Chiappori

⁴ Holmstrom (1982) provides the formalisation.

and Salanié (2003); see also Lazear's (1999) overview of personnel economics. Most of the research in this field is theoretical. Much of the empirical work relates to CEOs and the importance of share options. One notable exception to this is Lazear's (2000) study of the introduction of performance pay. There is very little evidence on the role and impact of performance pay in the public sector; Burgess and Ratto (2003) provide a review. In this section we briefly discuss some of this evidence, relating specifically to public sector organisations, and to team-based schemes.

One important source of data in the public sector has been the Job Training Partnership Agency (JTPA) programme. Under the JTPA system, local training centres receive monetary rewards based on the employment levels and wage rates attained by graduates of the programme. Heckman, Smith and Taber (1996) investigate the incentive to 'cream-skim' in the scheme by only taking the best training candidates. By contrast, they find that people with lower expected earnings are significantly more likely to be accepted into the programme. Courty and Marschke (1997, 2004) examine how the incentive structure of the JTPA programme leads to gaming of the system. They show that the structure affects the way in which the programme administrators report outcomes and the effect of their timing strategies on efficiency. Kahn, Silva and Ziliak (2001) examine the impact on the Brazilian tax collection authority of the introduction of performance pay. The reform involved the payment of financial incentives based on individual and team performance in detecting and fining tax evaders. The amounts involved were substantial, frequently providing bonuses over twice mean annual salary. The findings show that the scheme had a dramatic effect: fine collections per inspection were 75% higher than in the estimated counter-factual.

Turning to explicitly team-based schemes, Knez and Simester (2001) look at a firm-level incentive scheme introduced by Continental Airlines in 1995, offering bonuses to some 35,000 staff. One would expect free-riding to dominate any reaction to the scheme, since any one individual's influence on the outcome is bound to be negligible. Identification is achieved by comparing outsourced airports with non-outsourced airports, and the results show that performance did in fact increase by more in the latter. The positive effect is interpreted as the result of mutual (peer) monitoring. This is often thought to be ineffective in large firms, but Knez and

Simester argue that peer monitoring worked in Continental because employees worked in relatively small autonomous groups, within which monitoring and enforcement of group norms can be sustainable.

The importance of mutual monitoring and group norms has also been identified by Hamilton, Nickerson and Owan (2003). The authors provide an empirical analysis of the relationship between team incentives, worker participation, worker heterogeneity and productivity. They use data at a garment factory (Koret) in California, which between 1995 and 1997 offered the opportunity to its workers to engage in team production. The analysis, based on weekly productivity data over the years 1995-1997 for 288 employees, shows that the introduction of teams was associated with a 14% increase in productivity, on average. Moreover, some workers joined teams despite an absolute decrease in pay suggesting that teams offer non-pecuniary benefits to workers. These findings would suggest the presence of non-pecuniary benefits from teamwork, which might alleviate the free riding problem. Falk and Ichino (2003) provide experimental evidence on the importance of peer pressure in a work setting. They show that people working together tend to work at a similar rate, and that on average peer pressure tends to raise productivity.

Another part of this literature looks at professional practices and partnerships. Gaynor and Pauly (1990) investigate the determinants of productive efficiency in partnerships based on individual responses to compensation structures. The context is one with different partnerships having (endogenously) different correlations between compensation and productivity. They showed that output was greater where compensation was more directly related to productivity, and that practices with more members produced less output. Kandel and Lazear (1992) set out a model of peer pressure, norm creation and free-riding in this context. They discuss the importance of unit size in some detail and show that peer pressure rises to some point with size, and then declines. Encinosa, Gaynor and Rebitzer (1997) construct a model of the extent of performance pay in compensation contracts, allowing for group norms. This model is tested on medical group practices and it is found, as predicted, that the size of the group has a significant influence on the distribution of compensation. More recently, Gaynor, Rebitzer and Taylor (2001) look at incentives for doctors in HMO (health maintenance organisation) practices and find significant effects. Differences in HMO

size are key to identifying the impact of incentives, these being assumed random. So the variation in team size is what is driving differences in the degree of sharpness of incentives.

3. Modelling

(a) The Structure of Jobcentre Plus

The role of Jobcentre Plus (JP) is to help place people into jobs, to advise on training and to administer benefits. It was launched in October 2001, amalgamating the functions of two agencies: the Benefits Agency (BA), responsible for administering benefits to the unemployed, lone parents and others, and the Employment Service (ES), responsible for job placement. The method of delivering these services also began to change with 56 new ‘Pathfinder’ offices providing an integrated service, combining the work of the original, separate, benefits offices and employment offices. This process of change is slow, and most offices – there are 1464 in total – remain single service providers as ex-BA or ex-ES offices; more Pathfinder offices were created through the year of the pilot scheme. There are few operational links between offices in a district⁵ - that is, the work of non-Pathfinder offices is largely unaffected by the presence of Pathfinder offices in the district, and different offices are largely self-contained. The districts with the new-style offices were designated Pathfinder districts; at April 2002, these made up 17 out of 90 districts in total.

(b) The Nature of the Incentive Scheme

The initial drive for the introduction of financial incentives was political, originating in the White Paper “Modernising Government” (1999). This was followed up in the Makinson report (2000) for the Public Sector Productivity Panel, advocating incentive schemes for front line government workers; this study evaluates one of these schemes. The pilot incentive scheme ran from April 2002 to March 2003. The main relevant features of the scheme are as follows.

⁵ The incentive scheme introduces behavioural links.

Teams

This is a team-based scheme. The unit chosen as the basis for the team was the district: the targets are defined at district level, all workers in the district get the bonus if the target is hit, and the district manager is responsible for achieving the target.

These are big teams - there are only 90 districts covering the whole of the country, varying in size from 5 to 39 offices in the team, and from 264 and 1535 people within a team. We consider below issues of free-riding, peer monitoring and coordination across offices⁶.

The pilot scheme introduced the incentive structure in the 17 Pathfinder districts, leaving 73 districts as controls. So all offices in the Pathfinder districts were incentivised as part of the district team, and no offices in the control districts were incentivised. This raises two issues for identification: how districts were chosen as Pathfinder districts, and how we can distinguish between the effects of the incentive scheme and the effects of the introduction of Pathfinder offices. We discuss these below.

Threshold incentive payment

In common with many schemes, the form adopted was a step function, based on a threshold level of performance. Workers were paid straight salary up to the threshold, the bonus was paid for hitting the threshold, and then no further increase in remuneration for further output. Thus incentives are very sharp at output levels just below the threshold but weaker further below or above. The scheme was originally set up to offer 1% of salary for each target hit plus an additional 2.5% if all were hit, though the details evolved slightly through the period of the pilot⁷. The targets for the incentivised districts were set as percentage increases on the previous year⁸.

Multiple targets

⁶ See Knez and Simester (2001) for a recent paper analysing the impact of incentives within big teams.

⁷ The final bonus was actually based around standard rates, which varied with the job grade per target hit. If all five targets are achieved there is an extra 50% of the standard rate. This means that if all five targets were hit, a band A worker would earn an extra £750, whereas a band G job would get £3,750 more.

⁸ Note that all districts have clear targets set for all functions, but in the control districts these were not incentivised. The terminology of JP describes these base goals as targets and the higher levels as

One central issue in the design of incentive structures is the importance of multi-tasking. In particular, a trade-off between quantity produced and quality is often crucial⁹. This incentive scheme recognized that and included targets for five different functions. These were: job placements, customer service, employer service, other business delivery functions, and reducing benefit calculation error and fraud. Teams were offered 1% of salary for each target hit (provided a minimum of 2 were hit) and a further bonus of 2.5% on top of that if all five were hit. Clearly the definition and measurement of the target variables is important.

Job placements, job entries in JP's terminology, are measured as weighted numbers of clients who are found work by the office; this is the main quantity output measure. The weights are points based on a priority system imposed by the government – for example, a jobless lone parent attracts 12 points, compared to 4 points for a short-term unemployed claimant, and 1 point for an already-employed worker¹⁰. Extra points also accrue if the worker is not back on unemployment benefit within four weeks, and also in certain priority areas. Customer service captures aspects of quality: speed, accuracy, pro-activity of service, and the nature of the office environment. This relates to both clients and employers. It is measured by independent analysis of questionnaires to employers and 'mystery shopping' techniques. The employer outcome target is the flip side of job placements, a measure of whether and how quickly vacancies were filled. This was measured (again independently) by a survey of employers. It only has low correlation with placements. The business delivery target covers a wide range of other functions, and appears to be an attempt to measure everything else that the offices do. It includes two targets for benefit calculation accuracy, appropriate labour market interventions, and basic skills and incapacity screening. It is measured by checking samples of cases. The overall score on this performance measure is simply the average over the five categories. The final target is the Monetary Value of Fraud and Error, focused on two particular benefits – Income Support and Jobseeker's Allowance. This is measured by specialist teams visiting each district and examining samples of cases. In fact, the measurement and tracking of this particular target was and remains obscure. All 17 Pathfinder districts were treated

'stretch'. In this paper we keep to the standard economics terminology and describe the higher levels of output required to win the bonus as the targets.

⁹ See Paarsch and Shearer (2000) for an analysis of this issue.

as a single virtual region, and the reporting of the progress on achieving the target was very delayed. We ignore this particular target in the analysis below.

Hierarchy: measurement, reward and production.

One final relevant characteristic of the scheme is that the different targets were measured at different levels of the JP hierarchy. Job Entries were measured monthly at office level, Customer Service and Employer Outcome were measured quarterly at district level, and Business Delivery at district level at irregular intervals. This measurement structure has implications for our econometric analysis, but also for the likely behavioural response of workers in JP, which we discuss below. Note that this produces quite a complex hierarchical structure. Essentially the office is the core unit in terms of production (that is, there appear to be few operational reasons for cross-office interaction), the district is imposed as the decision-making unit in terms of reward, and the measurement structure works off the office for some targets and the district for others.

When output is measured at office level but assessed for bonus purposes at district level, the district manager assumes a crucial role in communicating how well offices are doing relative to the district threshold. The allocation by district managers of office-level targets may help alleviate the free-rider problem if the manager had good information about office capabilities, but could be counter-productive if based on poor information. When output is measured only at district level, there would not be the same information flow on progress. This suggests that there are differences in the free-rider issue within offices and between offices within districts. We therefore consider these separately in the empirical work.

(c) Theoretical issues

The design of an optimal incentive scheme is a complex matter. The nature of the organisation, the size of the team, the measurability of output, the multidimensionality and the nature of tasks are all elements to be considered in the design of team-based incentives and in any evaluation of a scheme. We consider the implications for worker behaviour of the way the scheme has been designed at JP. Of course, incentive

¹⁰ For full details on this and the other targets see Burgess et al (2003) cmpo dp.

schemes also impact on selection of workers into organisations (see Lazear, 2001, Dxiit, 2002, Besley and Ghatak, 2003), but the timescale of this pilot and the relatively low staff turnover suggest that in this context the main effects will come through changes in the behaviour of incumbent workers.

Team size and structure

Teams are defined to be very large, all the offices in a district. Note that this ‘team’ is created by the reward system only. Individual rewards depend upon the performance of the district, but there is no production basis for a district level team: whilst staff interact within offices, there is little need for interactions between team members located in different offices. Such a broad definition of teams may make it hard for team members to identify with their teams. This is likely to intensify free rider problems. Holmstrom (1982) provides one of the seminal contributions to the theory of incentives in teams and shows that a negative externality is created in an environment in which output is fully shared among team members. The agent who cheats will not pay in full for the consequences of her act, and hence she will choose an inefficiently low level of effort. This free-rider problem becomes more difficult to tackle the greater the uncertainty in output measurement and the greater the size of the team. In fact the free rider problem in the context of JP is not exactly equivalent to Holmström’s case, which is caused by the final output being fully shared among team members. The issue in JP is more akin to voluntary contributions of effort to a public good outcome (hitting the target). The theoretical predictions both from the Holmström paper and the literature on voluntary contributions to public goods¹¹ suggest that the greater the number of agents the more serious the free-rider problem. As noted above, Kandel and Lazear (1994) consider how the off-setting strength of peer pressure varies with unit size.

Multi-tasking and the Measurement Technology

JP is a complex organisation and staff are required to deliver a range of services. Theory suggests that this matters for the outcome of the scheme. In particular, if the different activities are substitutes, the use of high powered incentive schemes may have undesirable effects upon overall performance. Exerting more effort on one task

¹¹ See Olson (1971)

increases the marginal cost of any task that is a substitute and agents may focus their efforts upon one or a few tasks to the neglect of others. In this case each outcome cannot be rewarded in isolation and the principal should use lower incentives (Holmström and Milgrom, 1990, 1991).

An interesting case related to this situation is when activities are substitutes from the perspective of the agents (more time spent on one activity means less time on others), but they are complements from the perspective of the principal (the principal wants high performance in all of them). Hence the agent is willing to devote more time to the less difficult activities, whereas the principal prefers him to devote time to all activities. This situation is analysed by Marx and MacDonald (2001). They show that, if the principal is unsure about the agent's preferences over tasks, setting rewards on success on individual tasks may be suboptimal in that it may induce workers to focus and specialise in the less costly tasks.

It is likely that both positive and negative interdependencies are present in the JP context. For example, good performance on customer service target and employer outcome may have spillover effects on job entries¹². In contrast, more time spent on accurately processing benefit claims leaves less time to be devoted to job placements.

A crucial aspect in a multi-tasking context is how precisely the different dimensions of output are measured. If each outcome could be rewarded in isolation, the optimal scheme would set higher incentives on the better measured outcomes¹³. However, in a context with multiple dimensions of output, this would lead to a misallocation of effort by the agent. Therefore the principal has to weaken the incentives on the more accurately measured tasks. The prediction of the standard models on moral hazard when output is measured with error is that low powered incentive schemes should be used when the different outcomes are measured with differential precision (Dixit, 2000).

¹² There are two parts to this. First, a simple accounting feature that placing a job-seeker in a vacancy notified by a firm to the Jobcentre Plus office directly boosts performance on both targets. Second, a behavioural factor that understanding the customers' requests, meeting their individual needs and giving them accurate information (the "proactivity and accuracy" elements of the Customer Service target) may speed up the process of filling vacancy and facilitate the creation of job entries.

¹³ There is a trade-off between risk and incentives. See Prendergast (2002), Dixit (2000), for a general discussion.

The five targets in the JP scheme involve very different measurement precision. The main quantity factor, job entries, is measured most precisely – it is a well-defined concept, and is measured directly from the management information database monthly at office level. By contrast, the quality of service to job-seekers and employers is measured by sample survey. Furthermore it is measured at district level and quarterly. This can only give a much vaguer measure of a worker's effort on these tasks.

What is the optimal response of an employee given this reward structure and measurement technology? The rewards for hitting each target are the same, but the cost of employee effort on quantity and quality is unknown to us, as is the relative effort required to hit the target. It may be that these are known to the senior management of JP and are factored into the design of the scheme. In which case, workers will allocate their effort in line with the principal's optimum – possibly equally across tasks. If this is assuming too high a degree of sophistication in setting the parameters of the incentive scheme, absent substantial differences in effort costs across targets, we would expect a worker to focus more effort on the quantity target, because of the lower noise.

Threshold reward scheme

Given the difficulty of relating effort to measured performance on some targets, and that team bonuses are paid whenever two targets are hit, we can expect to observe gaming. Offices may focus their attention on a few targets rather than aiming to hit all five targets. Furthermore, the threshold nature of the scheme with additional performance beyond the target not being compensated means that workers will aim to just hit the target.

(d) Model

Output process

Focussing on the quantity target first, we adopt a standard job matching approach to the determination of job entries. Job entry points (*JEP*) in office *o* at time *t* depend on the available vacancies (*V*) and jobseekers (*U*), plus a set of office characteristics

$(W)^{14}$. The key variable for our purposes is the job placement effort of the office staff, α . This takes the following form, where v is noise:

$$JEP_{ot} = \alpha_{ot} V_{ot}^{\gamma} U_{ot}^{\delta} W_{ot} \exp(v_{ot}) \quad (1)$$

We allow the output of quality measures to depend on the same set of factors, though this is less obviously based in a standard economic approach. For quality of service, staff effort has an obvious impact, as potentially do a variety of office characteristics. The state of the labour market may well also matter. It may be that it has a direct impact on the quality of service, and it may be that it has an impact on perceptions of quality in that tight (slack) labour markets may lead to employers (workers) blaming the JP office for their lack of a match.

Effort Process

Bringing all aspects of the incentive scheme into a single model would be very cumbersome, for the gain of little additional insight. Instead we adopt a more straightforward approach. We briefly review the model for an individual's decision on a single output, and then discuss the extensions to a two-level team structure, and a multiple target setting. We omit subscripts here for clarity.

Let individual output be given by:

$$y = a.e + Z + \varepsilon \quad (2)$$

where y is output, a is the worker's ability, e is the worker's effort, Z is an exogenous, time-varying factor influencing output (the state of the labour market), and ε is noise.

Z has mean Z^e and is given by $Z \equiv Z^e + \xi$. The distribution function of $(\xi + \varepsilon)$ is F .

Worker utility is simply given by $U = w - c(e)$, where $c(e)$ is the cost of effort. The pay scheme is:

$$w = \bar{w} + k.\mathfrak{I}(y > \bar{y}) \quad (3)$$

where \bar{w} is straight salary, k is the bonus and $\mathfrak{I}(\cdot)$ equals 1 if (\cdot) is true, 0 otherwise.

The bonus threshold is assumed to be:

$$\bar{y} = \lambda y^0 = \lambda(a.e^0 + Z^e), \quad \lambda > 1 \quad (4)$$

with y^0 (e^0) being the 'traditional' or standard output (effort) level. So $(\lambda - 1)$ is the additional output required for the bonus. The worker's problem is to choose e to

¹⁴ See Petrongolo and Pissarides (2001) for a review of evidence on this.

maximise utility before ξ and ε are realised. It is useful to define x as the extra effort above e^0 : $e \equiv e^0 + x$. The solution for x^* is:

$$c'(e^0 + x^*) = k.a.f(-a.x^* + (\lambda - 1)(a.e^0 + Z^e)) \quad (5)$$

This equates the marginal cost of effort and the marginal benefit. The latter is the value of the bonus, k , multiplied by the effect of an extra unit of effort on the probability of hitting the threshold, and by ability.

So optimal individual effort depends on: two common policy parameters (k , the size of the bonus, and $(\lambda - 1)$ the required increase); environmental parameters (Z^e and $f(\cdot)$, the mean and variability of ξ and ε); and individually-varying parameters ($c'(\cdot)$, the cost of effort, a , ability and e^0 , past effort). Having set out this basic model, we consider two dimensions in which the pilot scheme is more complex than this.

Broadening out to the office level raises the issues of free-riding within the office, and free-riding across offices within districts. One simple way of summarising this is to assume that we can multiply the level of individual effort derived above by a factor β depending on office size, N , and on office and district characteristics, θ . This gives us a final effort equation:

$$e^* = \beta(N, \theta). \bar{e}^*(k, \lambda - 1, c'(\cdot), a, e^0, Z^e, f(\cdot)) \quad (6)$$

Office level effort, α , is simply the sum of this over all workers in the office. This can be substituted into (1) to determine office output.

This describes optimal effort on one task in isolation, but clearly individuals and offices need to consider the allocation of effort to the full range of tasks. This would involve solving jointly for the optimal effort on each task, as a function of the marginal benefit and cost of effort on all of the tasks. In general, this would mean that the determinants of optimal effort on any one task would be vectors of the set of variables on (6) defined across all tasks. However, the empirical implementation of this is less straightforward. The simple marginal benefit to each task is the same at 1% of salary per target hit. The relative marginal cost of effort on each task, past effort and worker ability are unobserved and all we can do is assume them to be constant across tasks. The state of the labour market is the same for all tasks. There is a clear difference between processes in the measurement precision, but that difference does

not vary across offices. Therefore the only aspect of the trade-off between effort on the different tasks that we can investigate is that arising from the system-wide difference in measurement precision.

(e) Empirical Model

Our empirical approach derives from (1) for the production function and (6) for the determination of effort. We address three main questions. First, is the behaviour of public sector workers influenced by financial incentives? Second, does free-riding matter in a team-based incentive scheme? We sub-divide this into the free-riding deriving from many workers in an office, and into that arising from many offices in a district. Third, does the differential measurement precision of the different targets influence behaviour?

Our dependent variables are (log) job entry points, and two measures of quality: quality of service to job seekers (“customer service”) and the quality of service to firms (the quality component of “employer outcome”). We take the inflow of new job-seekers and new vacancies as measures of the state of the labour market and use the (new vacancy/new job-seeker) ratio¹⁵. Office characteristics are the number of staff (see below), a measure of staff ability, whether the office is a Pathfinder office (established or created in-period), and whether the office is the district headquarters. We adopt a log functional form for job entry points, staff and the labour market characteristics.

We take two approaches. First, we run this model on each office’s annual total of job entry points. This provides our first main results. Second, we adopt a two-stage approach, run the regression on monthly data over the year to isolate an office average effect and then analyse that in relation to the office and district characteristics. This has the advantage of allowing us to estimate separately the two effects of the labour market (see below). The first stage is:

$$y_{odt} = (\mu_o + \Delta_d + \gamma S_d) + \beta N_{ot} + \alpha Z_{ot} + \delta_t + v_{ot} \quad (7)$$

¹⁵ The inflow of vacancies may be considered potentially endogenous – efficient offices attract more vacancies. We repeat the main analysis using just the job-seeker inflow, and the conclusions are largely unchanged (results are available from the authors).

where y is log total job entry points (tjep) in office o in district d at month t , N is the number of staff, and Z is the labour market variable. We allow for an office effect μ , a district effect Δ , and an effect from IS – incentive scheme status. Finally, δ is a set of time dummies, and ν is random noise. A fixed effects regression on (7) above will identify α , β , δ , and ϕ_o where:

$$\phi_o \equiv \mu_o + \Delta_d + \gamma IS_d \quad (8)$$

which is the second stage regression, supplemented by mean N and mean Z .

(f) Identification

We adopt two strategies for identification. Recall that assignment to the pilot scheme is at district level, and was based on a district being designated a Pathfinder district – a district with at least one Pathfinder office in. There are two points here. First and most important, assignment of offices other than the Pathfinder office itself to the pilot is random. Those offices are in the pilot on grounds entirely unrelated to their own performance and characteristics¹⁶. The second issue is how those Pathfinder offices were chosen. There are two factors in this – offices and districts. Individual offices were chosen for Pathfinder status on the grounds that their management would be able to cope well with the demands of the new structure – clearly this is likely to be correlated with other outcomes. From the set of such offices, districts were selected to be representative, designed to achieve a “cross-section of different communities and customer bases, i.e. from large inner-city offices to those in smaller towns, suburbs and rural areas.”¹⁷ This suggests that assignment at district level to the treatment category is stratified random. The two mechanisms together imply that for offices other than the Pathfinder office itself, assignment to the scheme at is random. Given random assignment, straightforward regression on the model above is then appropriate.

Second, to allow for any residual non-random assignment we also implement a propensity score matching approach. The assumption of mean independence

¹⁶ Offices are only linked together in districts through spatial proximity, not through performance levels, for example. Of course, there may be spatial factors influencing performance, but we control for the main factors here.

¹⁷ Private communication.

conditional on the estimated score is weaker than in the first strategy. As we explain below, the data do not permit a difference-in-difference analysis.

4. Data

We take data from JP's management information system and from their personnel database. We merge on to this unemployment and vacancy data from the local labour market, and also data on the local public/private wage differential, as a control for differences in the quality of staff.

Management information data records performance against the five targets. Job entry points (JEP) achieved for each office on a monthly basis are the measure of quantity. The quality outcomes are reported for each district on a quarterly basis. A basic description of this data is on Appendix 1. It shows wide variation in JEP across offices and time, but much less variation (and fewer observations) for the survey-based measures of quality.

This is predominantly a human capital intensive organisation, and we have detailed staff data: the number of staff in each grade for each office, recorded monthly. These are also described in Appendix 1. The numbers in different grades appear in more-or-less fixed proportions. For example, there is about one Executive Officer (EO) to two Administrative Officers (AO). Consequently, including numbers of each grade in the analysis leads to severe multicollinearity. We therefore define a measure of front-line staff: the office total of all numbers in EO and AO grades for use in the analysis.

Clearly the quality of the workers is an important consideration and there is no reason to expect it to be constant across the country. Traditionally, public sector jobs pay less than private sector jobs, and perhaps the key margin for JP workers is between these jobs and private sector jobs. We therefore take as a proxy for quality the local public/private sector wage differential (see Nickell and Quintini, 2002, for a detailed discussion of this). Our variable is derived from the Labour Force Survey Small Areas dataset. We constructed the wage gap between the private sector and the public sector

for each local authority looking at the relative hourly wage of full-time workers. This was matched to the office postcode.

We have no information on the state of the capital (principally computing and communications equipment) in offices. We do know which offices are Pathfinder offices, however. It is important to identify these for three reasons. First, they do have newer technology installed and generally refurbished premises. Second, they were also subject to restructuring in which the managers had to oversee the convergence of ex-ES and ex-BA offices. JP has estimated that Pathfinder offices took at least five months to adjust. Third, even beyond the adjustment period, Pathfinder offices fulfil more roles than regular ex-ES or ex-BA offices; consequently we would expect their productivity as measured on any one task to be lower. We separately identify the first tranche of Pathfinder offices created by October 2001, which would therefore have completed the process of readjustment by the start of the incentive scheme in April 2002, and Pathfinder offices created later.

Using the postcode (zip code) for each JP office, we locate each office in a Travel To Work Area¹⁸ (TTWA98). We then extract claimant inflow and vacancy inflow data from NOMIS¹⁹ for each TTWA and for each month. Whilst the matching function uses the unemployment and vacancy stocks, we cannot take these as exogenous as they are influenced by the outflow rate, our dependent variable. So we use the inflow, both of unemployed claimants and of vacancies, and take the latter divided by the former. Note that the state of the labour market plays two roles – first it provides the ‘raw material’ necessary for the office to produce job entries. Second, it proxies labour market tightness and hence the ease or otherwise of placing claimants in jobs.

The incentive scheme ran from April 2002 to March 2003, and this is the period of our data. Note that although Jobcentre Plus employees were informed about incentive scheme in April 2002, they did not know the specific targets until June 2002. It would obviously be very desirable to have data before the scheme was implemented to allow a difference-in-difference technique. Unfortunately this is simply not possible – the

¹⁸ These are largely self-contained local labour markets, defined by 75% of those living there also working there, and 75% of those working there also living there. There are some 400 covering Britain.

¹⁹ National Online Manpower Information Service, <http://www.nomisweb.co.uk/>.

district boundaries were re-drawn in 2002, and different PSA targets were in operation before April 2002, implying a different set of output measures.

5. Estimation results

We present results first for the main quantity variable, then for the quality measures; we finally estimate them jointly, allowing for production interdependencies.

(a) Quantity – job entry points

Annual Total

Figure 1 shows the distribution of the annual job entry points totals, and unconditional comparisons across different office and district types. Comparing offices in non-incentivised districts with non-Pathfinder offices in incentivised districts is closest to a like-with-like comparison, and we see that the distributions are fairly similar. Pathfinder offices are clearly associated with lower mean job entry figures.

We need to control for a number of different factors in order to isolate any potential incentive effect. Table 1 presents the results for the (log) annual totals of job entry points. We present a number of different specifications for the effects of the incentive scheme, along with the office characteristics.

We start with basic office characteristics in column 1. Big offices (defined in terms of front-line staff) produce more job entries. This is a strong and well-defined effect. The measure of the quality of staff also has a positive effect on job entries. This is measured as the private-public wage gap in the local area, and has a negative impact on job entries – a high wage gap implies lower quality workers in the public sector, including JP. District Offices (with central administrative functions) yield more job entries. An established Pathfinder office produces significantly fewer job entries than an otherwise equivalent office; an office becoming a Pathfinder office within the year

shows no effect on job entries²⁰. The state of demand in the labour market has no influence on job entries. As we shall see below, this is misleading and in fact reflects the two effects we noted above cancelling out. Once we run the two-stage approach below to isolate time-series variation from cross-section variation, we will find that labour market variation plays a strong role.

The main variables of interest relate to the incentive scheme. The simple comparison in column 2 shows that the scheme has an insignificant effect, reflecting the impression of Figure 1. However, in column 3 we allow for heterogeneity of response, possibly reflecting free-rider and peer pressure issues, by including an interaction of incentivisation status and office size. This yields a significant incentive effect, a positive effect that declines with office size. This fits the discussion above in that bigger offices face a greater free-rider problem and so the incentive payment is less effective in eliciting higher effort. In columns 4 and 5 we include the number of offices in the district²¹, and allow its effect to differ in incentivised and non-incentivised districts. It has no effect in the latter and a negative effect in the former. This suggests that there is little interaction between offices in non-incentivised districts, but that it is attempted in incentivised districts. The interaction is however far less effective in districts with many offices. Finally in column 5, we examine whether the number of high grade staff in the office has any independent effect but it appears not to. This regression explains about half of the variation between offices, and shows significant and heterogeneous effects from the incentive scheme.

Monthly data – two stage approach

The use of the annual totals rules out analysis of within-year time series variation in performance. This is essential to isolate the role of exogenous shocks to output from the labour market, and also to confirm the results in Table 1. We run fixed effects regressions on the 932 offices; time-varying variables are log front-line staff, log labour market, and a set of monthly time dummies²². The results, alongside the pooled OLS, are shown in Table 2. The pooled OLS shows a strong effect of staff and a

²⁰ This is presumably because staff in these offices are performing benefits-related activities as well as job entry tasks. The newer Pathfinder offices show no effect as they were changed on average late on in the year.

²¹ These are offices with positive job entries – not all JP offices.

positive effect of the labour market inflow ratio. The equivalent fixed effect regression shows a much diminished staff coefficient and a doubled labour market coefficient. Our interpretation of these results is that the pure time-series variation exploited in the fixed effect regressions isolates the environmental influence of labour market tightness on job entries. The cross-sectional influence of the labour market on job entries is captured by the fixed effect. Conversely, almost all of the variation in staff is across offices and very little over time within an office, and so the fixed effect regression picks up little impact of this on outcomes.

We extract the estimated office effects, and subject these to the same analysis as the annual totals, presented in Table 3. Note that these necessarily have mean zero, but we adjust them by adding back the grand mean to ensure they have the same mean as the equivalent raw data. The results are very similar. They will not be identical as the dependent variable in Table 3 is essentially the adjusted mean of the monthly log job entries, and the dependent variable in Table 1 is the log of the total job entry points (not the total of the monthly log values). One substantial difference is that in Table 3 the labour market variable is significantly negative, compared to insignificance in Table 1. This represents the cross-section variation in labour market tightness, with both the job entries and claimant inflows purged of time series variation. This negative effect is interpretable as the fact that, all else constant, areas with higher inflow of job seekers will have higher job entries. The main results on the impact of the incentivisation remain the same, or are slightly stronger.

We can also explore the role of labour market risk in this context. We expect that in markets where the noise involved in the production process is substantial, optimal effort will be lower. Accordingly we add to the specification the coefficient of variation of the labour market indicator and also interacted this with incentivisation status. The results show that the degree of risk matters significantly for all offices, but with no different effect for incentivised offices.

²² A variety of different functional forms were tried – see Burgess et al (2003).

Matching results

The above results correctly identify the effect of the scheme if allocation to the pilot incentive scheme was random as we believe. However, it is useful to consider another approach as well. We first compare the characteristics of incentivised and non-incentivised districts and offices in Table 4. This shows that incentivised districts are larger, both with more staff per office (35 compared to 29 on average), and more offices (16 compared to 12). They appear to face very similar labour market conditions.

We implement a matched estimator which yields an unbiased estimate if assignment to incentivisation status is conditional mean independent from the propensity score. We set up the propensity score matching as follows²³. Even though districts are the basis for assignment into the treated (incentivised) category, we compute propensity scores at office level because offices are the unit of analysis. We include all non-Pathfinder offices in incentivised districts and all offices in non-incentivised districts, as Pathfinder offices are unlikely to find a close match. We estimate the conditional probability of assignment to incentivisation status, based on a set of observable variables. These variables might influence choice of pilot areas and/or the outcome variables. The propensity score estimator (probit) is shown in Table 5. We see that there are some significant influences on an office's chance of being selected – the size of the office, the number of offices in the district, (both of which were controlled for above) and some regional effects. We adopt a smoothed and weighted matching estimator. Predicted probabilities from the probit are used as weights to construct a synthetic control observation for each incentivised office; the weights are proportional to an Epanechnikov kernel. We impose a requirement of common support on the match following Heckman, Ichimura and Todd (1997). All observations for the control group whose propensity score is outside the range of the propensity score in the treatment sample are deleted.

We first make a simple comparison of the overall treated sample and matched controls, using the estimated office fixed effect. This is in Table 6a and yields the

²³ We are very grateful to Barbara Sianesi for making her implementation of PS matching in *Stata* available.

average treatment on the treated. Since it is the average effect, it necessarily omits heterogeneity that we found to be important above. Unsurprisingly it yields an insignificant negative effect²⁴, as in column 2 of tables 1 and 3. However, the heterogeneity of response in this study is a central part of the argument rather than a nuisance, linked to free-rider issues and optimal team structure. Accordingly, we partition the sample into eight cells, split into the quartiles of the office size distribution, and above and below mean number of offices per district. We recalculate the matched difference separately within each of these cells; the results are in Table 6b. We find some supporting evidence of the findings above: the estimated incentive effect is generally declining in office size and is generally lower in districts with fewer offices. The effect is positive in all small and very small offices, and negative in all large and very large offices. The top left cell does not fit in with the pattern. Significance levels are low.

However, conducting these eight independent tests reduces efficiency levels. In order to better parameterise the response heterogeneity (and to control for any remaining differences between treated and control offices), we run the regressions in Table 1 on the matched sample only²⁵. This is in the spirit of Heckman, Ichimura and Todd (1997) for non-experimental data. The results are in table 7 and confirm a significant positive effect of the scheme, declining with office size. The number of offices in the district does not appear to be important here.

To summarise the findings on the impact of the scheme on the main quantity variable targeted: we find evidence of a significant positive effect of incentivisation on output, declining as the size of the team increases. This fits with previous results on free-riding and Kandel and Lazear's (1992) result of peer pressure decreasing after some critical unit size. There is also some evidence that a large number of offices in a district attenuates the impact. It is perhaps surprising that the incentive scheme generates negative as opposed to zero effects for large offices. To some extent this is misleading – only one office has a predicted incentive effect that is significantly below zero. But it may also be reflecting a real phenomenon. It may be that the

²⁴ Bootstrapped standard errors were derived by 1000 replications.

²⁵ The sample is made up of treated (non-PF) offices and matched controls using nearest-neighbour matching.

contractualisation of effort in a public service conflicts with individual's intrinsic motivation and reduces work effort. It may also be that the extra managerial effort required to implement and monitor the scheme in large efforts subtracts significantly from direct work effort; conversations with JP officials lend credence to this story.

(b) Quality – job-seeker and employer service

We adopt a similar approach to modelling quality outcomes. We choose to model the quality of service to job-seekers (JSQ) and to employers (EMQ). Recall that these outcomes are monitored differently – they are only measured quarterly (cf. monthly for quantity) and at district (cf. office) level. This has implications for behaviour as set out above, but also for our estimation: this reduces our observations from over 900 offices to just 90 districts.

Annual Average

The Appendix Table gives the mean response on these two quality measures, 84.4% success rating for JSQ and 88.6% for EMQ. The table also shows little variation in these scores across districts for JSQ, but more for EMQ. In fact, all districts hit their targets for JSQ, whilst only 64% did for EMQ.

Table 8 shows the results for the district annual averages for JSQ and EMQ. Few variables are estimated to have a significant effect, due in part to the small number of observations and the lack of variation in JSQ. For both, the number of staff in the office has a negative effect. This may arise from a more personal service in smaller offices. The tightness of the labour market has a negative impact on JSQ and a stronger negative effect on EMQ. This makes sense: a tight labour market means a difficult time for employers to fill vacancies. It may also mean that JP staff are more pressured for both client groups. There is no significant impact of any term involving incentivisation status on JSQ, and one weakly positive effect on EMQ.

Monthly data – two stage approach

We repeat the two-stage approach for these two quality outcomes. The results are in Table 9, exploiting the quarterly variation and estimating district fixed effects, and in Table 10, analysing these fixed effects. The results differ little from the annual

average estimates above, and find no effect of incentivisation on JSQ. There is a positive effect of incentivisation on EMQ, though only significant at 10%. The private-public wage gap has a negative effect on JSQ but no significant effect on EMQ.

PS Matching

We finally repeat the matching procedure outlined above on the quality outcomes. The results are in Table 11, and the regression on the matched sample is in Table 12. These results confirm that there appears to be no impact of the scheme on quality of service under JSQ and a positive effect in EMQ, significant only at 10%.

Summarising, we find little significant effect of incentivisation on quality outcomes. This can be taken in two different ways. First, it could be argued that the scheme failed to elicit any increase. This is not surprising in that the precision of the monitoring technology for quality was low (measured infrequently at a high level of aggregation), implying a low optimal effort allocation to this component of the job. Second, it could be viewed more positively as showing that despite the greater effort on quantity, quality did not actually fall, a standard failing of many incentive schemes. Whether this is due to the incentive scheme explicitly targeting quality, or whether due to the existence of sufficient slack to permit the increase in quantity, is difficult to say.

(c) Quantity and Quality together

There are two points to make in this section. First, the contrast between the significant effect of the scheme on quantity and lack of effect on quality is interesting. It may be that this arises from the differing measurement precision for the two aspects of the job, or it may simply be statistical – 90 observations in one case compared to over 900 in the other. Since this matters for incentive scheme design, we get at this by re-running the quantity regression at district level. To do this, we used as the dependent variable (log) district annual job entry points. The results are in Table 13. They continue to show a positive impact of incentivisation, declining with office size. This suggests that there is something different about the behavioural response to the quantity and quality targets. This is straightforwardly explicable in terms of the differences in the precision of the monitoring technology. It fits well with the results

of Gaynor and Pauly (1990) noted above: they show that output was significantly higher in medical practices in which compensation was more directly related to productivity.

The second issue is an estimation one. Since time allocated to quantity or to quality is determined jointly, we need to take account of that in estimation. The first step is simply to establish whether good performance on one dimension is positively or negatively correlated with good performance on the other. Using the annual totals (averages), the results are:

Correlation of district performance on quantity and quality

	Quality – JSQ	Quality – EMQ
Quantity	-0.067 -0.081 (weighted)	0.069 0.139 (weighted)
Quality – JSQ		-0.018 -0.163 (weighted)

Weighted by district staff numbers.

In fact, we see that there is little correlation. If we take EMQ as more useful a measure given the low variation in JSQ, there is a low positive association.

Second, we estimate the district-level annual quantity and EMQ models jointly using SUR. Given the low correlations noted above, we would not expect a large change in the standard errors and this is indeed the case, as Table 14 shows.

6. Valuing the impact of the incentive scheme

We evaluate the quantitative importance of the change in the quantity outcome in three ways. Since we do not find much effect on quality, we focus only on quantity. The evaluations below focus just on the incentivised districts, not the whole system.

(a) Quantification of Scheme Effect

Given the estimates in Table 3, column 4, we can straightforwardly calculate the distribution of change in job entry points associated with the incentive scheme. Since the impact varies according to office and district size, we report this distribution as

well as the mean – see Table 15. As would be expected from Tables 1 and 3 column 2, the overall effect of incentivisation is about zero. There is a substantial positive effect in small offices and districts, offset by a negative effect in larger units. However, it makes sense to quantify the impact of the scheme in the contexts in which it worked (whilst bearing in mind that the overall effect was zero) in order to appraise the potential quantitative significance of such schemes. In fact, the mean in small offices (in any district) and the mean in small districts (across all offices) are both around 10%, so this seems a reasonable value to follow up. A 10% overall increase translates to around 17,000 job entry points in small districts, or approximately converting back into people, about 3,400 extra people. This is simply a first stage summary – we look at the impact on steady-state unemployment below.

The *ex post* cost of the job entry component of the scheme was around £272,100, 0.21% of the salary bill for the 17 incentivised districts. This derives from 5 of the 17 hitting their job entry target and earning 1% of salary (allowing for different numbers of staff). All 5 who hit were small districts. The *ex ante* cost depends on the level that the target thresholds were set at, and the implicit success probability management thereby set.

(b) Impact of the Scheme on Unemployment

It is clear that the operation of the incentive scheme does not create new jobs. Nor does it help into employment people who would otherwise have remained unemployed forever. So a ‘cost per job created’ is not directly appropriate. The scheme accelerates movement into work of its clients, and one way to evaluate that is to return to the labour market model in (1) and analyse the implied unemployment rate. Given an unemployment exit rate of x , and an inflow rate of i , the steady-state unemployment rate is given by $u^* = i/(i + x)$. We assume that the inflow rate is unaffected by the JP incentive scheme, and that the exit rate is $x = k \cdot (JE/U)$, where $k > 1$ allows that not all those leaving unemployment do so to jobs, JE is job entries as modelled above, and U is the stock of unemployed. From (1), $x = k \cdot A \cdot (V/U)^{1-\alpha}$, and $A = \alpha W$, is the JP office effect, with α (office effort) depending on the incentive scheme. It is easy to show that:

$$\eta_{u^*, \alpha} = -x/(x+i) = -(1-u^*) \quad (9)$$

where η denotes elasticity. Given an overall mean effect of zero, clearly the mean effect on unemployment is also zero. However, taking the mean value of a 10% increase from the small districts, this produces a mean percentage decrease in unemployment of 9.5% given a national mean unemployment rate of 5.1% at end 2002; this is a fall of almost one half a percentage point.

(c) Effort, Quality or Quantity of Staff? Incentive scheme, general pay rise or more staff

A final metric for the evaluating the size of the incentivisation effect is to compare it to raising the quality of staff through a general pay rise, and to simply having more staff. We can do this straightforwardly through the estimated production function. Using column 4 of table 3, we compute the change in the private-public pay differential (£ per hour) required to produce a 10% increase in job entry points. This is given by $\ln(1.1)/(-0.018)$, equal to -5.29 . Thus a £5 an hour pay increase in the public sector would, through recruitment of higher quality staff, on average elicit the same output improvement as this scheme does (in small districts). Given that the average hourly pay in the organisation was £8.70 at the time, a £5 per hour increase would be extremely expensive, and way above the cost of the incentive scheme.

Turning to an increase in staff, a similar calculation shows that a 10% increase in output would require a 19.4% increase in staff and hence the salary bill. Again this is very high compared to the cost of the incentive scheme of 0.213% of the salary bill in small districts.

Note that even if we have dramatically over-stated the output effect of the incentive scheme, even a 1% increase in output would have made this scheme cost effective on the basis of the estimated parameters.

7. Conclusion

There is a dearth of evidence on the role and impact of performance pay in the public sector – a sector that employs as many people in the UK as manufacturing does. This paper starts to fill that gap by providing an evaluation of a pilot scheme of financial

incentives in a major UK government agency, Jobcentre Plus. The incentive scheme was based on teams rather than individuals and covered five different targets, measured with varying degrees of precision. It offered a maximum bonus of 7.5% of salary if all targets were hit. Using data from the agency's performance management system and personnel records, plus matched labour market data, we evaluate the impact of the scheme. We focus on three main issues: whether performance pay matters for public service workers, what the team basis of the scheme implies, and the impact of the differential measurement precision.

We show that the use of performance pay did have a significant effect on the main quantity measure (job placements), but that there was important heterogeneity of response. This heterogeneity was patterned as one would expect from a free rider *versus* peer monitoring perspective. We found that the incentive had a substantial positive effect in small offices, and in offices in small districts. In districts with many offices, and in large offices, the scheme reduced output. Our interpretation of this is that peer monitoring and better information flows can overcome free rider problems in small units, but that it fails in teams made up of many people, or dispersed among many offices (see Kandel and Lazear, 1992, and papers by Gaynor and co-authors). The impact of performance pay on quantity was not matched by any impact on different quality measures. One key difference between the quantity and quality targets is the precision of measurement. Job placements are measured monthly at office level, with a clear, direct effect of an individual's effort to the target measure. Quality measures are based on samples of different clients' experiences, and are only measured quarterly at district level. In this case, an individual's effort is only measured probabilistically, and is in any case submerged in a much broader total. Our findings suggest that individuals responded to this by focussing their effort on quantity rather than quality (see Pauly and Gaynor, 1990).

We quantify the economic significance of the scheme in a number of ways. Given the heterogeneity of response, in fact the overall mean impact is close to zero. However, we examine the size of the positive effect in the small districts where the scheme did work. We compute the implied number of additional job placements, and the reduction in the equilibrium unemployment rate. These are non-trivial numbers for a very inexpensive scheme. We also compare the cost-effectiveness of performance pay

with a general pay rise (raising worker quality) and simply the addition of more staff. Our estimates suggest that the use of incentive pay delivers equivalent output increases as the other two at very significantly lower cost.

Indeed, the mean increase in small districts of job placements was around 10%²⁶, and this may seem too high for the modest size of the bonus. There are a number of caveats that need to be mentioned. First, the scheme only operated for one year, and so the results may include a “first year” novelty effect in addition to the pure incentive effect. Furthermore, if a ‘ratchet’ design of continual percentage improvements were repeated in a dynamic setting, the optimal response would be different to the response we have measured to a possibly once-only pilot. Second, the outcome may be the result of performance management per se, rather than the financial reward attached. This seems unlikely in that the same performance management system was in place everywhere, across the control offices as well as the pilot offices. It may be that the financial incentives led managers to take the existing framework more seriously but that is surely part of the aim of performance pay. Third, Jobcentre Plus may be an organisation with a lot of slack in it. Unemployment and job-seeking has fallen considerably since the peaks of the 1980s (though it has been stable at a low level in recent years), and it may be that staff are less hard-pressed than before. Finally, it may be that the assignment to the pilot was not completely random and differentially included high-performing offices. Whilst possible, this seems to be unlikely given the nature of the assignment process. Districts were included in the pilot if one office in that district had been selected to be a Pathfinder office. Given the few operational links between offices, this is essentially random assignment for other offices in that district.

There are a lot more issues that we can address with this data. We have ignored within-year and between-office effects on output in this paper, but this seems a likely source of strategic behaviour. The awarding of differential job entry points for different client groups is also of interest, and given available data on the inflow of these groups, we can analyse the degree to which offices do respond differently to people and points. Finally, whilst we have controlled for the impact of the

²⁶ Balanced by a mean in large districts of -5%.

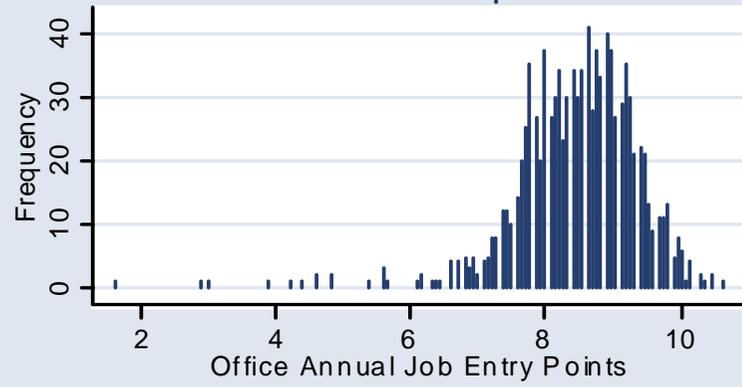
environment (the state of the labour market) on the outcome, we have not allowed differential response across offices to the ups and downs of the local labour market. These are all on our agenda.

We finally draw some tentative conclusions for the design of performance pay schemes in the public sector. There are some obvious conclusions – team size needs to be small, and preferably not dispersed over many sites. The connection between effort and output needs to be as clear and well-measured as possible. There are trade-offs here: precise measurement may be very expensive if conducted for many small teams. Less obvious conclusions are the role of environmental risk. In the context of this organisation, changes in the local labour market affect the targeted outcome: a one-standard deviation change in the labour market variable has on average a 1.1% impact on job entry points²⁷. This is not trivial at the margin, and suggests that some broad conditionality of the scheme may be useful. Finally, there are lessons for the structure of organisations as well as for the nature of optimal incentive schemes. Dewatripont, Jewitt and Tirole (1999) make this point in the context of mission definition, but it also applies here to team size and task measurement. If incentives are indeed a very cost-effective way of inducing greater output given the right team size, then organisations could be re-structured to create natural teams of the appropriate size. Such re-structuring could also allow for relative performance evaluation to filter away common uncertainty. These points also fit well with the general ethos of devolved agency inherent in many current public service reforms.

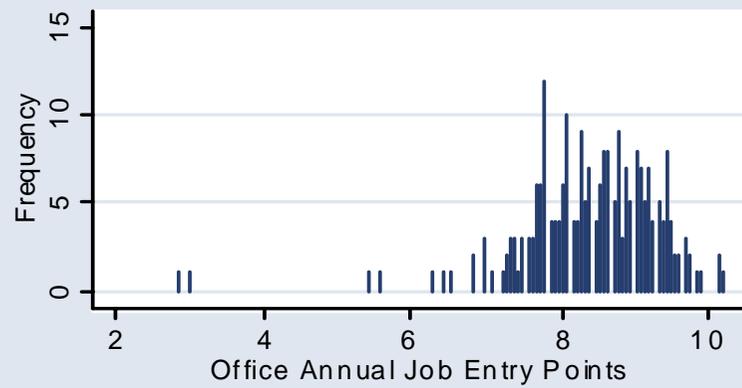
²⁷ Using the estimates from table 3, column 4, and given that the mean (across offices) standard deviation of the time series variation in our labour market variable is 0.0567.

Figure 1. Distribution of Annual Job Entry Outcome

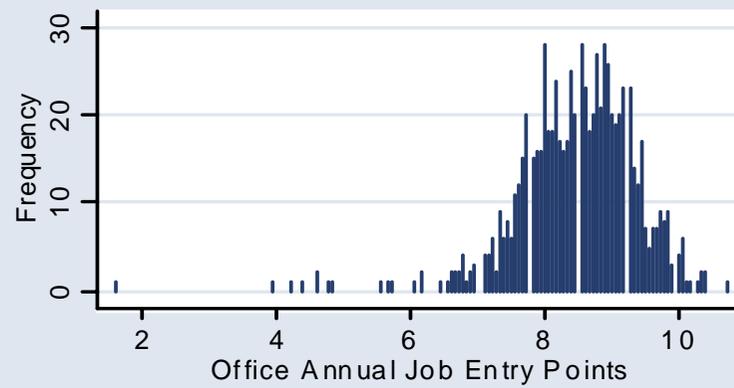
Total Sample



Incentivised Districts

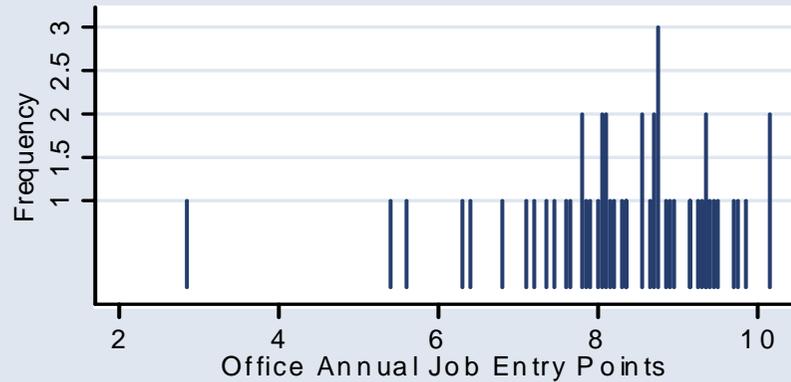


Non Incentivised Districts

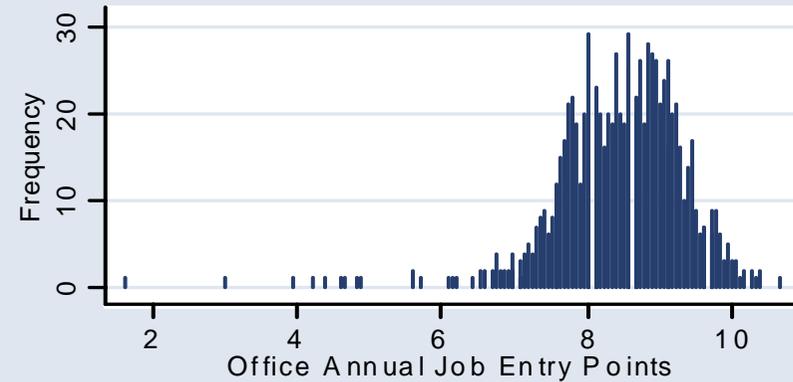


Distribution of Annual Job Entry Outcome

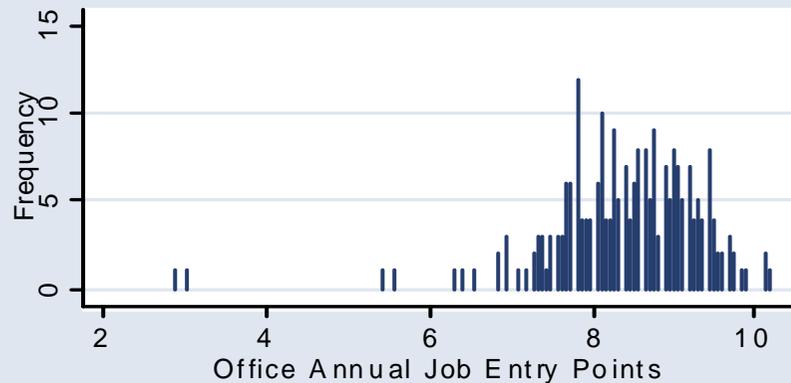
PF Offices



Non PF Offices



Non PF Offices, Incentivised Districts



Offices in Non Incentivised Districts

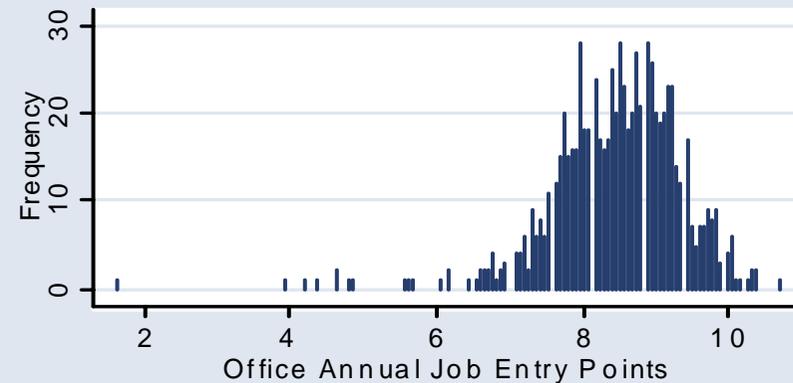


Table 1: Ooffice annual job entry analysis

Dependent variable is log (annual total of job entry points)

	(1)	(2)	(3)	(4)	(5)
Pathfinder Status	-0.607 (0.079)***	-0.579 (0.091)***	-0.520 (0.095)***	-0.585 (0.098)***	-0.582 (0.098)***
District Office	0.047 (0.068)	0.046 (0.068)	0.046 (0.067)	0.050 (0.067)	0.052 (0.067)
JCP Status	-0.053 (0.064)	-0.023 (0.079)	-0.026 (0.079)	-0.065 (0.080)	-0.063 (0.080)
Private Public Wage Gap	-0.030 (0.007)***	-0.030 (0.007)***	-0.030 (0.007)***	-0.029 (0.007)***	-0.029 (0.007)***
Log Annual Frontline Staff	0.720 (0.021)***	0.720 (0.021)***	0.743 (0.024)***	0.736 (0.024)***	0.735 (0.024)***
Log Mean Labour Market	0.033 (0.056)	0.033 (0.056)	0.039 (0.056)	0.048 (0.055)	0.048 (0.055)
Incentivisation Status		-0.035 (0.054)	0.274 (0.160)*	0.489 (0.212)**	0.481 (0.212)**
Incentivisation * Mean Frontline Staff			-0.100 (0.049)**	-0.100 (0.049)**	-0.101 (0.049)**
No. Offices per District				-0.009 (0.005)*	-0.009 (0.005)*
Incentivisation * No. Offices per District				-0.011 (0.009)	-0.010 (0.009)
Office Mean % High Grade Staff					-0.958 (0.582)
Constant	6.318 (0.070)***	6.322 (0.070)***	6.251 (0.078)***	6.370 (0.104)***	6.405 (0.106)***
Observations	932	932	932	932	932
R-squared	0.56	0.56	0.56	0.57	0.57

Standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 2: Monthly job entry first stage
 Dependent variable is log (job entry points)

	(1)	(2)
Log Frontline Staff	0.690 (0.008)***	0.199 (0.025)***
Log Labour Market	0.072 (0.018)***	0.190 (0.016)***
July 2002	-0.023 (0.028)	-0.013 (0.013)
August 2002	0.173 (0.028)***	0.187 (0.013)***
September 2002	0.099 (0.028)***	0.071 (0.013)***
October 2002	0.274 (0.028)***	0.265 (0.014)***
November 2002	0.063 (0.028)**	0.079 (0.013)***
December 2002	-0.618 (0.028)***	-0.601 (0.013)***
January 2003	0.097 (0.032)***	0.188 (0.019)***
February 2003	-0.155 (0.028)***	-0.145 (0.014)***
March 2003	-0.256 (0.028)***	-0.289 (0.013)***
Constant	3.862 (0.032)***	5.361 (0.079)***
Observations	9312	9312
R-squared	0.51	0.46
Number of offices		962

Standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 3: Analysis of the office fixed effects

	(1)	(2)	(3)	(4)	(5)	(6)
Pathfinder Status	-0.743 (0.093)**	-0.717 (0.108)**	-0.552 (0.121)**	-0.631 (0.123)**	-0.630 (0.123)**	-0.631 (0.123)**
District Office	0.058 (0.081)	0.058 (0.081)	0.054 (0.081)	0.056 (0.080)	0.056 (0.081)	0.048 (0.080)
JCP Status	-0.058 (0.077)	-0.031 (0.096)	-0.039 (0.096)	-0.121 (0.099)	-0.120 (0.099)	-0.120 (0.099)
Private Public Wage Gap	-0.019 (0.009)**	-0.018 (0.009)**	-0.019 (0.009)**	-0.018 (0.009)**	-0.018 (0.009)**	-0.024 (0.009)**
Log Annual Frontline Staff	0.462 (0.025)**	0.462 (0.025)**	0.494 (0.027)**	0.492 (0.027)**	0.492 (0.027)**	0.474 (0.028)**
Log Mean Labour Market	-0.199 (0.047)**	-0.199 (0.047)**	-0.193 (0.047)**	-0.185 (0.047)**	-0.184 (0.047)**	-0.163 (0.048)**
Incentivisation Status		-0.030 (0.065)	0.108 (0.079)	0.537 (0.176)**	0.534 (0.176)**	0.427 (0.231)*
Incentivisation * Mean Frontline Staff			-0.005 (0.002)**	-0.005 (0.002)**	-0.005 (0.002)**	-0.005 (0.002)**
Number of Offices per District				-0.003 (0.006)	-0.003 (0.006)	-0.002 (0.006)
Incentivisation * No. Offices per District				-0.025 (0.010)**	-0.025 (0.010)**	-0.025 (0.010)**
Office Mean % High Grade Staff					-0.138 (0.610)	-0.085 (0.609)
Labour Market Time Series Variation						-0.721 (0.289)**
Incentivisation * Labour Market Time Series Variation						0.390 (0.570)
Constant	4.803 (0.106)***	4.808 (0.106)***	4.703 (0.111)***	4.730 (0.136)***	4.735 (0.138)***	4.946 (0.161)***
Observations	962	962	962	962	962	962
R-squared	0.31	0.31	0.32	0.33	0.33	0.33

Standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 4: Characteristics of the districts and offices by incentive status

		% Pathfinder Office	Frontline Staff	Number of offices in District	Mean labour market conditions
Non-Incentivised Districts	Mean		288.343	12.112	1.197
	Median		262	12	1.235
Incentivised Districts	Mean	14.319	482.353	15.936	1.148
	Median	14	483	18	1.241

		Pathfinder Office (%)	Frontline Staff	Mean labour market conditions
Offices in Non-Incentivised Districts	Mean		29.000	1.213
	Median		23	1.101
Offices in Incentivised Districts	Mean	14.617	35.418	1.174
	Median	14	27	1.115

Table 5: Probit to estimate propensity score
 Dependent variable is Incentivisation status

Mean Frontline Staff	-0.032 (0.015)**
Office Frontline Staff Variance	0.025 (0.019)
Office Frontline Staff Squared	-0.000 (0.000)
Frontline Staff * No. Offices	0.001 (0.001)
Office Mean Labour Market	-0.018 (0.313)
Office Labour Market Variance	-0.340 (0.574)
Labour Market * Frontline Staff	0.018 (0.010)*
No. Offices per District	-0.615 (0.085)***
No. Offices per District Squared	0.028 (0.003)***
<u>Regional Variables</u>	
East of England	0.345 (0.355)
London	0.881 (0.344)**
North East	0.168 (0.444)
North West	0.697 (0.319)**
Office for Scotland	0.662 (0.305)**
Office for Wales	0.297 (0.336)
South East	-3.056 (0.470)***
South West	-0.701 (0.380)*
West Midlands	0.693 (0.326)**
Yorkshire	-0.122 (0.404)
Constant	1.640 (0.666)**
Observations	912
Pseudo R ² = 0.2813	

Standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6a: Matching on whole sample

Total Sample

	Off support	On support	Total
Untreated	0	739	739
Treated	65	108	173
	65	847	912

Average Treatment of the Treated	-0.052
	(0.071)

Standard error in parentheses

Table 6b: Matching separately by office and district size

		Office Size			
		Very small	Small	Large	Very Large
District Size	Small	0.0847	0.2549	-0.0382	-0.1195
		(0.128)	(0.300)	(0.108)	(0.108)
	Obs	90	98	120	129
	Large	0.2922	0.1536	-0.2052	-0.1329
	(0.084)**	(0.175)	(0.207)	(0.211)	
	obs	130	102	101	62

Standard errors in parentheses

Table 7: Annual job entry analysis on matched sample
 Dependent variable is log (annual total job entry points)

	(1)	(2)	(3)	(4)	(5)
District Office	0.137 (0.136)	0.137 (0.137)	0.115 (0.132)	0.116 (0.135)	0.131 (0.135)
Private Public Wage Gap	-0.028 (0.014)**	-0.030 (0.014)**	-0.014 (0.015)	-0.013 (0.015)	-0.016 (0.016)
Log Annual Frontline Staff	0.784 (0.041)***	0.785 (0.041)***	0.927 (0.059)***	0.925 (0.062)***	0.927 (0.062)***
Log Mean Labour Market	0.176 (0.155)	0.176 (0.156)	0.296 (0.154)*	0.293 (0.156)*	0.261 (0.159)
Incentivisation Status		0.021 (0.072)	0.817 (0.257)***	0.850 (0.318)***	0.867 (0.318)***
Incentivisation * Mean Frontline Staff			-0.264 (0.082)***	-0.260 (0.085)***	-0.264 (0.085)***
No. Offices per District				0.002 (0.013)	0.001 (0.013)
Incentivisation * No. Offices per District				-0.004 (0.019)	-0.003 (0.019)
% High Grade Staff per District					-1.184 (1.066)
Constant	5.921 (0.133)***	5.909 (0.140)***	5.470 (0.192)***	5.457 (0.219)***	5.489 (0.220)***
Observations	124	124	124	124	124
R-squared	0.76	0.76	0.78	0.78	0.79

Standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 8: District annual JSQ and EMQ analysis

Dependent variables are log (annual district average JSQ outcome) and log (annual district average EMQ outcome)

	(1) JSQ	(2) EMQ
% PF offices per District	0.004 (0.007)	-0.014 (0.007)**
% JCP offices per District	0.001 (0.003)	-0.002 (0.003)
Private Public Wage Gap	-0.008 (0.001)***	0.001 (0.001)
Log Mean District Frontline Staff	-0.026 (0.008)***	-0.021 (0.008)**
Log Mean District Labour Market	-0.024 (0.012)**	-0.059 (0.012)***
Incentivisation Status	-0.034 (0.076)	0.038 (0.125)
Incentivisation * District Frontline Staff	-0.004 (0.031)	0.021 (0.016)
No. Offices per District	0.002 (0.001)**	0.001 (0.001)
Incentivisation * No. Offices per District	-0.000 (0.001)	-0.002 (0.001)*
Constant	0.001 (0.070)	0.068 (0.070)
Observations	90	90
R-squared	0.47	0.33

Standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 9: District quarterly first stage regressions

Dependent variables are log (district JSQ outcome) and log (district EMQ outcome)

	(1)	(2)	(3)	(4)
	District Log JSQ		District Log EMQ	
	OLS	Fixed effect	OLS	Fixed effect
Log District Frontline Staff	-0.018 (0.006)***	0.039 (0.020)*	-0.011 (0.005)**	0.005 (0.038)
Log District Labour Market	0.011 (0.011)	0.017 (0.011)	-0.043 (0.009)***	0.008 (0.022)
September 2002	-0.022 (0.003)***	-0.021 (0.003)***	-0.054 (0.008)***	-0.056 (0.007)***
December 2002	-0.009 (0.003)***	-0.008 (0.004)**	-0.071 (0.007)***	-0.067 (0.007)***
March 2003	-0.004 (0.004)	-0.003 (0.003)	-0.057 (0.008)***	-0.055 (0.007)***
Constant	-0.186 (0.009)***	-0.118 (0.025)***	-0.083 (0.009)***	-0.075 (0.047)
Observations	359	359	359	359
R-squared	0.12	0.16	0.30	0.33
Number of District Identification Number		90		90

Standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 10a: Regressions on the district JSQ fixed effects

	(1)	(2)	(3)	(4)
% PF Offices per District	0.000 (0.001)	0.003 (0.007)	0.004 (0.007)	0.005 (0.007)
% JCP Offices per District	0.002 (0.003)	0.002 (0.003)	0.002 (0.003)	0.001 (0.003)
District Private Public Wage Gap	-0.008 (0.001)***	-0.008 (0.001)***	-0.008 (0.001)***	-0.008 (0.001)***
Log District Frontline Staff	-0.049 (0.007)***	-0.050 (0.007)***	-0.049 (0.007)***	-0.059 (0.009)***
Log District Mean Labour Market	-0.008 (0.007)	-0.008 (0.007)	-0.008 (0.007)	-0.014 (0.007)**
Incentivisation Status		-0.035 (0.078)	-0.037 (0.079)	-0.048 (0.079)
Incentivisation * District Mean Frontline Staff			-0.005 (0.033)	-0.010 (0.039)
No. Offices per District				0.002 (0.001)**
Incentivisation * No. Offices per District				-0.000 (0.001)
Constant	-0.240 (0.032)***	-0.234 (0.034)***	-0.233 (0.035)***	-0.253 (0.035)***
Observations	90	90	90	90
R-squared	0.65	0.65	0.65	0.67

Standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 10b: Regressions on the district EMQ fixed effects

	(1)	(2)	(3)	(4)
% PF Offices per District	0.001 (0.001)	-0.011 (0.006)*	-0.011 (0.007)	-0.011 (0.007)*
% JCP Offices per District	-0.004 (0.002)	-0.001 (0.003)	-0.001 (0.003)	-0.002 (0.003)
District Private Public Wage Gap	0.002 (0.001)	0.002 (0.001)	0.002 (0.001)	0.002 (0.001)
Log District Frontline Staff	-0.024 (0.006)***	-0.022 (0.006)***	-0.022 (0.007)***	-0.025 (0.008)***
Log District Mean Labour Market	-0.035 (0.006)***	-0.035 (0.006)***	-0.035 (0.006)***	-0.036 (0.007)***
Incentivisation Status		0.139 (0.075)*	0.139 (0.076)*	0.158 (0.077)**
Incentivisation * District Mean Frontline Staff			-0.001 (0.032)	0.026 (0.038)
No. Offices per District				0.001 (0.001)
Incentivisation * No. Offices per District				-0.002 (0.001)
Constant	-0.038 (0.031)	-0.061 (0.033)*	-0.061 (0.033)*	-0.070 (0.035)**
Observations	90	90	90	90
R-squared	0.32	0.34	0.34	0.36

Standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 11: Propensity score matching results for JSQ and EMQ outcomes

	Off support	On support	Total
Untreated	0	73	73
Treated	2	15	17
	2	88	90

JSQ	
Average Treatment of the Treated	-0.020 (0.0140)

EMQ	
Average Treatment of the Treated	0.009 (0.0109)

Table 12a: Annual District JSQ analysis on matched sample

Dependent variable is log (annual district average JSQ outcome)

	(1)	(2)	(3)	(4)	(5)	(6)
% PF Offices per District	0.001	0.001	-0.001	-0.002	0.002	0.003
	(0.001)	(0.001)	(0.010)	(0.011)	(0.010)	(0.011)
% JCP Offices per District	0.010	0.010	0.011	0.012	0.008	0.009
	(0.005)*	(0.005)*	(0.008)	(0.008)	(0.008)	(0.008)
District Private Public Wage Gap	-0.004	-0.004	-0.004	-0.004	-0.005	-0.004
	(0.002)*	(0.002)*	(0.002)*	(0.002)*	(0.002)**	(0.002)*
District Log Frontline Staff	-0.024	-0.024	-0.024	-0.033	-0.057	-0.057
	(0.010)**	(0.010)**	(0.011)**	(0.017)*	(0.019)**	(0.020)***
District Log Mean Labour Market	0.014	0.014	0.013	0.013	0.006	0.009
	(0.011)	(0.011)	(0.011)	(0.012)	(0.012)	(0.014)
Incentivisation Status			0.017	0.022	-0.009	-0.021
			(0.120)	(0.121)	(0.116)	(0.120)
Incentivisation * District Mean Frontline Staff				0.040	0.086	0.086
				(0.052)	(0.060)	(0.061)
No. Offices per District					0.003	0.003
					(0.001)***	(0.002)*
Incentivisation * No. Offices per District					-0.003	-0.002
					(0.002)	(0.002)
% High Grade Staff per District						-0.394
						(0.691)
Constant	-0.333	-0.333	-0.343	-0.368	-0.389	-0.387
	(0.064)***	(0.064)***	(0.095)***	(0.101)***	(0.097)***	(0.098)***
Observations	30	30	30	30	30	30
R-squared	0.41	0.41	0.41	0.42	0.53	0.54

Standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 12b: Annual District EMQ analysis on matched sample

Dependent variable is log (annual district average EMQ outcome)

	(1)	(2)	(3)	(4)	(5)	(6)
% PF Offices per District	0.002 (0.001)**	0.002 (0.001)**	-0.017 (0.010)	-0.019 (0.010)*	-0.017 (0.011)	-0.019 (0.010)*
% JCP Offices per District	-0.007 (0.005)	-0.007 (0.005)	0.005 (0.008)	0.006 (0.008)	0.005 (0.008)	0.003 (0.008)
District Private Public Wage Gap	0.001 (0.002)	0.001 (0.002)	0.001 (0.002)	0.001 (0.002)	0.001 (0.002)	-0.000 (0.002)
District Log Frontline Staff	-0.022 (0.011)*	-0.022 (0.011)*	-0.020 (0.010)*	-0.029 (0.016)*	-0.028 (0.020)	-0.026 (0.019)
District Log Mean Labour Market	-0.024 (0.012)**	-0.024 (0.012)**	-0.027 (0.011)**	-0.027 (0.011)**	-0.021 (0.013)	-0.030 (0.013)**
Incentivisation Status			0.220 (0.118)*	0.224 (0.119)*	0.215 (0.120)*	0.251 (0.117)**
Incentivisation * Mean Frontline Staff				0.036 (0.051)	0.070 (0.062)	0.071 (0.060)
No. Offices per District					-0.000 (0.002)	0.000 (0.001)
Incentivisation * No. Offices per District					-0.002 (0.002)	-0.003 (0.002)
% High Grade Staff per District						1.119 (0.673)
Constant	-0.046 (0.068)	-0.046 (0.068)	-0.173 (0.094)*	-0.195 (0.100)*	-0.194 (0.100)*	-0.200 (0.096)*
Observations	30	30	30	30	30	30
R-squared	0.40	0.40	0.48	0.49	0.54	0.60

Standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 13: District level annual job entry analysis

Dependent variable is log (district total annual total job entry points)

	(1)	(2)	(3)	(4)
% PF Offices per District	-0.010 (0.006)	-0.111 (0.072)	-0.081 (0.077)	-0.050 (0.066)
% JCP Offices per District	0.020 (0.028)	0.038 (0.031)	0.038 (0.031)	0.032 (0.026)
District Private Public Wage Gap	-0.023 (0.015)	-0.025 (0.015)	-0.025 (0.015)	-0.023 (0.013)*
District Log Annual District Mean Frontline Staff	0.659 (0.073)***	0.671 (0.073)***	0.710 (0.081)***	0.498 (0.081)***
District Log Mean Labour Market	0.039 (0.130)	0.026 (0.130)	0.037 (0.130)	-0.201 (0.119)*
Incentivisation Status		1.196 (0.849)	2.336 (1.334)*	2.737 (1.230)**
Incentivisation * District Mean Frontline Staff			-0.177 (0.160)	-0.275 (0.162)*
No. Offices per District				0.033 (0.007)***
Incentivisation * No. Offices per District				0.004 (0.013)
Constant	7.988 (0.653)***	7.682 (0.685)***	7.379 (0.737)***	8.819 (0.692)***
Observations	90	90	90	90
R-squared	0.56	0.57	0.57	0.70

Standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 14: Joint (SUR) analysis of quantity and quality

	(1) Log mean EMQ	(2) Log annual Job Entry Points
% PF Offices per District	-0.014 (0.006)**	-0.050 (0.063)
% JCP Offices per District	-0.002 (0.003)	0.032 (0.025)
Private Public Wage Gap	0.001 (0.001)	-0.023 (0.012)*
Log Mean District Frontline Staff	-0.021 (0.008)***	0.498 (0.076)***
Log Mean District Labour Market	-0.059 (0.011)***	-0.201 (0.112)*
Incentivisation Status	0.038 (0.118)	2.737 (1.159)**
Incentivisation * District Frontline Staff	0.021 (0.016)	-0.275 (0.153)*
No. Offices per District	0.001 (0.001)	0.033 (0.007)***
Incentivisation * No. Offices per District	-0.002 (0.001)*	0.004 (0.012)
Constant	0.068 (0.066)	8.819 (0.652)***
Observations	90	90

Standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 15: Mean incentivisation effect

	Frontline Staff	≤25	26-50	≥51	Mean
Small districts	Mean staff all	13.2	38.6	100.2	49.3
	Mean estimated Incentivisation effect	27.6	14.8	-13.6	10.2
	No. offices	22	26	21	69
Large districts	Mean staff all	13.7	36.4	83.5	34.1
	Mean estimated Incentivisation effect	4.6	-9.3	-24.0	-5.1
	No. offices	78	45	30	153
	Mean estimated Incentivisation effect	9.7	-0.45	-19.7	-0.33

References

- Besley, T. and Ghatak, M. (2003) “Incentives, Choice and Accountability in the Provision of Public Services” *Oxford Review of Economic Policy* vol. 19 no. 2 pp. 235 – 249
- Burgess, S., Propper, C., Ratto, M.L. and Tominey, E. (2003) “Incentives in the public sector: some evidence from a UK government agency”, CMPO Working Paper Series No: 03/080.
- Burgess, S., and M. L. Ratto (2003) “The role of incentives in the public sector: issues and evidence”, *Oxford Review of Economic Policy* vol. 19 no. 2.
- Chiappori, P.A. and Salanié, B. (2003) “Testing Contract Theory: a Survey of Some Recent Work”, in *Advances in Economics and Econometrics – Theory and Applications*, Eighth World Congress, M. Dewatripont, L. Hansen and P. Turnovsky, ed., Econometric Society Monographs, Cambridge University Press, Cambridge, 115-149.
- Courty, P., and Marschke, G. (1997) “Measuring Government Performance: Lessons from a Federal Job-Training Program”, *American Economic Review*, 87, 1997, pp.383-388.
- Courty, P., and Marschke, J. (2004) “An empirical investigation of gaming responses to explicit performance incentives”, *Journal of Labor Economics*. Vol. 22, no. 2.
- Dewatripont, M., Jewitt, I. and Tirole, J. (1999) “The Economics of Career Concerns, Part II: Application to Missions and Accountability of Government Agencies”. *The Review of Economic Studies*, Vol. 66, No. 1, Special Issue: Contracts. pp. 199-217.
- Dixit, A. (2002) “Incentives and organisations in the public sector: an interpretative review” *Journal of Human Resources*, 37(4), pp.696-727.
- Encinosa III, W., Gaynor, M., and Rebitzer, J. (1997) “The sociology of groups and the economics of incentives”, NBER Working Paper 5953
- Falk, A. and Ichino, A. (2003) “Clean Evidence on Peer Pressure” IZA Discussion paper no. 732.
- Gaynor, M., M. Pauly (1990) “Compensation and productive efficiency in partnerships: evidence from medical group practice” *Journal of Political Economy*, vol 98:33, 1990, 544-573.
- Gaynor, M., Rebitzer, J. B., and Taylor, L. J. (2001) “Incentives in HMOs”, Economics Working Paper Archive 340, Levy Economics Institute
- Hamilton, B. H., Nickerson, J.A., and Owan, H. (2003) “Team incentives and worker heterogeneity: an empirical analysis of the impact of teams on productivity and participation”, *Journal of Political Economy*, 111(3), pp. 465-497.
- Heckman, J. Ichimura, H. and Todd, P. (1997) Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme, *Review of Economic Studies*, 64, 605-654

- Heckman, J. Smith, J., and Taber, C. (1996) "What do bureaucrats do? The effects of performance standards and bureaucratic preferences on acceptance into the JTPA program", NBER Working Paper 5535
- Holmström, B. (1982) "Moral hazard in teams", *Bell Journal of Economics*, 13, pp. 324-340.
- Holmström, B. and Milgrom, P. (1990) "Regulating trade among agents", *Journal of Institutional and Theoretical Economics*, vol. 146(1), pp. 85-105.
- Holmström, B., and Milgrom, P. (1991) "Multi-task principal-agent analyses: Linear contracts, asset ownership and job design", *Journal of Law, Economics and Organisation*, vol 7, 24-52.
- Itoh, H. (1991) "Incentives to help in multi-agent situations", *Econometrica*, vol 59, (3), 611-636.
- Kahn, C. M., Silva, E. C. D., and Ziliak, J. P (2001) "Performance-based wages in tax collection: The Brazilian Tax Collection Reform and its effects", *Economic Journal*, 111(468), pp.188-205.
- Kandel, E. and Lazear, E. (1992) "Peer Pressure and Partnerships", *Journal of Political Economy*, 100 (4), pp. 801-817.
- Knez, M., and Simester, D. (2001) "Firm-wide incentives and mutual monitoring at Continental Airlines", *Journal of Labor Economics*, 19(4), pp.743-772.
- Lazear, E. (1999) "Personnel economics: Past lessons and future directions", *Journal of Labor Economics*, 17, pp.199-236.
- Lazear, E. (2001) "Performance pay and productivity", *American Economic Review* 90:5 pp. 1346-1361.
- MacDonald, G. and Max, L.M. (2001) "Adverse Specialization", *Journal of Political Economy*, vol 109 (4), 864-899.
- Makinson, J. (2000) "Incentives for change. Rewarding performance in national government networks", Public Service Productivity Panel. HMSO.
- Malcomson, J. (1999) "Incentive contracts in labor markets" in *Handbook of Labor Economics*, Vol.3, O.Ashenfelter and D.Card (eds), North Holland, Amsterdam
- Murphy, K, (1999) "Executive compensation", in *Handbook of Labor Economics*, Vol. 3, O.Ashenfelter and D. Card, eds North-Holland, Amsterdam
- Nickell, S. and Quintini, G. (2002) "The consequences of the decline in public sector pay in Britain: a little bit of evidence", *Economic Journal*, vol 112(477), 107-118.
- Olson, M (1971) *The logic of collective action*. Harvard University Press, Cambridge
- Osborne, D. and Gaebler, T. (1993) "*Reinventing Government: How the Entrepreneurial Spirit Is Transforming the Public Sector*" Plume Books, (Penguin Group) New York.
- Paarsch, H. and Shearer, B. (2000) "Piece rates, fixed wages, and incentive effects: statistical evidence from payroll records", *International Economic Review*, vol 41(1), 59-92.

- Petrongolo, B. and Pissarides, C. (2001) “Looking into the black box: a survey of the matching function”, *Journal of Economics Literature*, vol 39, 390-431.
- Prendergast, C. (1999) “The provision of incentives in firms”, *Journal of Economic Literature*, vol 37, 7-63.
- Prendergast, C. (2002) “The tenuous trade-off between risk and incentives”, *Journal of Political Economy*, 110(5), pp. 1071-1102.
- White Paper “Modernising Government” (1999). www.archive.official-documents.co.uk
- World Development Report (2003) <http://econ.worldbank.org/wdr/wdr2003/>

Appendices

Appendix 1: Data Descriptives

Variable	Mean	Standard Deviation		
		Total	Between	Within
Office Level Variables				
Log Office Monthly Job Entry Points	5.8989	0.9709	0.8740	0.4323
Office Pathfinder Status	0.0540	0.2261	0.2242	0.0000
Office JCP Status	0.0621	0.2413	0.2558	0.0298
District Office	0.0555	0.2290	0.2441	0.0000
Private Public Wage Gap	-0.5504	2.4171	2.3819	0.0000
Log Office Frontline Staff	3.0847	0.8562	0.8571	0.1965
Office Frontline Staff Variance	4.5027	6.3942	6.1628	0.0000
Log Office Labour Market	0.0906	0.4407	0.3114	0.3120
Incentivisation Status	0.2409	0.4277	0.4256	0.0000
Office Mean % High Grade Staff	0.0338	0.0340	0.0340	0.0000
Labour Market Time Series Variation	0.2624	0.0567	0.0567	0.0000
District Level Variables				
Log District Annual Job Entry Points	13.8177	0.3592	0.3610	0.0000
Log District EMQ	-0.1267	0.0514	0.0248	0.0451
Log District JSQ	-0.1703	0.0348	0.0278	0.0218
% PF Offices per District	49.4324	103.9234	86.6939	17.8404
% JCP Offices per District	0.0052	0.0217	0.0159	0.0166
Log District Mean Frontline Staff	8.3573	0.4617	0.4511	0.0000
Log District Labour Market	0.1304	0.3529	0.2168	0.2850
No. Offices per District	11.6473	3.9807	4.2650	0.0000
M * No. Offices per District	3.4562	6.6076	6.6025	0.0000

Appendix 2: Job Entry Priority Group Categories

Priority Client Group 1	Job entry points score 12
Jobless Lone Parents including people on the New Deal for Lone Parents Those on the New Deal for Disabled People People with Disabilities in receipt of a specified primary benefit Other people in receipt of a specified primary benefit	
Priority Client Group 2	Job entry points score 8
People on the New Deal 50 plus People on the New Deal 25 plus Those on the New Deal for Young People Employment Zones Other People with Disabilities not included in Priority Client Group 1 Jobseeker's Allowance (JSA) long term claimants	
Priority Client Group 3	Job entry points score 4
JSA short term claimants	
Priority Client Group 4	Job entry points score 2
Unemployed non claimants	
Priority Client Group 5	Job entry points score 1
Employed People	