



UNIVERSITY OF SOUTHERN DENMARK

COHERE - Centre of Health Economics Research, Department of Business and Economics
Discussion Papers, No. 2013:5
ISSN: 2246-3097



Who to pay for performance? The choice of organisational level for hospital performance incentives

By

Søren Rud Kristensen^{a,b}, Mickael Bech^b, Jørgen Lauridsen^b

^aCentre for Health Economics, University of Manchester, UK

^bCOHERE, University of Southern Denmark

COHERE, Department of Business and Economics
Faculty of Business and Social Sciences
University of Southern Denmark
Campusvej 55
DK-5230 Odense M
Denmark
www.cohere.dk

Who to pay for performance?

The choice of organisational level for hospital performance incentives

Søren Rud Kristensen^{a,b,*}, Mickael Bech^b, Jørgen Lauridsen^b

^a*Manchester Centre for Health Economics, University of Manchester, United Kingdom*

^b*COHERE, University of Southern Denmark, Denmark.*

Abstract

When implementing a pay for performance (P4P) scheme, designers must decide to whom the financial incentive for performance should be directed. This paper compares department level hospital reported performance on the Danish Case Management Scheme at hospitals that did and did not redistribute performance payments to the department level. Across a range of models we find that hospital reported performance at departments that operate under a direct financial incentive is about 5 percentage points higher than performance at departments at hospital where performance payments are not directly redistributed to the department level. This result is in line with the theoretical expectations but due to the non-experimental design of the study, our results only have a causal interpretation under certain assumptions discussed in the paper.

Keywords:

Pay for performance, P4P, Hospital incentives, Incentive design, Team production

JEL classification: I1, L2, M5

*Correspondence to: Søren Rud Kristensen, Manchester Centre for Health Economics, University of Manchester, Oxford Road, Manchester M13 9PL, United Kingdom, soren.kristensen@manchester.ac.uk

1. Introduction

Pay for performance (P4P) schemes that link a proportion of health care providers' income to their performance on quality indicators are becoming the preferred instrument for third-party payers wishing to incentivise higher quality of health care. Although P4P has a high face validity, the evidence of the effectiveness of such schemes in improving quality remains mixed (Rosenthal and Frank, 2006; Mehrotra et al., 2009; Eijkenaar et al.). Designers of P4P schemes face a multitude of design choices (Ryan, 2009; Van Herck et al., 2010; Maynard, 2012; Eijkenaar, 2013) and the key importance of these design choices in determining the effectiveness of P4P schemes has recently been emphasised (Epstein, 2012; Roland, 2012). Knowledge of the effect of these design choices on performance is still limited and more empirical evidence on the impact of specific design choices has been called for (Van Herck et al., 2010; Maynard, 2012; Epstein, 2012; Eijkenaar, 2013).

This paper is concerned with one of the design issues highlighted by the literature, namely to which organisational level financial performance incentives should be directed. Although individual-level incentives has a high level of accountability, due to the joint production nature of health care group-level incentives are typically preferred in health care where the performance of individuals may be impossible or prohibitively costly to identify or can lead to unintended effects (Duckett, 2008; Van Herck et al., 2011; Eijkenaar et al.). In a hospital setting, hospitals have the choice of keeping performance payments at the hospital level or redistribute performance payments within the hospital (Ryan, 2009).

The aim of this paper is to assess whether redistribution of performance payments to the department level is associated with higher performance than keeping performance payments at the hospital level with no explicit redistribution to departments. Our performance indicator is hospital reported performance on the The Danish Case Management Scheme (DCMS) measured by registrations in the patients' medical records. Our work is thus related to Sutton et al.'s (2010) work on 'record rewards'. Sutton et al. investigated the effect of the financial incentives for recording patient's risk factors such as blood pressure and smoking status embedded in the UK P4P scheme for primary care the Quality and Outcomes Framework. Sutton et al. found that incentivising the recording of risk factors significantly increased recording efforts with positive spill overs to unincentivised areas. Focusing on general practices, Sutton et al. did not consider the effect of paying for

performance at different organisational levels.

Our work is also related to the literature that has analysed the difference between rewarding individual physicians and teams. Town et al. (2004) and Conrad and Christianson (2004) considered the theoretical aspects of the issue. Newhouse (1973) tested the difference in overhead costs and hours worked for physicians who did and did not work under a revenue sharing scheme. He found significantly higher costs but a small and statistically insignificant decrease in hours worked for physicians in revenue sharing practices compared to physicians who did not share revenue. Prendergast (1999) notes that as Newhouse’s study is cross-sectional, selection effects might explain the revealed pattern. Gaynor and Gertler (1995) found that revenue sharing in medical groups significantly reduced physician effort. Again, the conclusions are drawn on the basis of cross-sectional data, and there may be other explanations for Gaynor and Gertler’s result. Reviewing a large body of literature, Van Herck et al. (2010) conclude that incentive schemes directed toward individuals or teams generally have a larger effect than programmes directed at hospitals. However, this conclusion is derived on the basis of studies assessing different programmes in different contexts, and other factors may explain the apparent difference in performance.

Our estimand of interest is the difference in performance at hospital departments where performance payments are redistributed directly to the department level versus performance at hospitals that do not redistribute payments to lower organisational level. Using a variety of difference-in-differences estimators to account for the nature of our data and the non-experimental introduction of the P4P scheme we find that hospital reported performance was on average 5 percentage point higher at hospital departments facing a direct financial performance incentive.

We proceed by briefly describing the setting of our study. We then generate a working hypothesis on the basis of an analytical framework and discuss our study design, data and methods used. We then present our results and end with some concluding remarks and a discussion of the generalisability and limitations of our findings, especially a discussion of when a causal interpretation of our findings is valid.

2. The Danish Case-Management Scheme

Since 2001, through an agreement between the Danish government and the Danish Association of Counties (now regions), all inpatients and outpa-

tients in long-term treatment have had the right to a case manager. The assigned case manager must be a member of the team of health care professionals treating the patient and may either be a physician or, more likely, a nurse. The stated objective of providing all patients with a case manager is to improve the quality of treatment by improving continuity of care and increasing patients' feelings of safety (Ministry of Finance, 2004). It is the case manager's task to maintain an overview of the patient's treatment, secure coordination and continuity and act as a liaison between the hospital and the patients and their relatives.

Patients self-reporting a case manager also report being more satisfied with the hospital staff's level of knowledge on the patient's course of illness. These patients experience a more coherent patient course, less unnecessary waiting time during their admission, and greater satisfaction with their level of involvement in decision-making than patients who do not report having a case manager (The Unit of Patient-Perceived Quality, 2009). However, as suggested by Hasnain-Wynia and Jean-Jacques (2009), more research is needed into the relationship between patient-centered care and clinical outcomes.

In 2009, assigning case managers to patients became a legal requirement in Denmark. In addition, the provision of case managers to patients has become part of the Danish national health care quality accreditation programme (The Danish Institute for Quality and Accreditation in Healthcare, 2011).

In 2004, quarterly monitoring of hospital's performance on the Danish case management scheme (DCMS) was initiated. Adherence to the performance scheme is measured by what is known as the medical record indicator. This is a process indicator measuring the extent to which, according to a note in patients' medical record, hospitals have assigned case managers to patients. In the following we shall refer to this measure as hospital reported performance (HRP)

The P4P scheme analysed in this paper was introduced in the region of Southern Denmark in 2009. This scheme distributes 8m DKK \approx 1.1m EUR to the 4 hospitals in the region according to the hospitals' absolute performance on the DCMS. Payments are based on HRP and are disbursed to the hospitals on the basis of the mean score of the hospital's first three quarterly performance assessments. The maximum attainable amount differs for each hospital according to the production value as measured by the diagnosis related group (DRG) system. Payments are distributed according to the degree to which hospital performance is above or below certain thresholds

Table 1: Bonus Allocation Rules in the regional P4P scheme

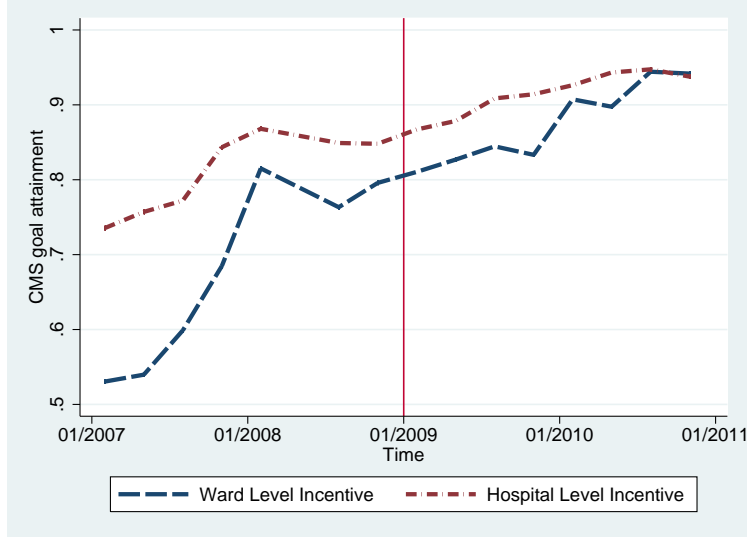
HRP	H 1	H 2	H 3	H 4
Above 97.5	3.607	1.990	1.196	1.206
95.0–97.49	3.006	1.659	0.997	1.005
92.5–94.99	2.605	1.437	0.864	0.804
90.0–92.49	1.804	0.995	0.598	0.603
87.5–89.99	1.202	0.663	0.399	0.536
85.0–87.49	0.601	0.332	0.199	0.469
82.5–84.99	0.000	0.000	0.000	0.402
80.0–82.49	−1.804	−0.995	−0.598	0.335
77.5–79.99	−3.607	−1.990	−1.196	0.268
75.0–77.49	−3.607	−1.990	−1.196	0.201
72.5–74.99	−3.607	−1.990	−1.196	0.134
70.0–72.49	−3.607	−1.990	−1.196	0.067
67.5–69.99	−3.607	−1.990	−1.196	0.000
65.0–67.49	−3.607	−1.990	−1.196	−0.603
Below 65	−3.607	−1.990	−1.196	−1.206

Note: Bonus/penalty in DKK (100DKK \approx)13.4 EUR) per hospital(H) for different performance levels. Payments were redistributed to the department level at hospital 2 and 4

(see Table 1). These performance goals was set by the region.

Three of the hospitals in the region receive payments if the HRP score is above 82.5-84.9 percent. For performance below this level the hospitals must pay a fine of up to the same amount as the hospital maximum reward. Each step is equal to a 2.5 percentage point performance increase. The maximum payment is received for performance of 97.5 percent or above. For the largest hospital, this amounts to 3.6m DKK \approx 0.5m EUR. One hospital (H4) which scored significantly lower on the MRI in 2008, has a special allocation rule with a lower performance payment threshold at 67.5-69.9 percent. The potential performance payments make up a only a tiny fraction of the hospital budget (the average hospital budget size in the region was 3bn DKK in 2010) and is given to the hospitals without restrictions on use. Hospitals are for example free to redistribute performance payments to the department level. Two of the four hospitals in the region chose to do so. The decision was hospital wide and thus exogenous to the department

Figure 1: Measured performance by hospital and redistribution scheme
 Quarterly mean department level performance for hospital departments to which performance payments are and are not redistributed. The red vertical line indicates the introduction of the P4P scheme. The lines are connected through a data breakage (due to a hospital worker strike) in the second quarter of 2008



level which is our unit of analysis. However, as we are unable to observe the basis on which hospitals chose to redistribute performance payments to the department level, our results will potentially suffer from endogeneity bias. We return to this issue later. The performance payments could be used at the hospital departments discretion but was not used as salary/bonus payments to individual members of staff.

Figure 1 displays HRP on the case management scheme for hospital departments in the region under study from 2007-2010. The blue dashed line displays the mean performance at the two hospitals where hospital departments were given a direct financial incentive when the P4P scheme was established. The red dot-dashed line represents departments at hospitals where the performance payments were maintained at hospital level. HRP follow similar trends in both types of hospitals before the introduction of the scheme and goal attainment has increased over time. After the introduction of the scheme (indicated by the vertical line) the increase in performance appears to be stronger in the group that received payments at the department level. The group that did not receive direct incentive payments was closer to max-

imum performance at the outset. Thus, additional performance increases may have been more difficult.

The goal of the rest of this paper is to establish whether differences in the organisational level of payment in the P4P scheme is associated with differences in hospital reported performance on the DCMS

3. Analytical framework

A framework for our analysis may be found in the agency literature on team production. Alchian and Demsetz (1972) pioneered this literature by describing the problem of joint production as one in which the outcome is non-separable; that is, the individual's marginal contribution to the outcome cannot be identified or is prohibitively costly to identify. Holmström (1982) described the problem as one of free-riding, when risk-averse agents to whom effort is costly exert insufficient effort when individual performance is not monitored, and output is shared equally among the agents. For that reason, the problem has also been described as the $\frac{1}{N}$ problem (Prendergast, 1999).

In the DCMS, performance is measured at an aggregated level—either the hospital or the department level. The total output can be seen as the performance payments to the level at which performance is measured.

In the illuminating presentation of the $\frac{1}{N}$ -problem by Kandel and Lazear (1992) the total output is seen as a function of the individual effort of N identical hospital employees, $e_i \geq 0$, denoted $f(\mathbf{e})$ which is assumed non-separable in e_i and where \mathbf{e} is an N -dimensional vector of each employee's effort. The output is assumed distributed equally between the employees so that each employee wishes to maximise

$$\max_{e_i} \frac{f(\mathbf{e})}{N} - C(e_i) \quad (1)$$

where $C(e_i)$ is the cost of effort assumed to be increasing in e_i at an increasing rate with first and second order derivatives: $C_{e_i} > 0$ and $C_{e_i^2} > 0$.

Compare this to the hospital's or department's maximisation problem focusing on total output:

$$\max_{e_1, e_2, \dots, e_N} f(\mathbf{e}) - \sum_{i=1}^N C(e_i) \quad (2)$$

The classical free-riding problem arises because performance is measured at the organisational level and the total performance payment is not distributed according to individual effort, but shared equally among the employees. Each worker bears the full costs of his own effort, but receives only a fraction of the performance payments.

Barua et al. (1995) show that the effort level that solves the organisation's maximisation problem e_i^* must be greater than the effort level that solves the individual's maximisation problem e_i' under the assumption that $f_{e_i}(\mathbf{e}) > 0$ and $f_{e_i^2}(\mathbf{e}) > 0$ when $C_{e_i^2} > 0$.

Adams (2002) point out, that whether worker i 's effort level decreases in N depends on whether f_{e_i} is constant in N . For example, effort levels would decrease in N if the production function is additive. If, on the other hand, the employee's effort levels are complementary, then the opposite might be true. In the case of recording case managers in patients' medical records, we find additivity a reasonable property to assume about the production function, as worker i 's recording of case managers is an activity that should be independent on the effort levels of other workers. For that reason we expect the risk of free-riding to increase with team size.

Kandel and Lazear suggest peer pressure as a remedy for the free-riding problem. Knowing that their expected output depends on the effort of their colleagues, employees may monitor their co-workers to secure that they put in a sufficient level of effort.

Kandel and Lazear define the expected penalty of being caught shirking by a coworker a function of the effort agent i puts forward, and the monitoring level of his co-workers. Thus, peer pressure adds a costly monitoring effort into agent i 's cost function.

Kandel and Lazear conclude that whether effort decreases or increases in N depends on the actual shapes of the production function and the peer pressure function. In their subsequent discussion, Kandel and Lazear also suggest that the possibility of being caught shirking in a company is likely to decrease in N . However, as pointed out by Backes-Gellner et al. (2004), this does not follow directly from the Kandel-Lazear model. Backes-Gellner et al.'s extension analyses the relationship between peer pressure and team size with the goal of investigating the joint effect of free-riding and peer pressure given a specific team-size. Under the assumptions that the probability of being caught shirking increases in N because monitoring increases effort but at a decreasing rate and the individual cost function is independent of N , an effort function that is concave in N , and the existence of an optimal N that

maximises effort level, they suggest that the peer pressure effect dominates the free riding effect for small team sizes. In their model, for team sizes above the optimal size, N^* , the free riding effect will dominate the peer pressure effect, and effort will decrease in N for $N > N^*$.

3.1. Hypothesis

The team production literature proposes that when agents receive only a fraction of a jointly produced output and individual performance is thus not rewarded, there is a risk of free riding. Under assumptions that adequately describe the DCMS, the problem is likely to increase with team size. Although peer monitoring has been suggested as a remedy, the effectiveness of this tool might decrease as team size increases, because the free-riding effect will dominate the effect of peer pressure for teams above a certain size.

Against this background we hypothesise that redistribution financial incentive for performance to the department level is associated with higher performance than keeping performance payments at the hospital as an organisational unit. Even if the achievable reward faced by each agent is the same in the big N and the small N case, we expect higher performance in the small N case, as the ability for each agent to monitor the effort levels of the other agents is larger in the case of smaller teams.

4. Data

Our dataset is an unbalanced panel of 94 hospital departments from the four hospitals in the region of Southern Denmark. We have quarterly observations of HRP, our dependent variable, at the department level from 2007-2010. The mean numbers of departments that submitted quarterly performance data at each of the four hospitals were 32, 21, 17 and 12.

At the department level, the basis for each observation of HRP is a minimum of 15 randomly drawn medical records from the patient administrative systems in each hospital department. The variable is defined as

$$HRP_{it} = \frac{y_{it} = 1}{n_{it}}, \quad (3)$$

where $y_{it} = 1$ denotes medical records where the assignment of a case manager has been recorded, and n_{it} is the number of random draws from department i at time t .

Table 2: Mean performance on the case management scheme by hospital and year
Mean performance in pct. at the 4 hospitals. P4P was introduced from 2009. Hospital 2 and 4 redistributed performance payments to the department level

Hospital	Year				Total
	2007	2008	2009	2010	
Hospital 1	78	85	87	92	86
Hospital 2	65	83	84	93	81
Hospital 3	78	87	92	97	89
Hospital 4	50	73	82	92	75
Total	71	83	87	93	84

Table 2 below displays the mean performance aggregated at the hospital level over time and the number of observations per hospital per year. It can be seen that the number of observations is lower for all hospitals in 2008. This is due to a nation wide strike among hospital employees, which means that there are no observations for the second quarter of 2008.

5. Empirical strategy

To assess the difference in performance associated with redistributing performance payments directly to the department level we begin with a difference in differences model.

Our dependent variable is the hospital reported performance on the DCMS HRP_{it} (defined in the previous section) in department i in quarter t . A dummy variable D_{jt}^1 takes a value of 1 for hospital departments at hospitals that redistribute performance payments to the department level from 2009, and another dummy variable D_{jt}^2 indicates whether an observation is from this period. We further include hospital fixed effects (u_h) and year fixed effects (v_y):

$$HRP_{it} = \alpha + \tau D_{jt}^1 D_{jt}^2 + u_h + v_y + \epsilon_{jt}. \quad (4)$$

Our primary interest is in the estimate of τ which is the difference in mean HRP at departments with and without a direct financial incentive for

performance after the P4P scheme was introduced.

5.1. Estimation issues

5.1.1. Fractional dependent variable

Because $0 < HRP_{it} < 1$, the effects of any explanatory variable cannot be constant across all values. In addition, OLS on an untransformed dependent variable may lead to predictions of MRI outside of the $[0,1]$ interval. For this reason, we assume a logistic distribution so that

$$HRP_{it} = \Lambda_{it}(\mathbf{x}'_{it}\beta) + \epsilon_{it} = \pi_{it} + \epsilon_{it}. \quad (5)$$

This can be estimated with OLS after performing a logit transformation, taking the log of odds:

$$\ln\left(\frac{HRP_{it}}{1 - HRP_{it}}\right) = \mathbf{x}'_{it}\beta. \quad (6)$$

A non-trivial proportion of the observations display a goal attainment of 0 or 100 percent. In that case, (6) does not hold because the natural logarithms for 0 and 1 are undefined. A possible solution, advocated by Greene (2000), is to modify the dependent variable by subtracting/adding 0.001 to the HRP to make logit transformation possible and avoid losing observations. Alternatively, Papke and Wooldridge (1993, 1996, 2008) suggest estimating a quasi maximum likelihood model using the generalised linear model (GLM) framework instead.

We apply the generalised linear model (Nelder and Wedderburn, 1972; McCullagh and Nelder, 1989; Hardin and Hilbe, 2007) and follow Papke and Wooldridge's suggestion for estimating proportions by assuming a binomial distribution and a logit link function. This approach ensures that the predictions are within the $[0,1]$ range and has the additional benefit of allowing for the inclusion of observations in which MRI takes the value 0 percent or 100 percent without any manipulation of data.

5.1.2. Heteroscedastic error term

The variance of the error term is heteroscedastic because it depends on the denominator of the proportions, n :

$$\text{Var}[\epsilon_{it}] = \frac{1}{n_{it}HRP_{it}(1 - HRP_{it})}. \quad (7)$$

To address this issue we apply Berkson’s minimum chi-square estimator through weighted least squares. In the first step, we estimate the $H\hat{R}P_{it}$ as

$$M\hat{R}I_{it} = \frac{\exp(\mathbf{x}'_{it}\hat{\beta})}{1 + \exp(\mathbf{x}'_{it}\hat{\beta})} \quad (8)$$

In the second step we perform the regression again, but this time, we use analytical weights, defined as

$$n_{it}H\hat{R}P_{it}(1 - H\hat{R}P_{it}). \quad (9)$$

5.1.3. *Unobserved heterogeneity*

The panel nature of our data allow us to control for unobserved department-specific heterogeneity. In a fixed effects (FE) model, the department level effects are allowed to be correlated with the explanatory variables. The model is less efficient than the random effects (RE) model which is as consistent as the FE model if the stricter assumption of no correlation between the department level effects and the explanatory variables holds. As we have no reason to suspect correlation between the department level effect and the explanatory variables we use a Hausman test to guide our choice between a fixed and random effects and use the more efficient RE model as this was not rejected by the Hausman test.

In the GLM case, we rely on an extension of the GLM for panel data using a generalised estimating equations (GEE) estimator which requires the same assumption of no correlation between the unobserved department effect and the explanatory variables as the random effects model (Liang and Zeger, 1986; Papke and Wooldridge, 2008).

5.1.4. *Potential endogeneity*

The decision to redistribute performance payments to the department level was taken by the hospital. Although exogenous at the department level (hospitals used the same incentive scheme for all departments, and the departments were not asked about which payment scheme the preferred) there is a potential risk of biased estimates due to endogeneity. We do not observe the basis on which the hospitals made their decision, but in an attempt to address the issue, we include past department-level performance in some regressions, but only up to the time that the P4P scheme was introduced.

5.1.5. Interaction terms in non-linear models

Our central estimate of interest τ is the interaction of two dummy variables (time and group). As highlighted by Ai and Norton (2003), the sign and statistical significance of an interaction term in a non-linear model can be different for different values of the explanatory variables. We thus supplement our estimates by calculating the cross-partial derivative of the expected value of HRP. Following Norton et al. (2004), in the case of two dummy variables, the cross-partial derivative of the expected value of HRP is the discrete double difference defined as:

$$\frac{\Delta^2 F(u)}{\Delta D^1 \Delta D^2} = \frac{1}{1 + e^{-(\tau + u_h + v_y + \mathbf{x}'\beta)}} - \frac{1}{1 + e^{-(u_h + \mathbf{x}'\beta)}} - \frac{1}{1 + e^{-(v_y + \mathbf{x}'\beta)}} + \frac{1}{1 + e^{-(\mathbf{x}'\beta)}}, \quad (10)$$

where $F(u)$ is the conditional mean of the HRP:

$$F(u) = \frac{1}{1 + e^{-(\tau + u_h + v_y + \mathbf{x}'\beta)}}. \quad (11)$$

We calculate the statistical significance of this value using the Delta method¹, and plot the size and significance of the department level effect against predicted performance.

6. Results

6.1. Performance increase in departments subject to a direct financial incentive

The estimation results are shown in Table 3. In all of the models, the sign of the coefficient measuring the difference in performance between hospital departments with and without a direct incentive for performance is positive and statistically significant. We refer to this difference as the incentive effect, but discuss when this effect should be given a causal interpretation in the next section.

In all but the first model which uses the untransformed dependent variable, the reported coefficients are the log relative odds. These are non-trivial to interpret because the coefficients do not simply represent the change in the MRI for a discrete change in the variable but depend on the values of

¹Both implemented using Stata's nlpredict

the explanatory variables. For that reason we also present the average partial effects (Greene, 2011; Wooldridge, 2002) which are comparable across models.

Model 1 uses OLS on the untransformed MRI to estimate the effect of redistributing performance payments to the department level. According to this model, the increase in performance resulting from directing incentives to the department level compared to the hospital level is approximately 7 percentage point.

Because performance on the MRI is bound between 0 percent and 100 percent, we move to a logistic scale from model 2 and onward. Model 2 is also estimated using OLS but it employs the logit transformed dependent variable. This yields a slightly small estimate of the difference in performance between departments with and without a direct performance incentive. When we control for unobserved heterogeneity at the department level using an RE model in model 3 there is a further decrease in the estimated difference in performance.

Model 4 represents our attempt to control for potential endogeneity in the decision to redistribute performance payments to the department level. We use the same specification as model 2 but add the lagged department-level performance that takes the values of MRI_{it-1} before the introduction of the P4P scheme and 0 afterwards. This inclusion explains the lower number of observations. The inclusion of lagged performance in Model 4 seems to lead to lead to biased estimates.

Model 5 is the Berkson minimum chi-square estimator, using a weighted least squares model to address the heteroscedastic error term arising from the fact that the dependent variable is a ratio for which the sample size n_{it} varies between observations. As expected this leads to an increased precision of the estimates and an estimate of τ within the range of the previously estimated models.

Finally model 6 and 7 represents our attempt to deal simultaneously with the limited dependent variable, potential heteroscedasticity, endogeneity and in model 7 also heterogeneity. In model 6 we use a GLM model with a binomial distribution of the dependent variable and a logit link and model 7 is the estimates from the GLM equivalent of a RE model, the population-averaged panel data model. Again we find an estimated increase in performance from redistributing performance payments directly to the department level of about 5 percentage points.

Table 3: Results

	(1) OLS untrnsf.	(2) OLS	(3) RE	(4) OLS	(5) Berkson's MCS	(6) GLM	(7) Pop. avg. GEE
Dept. Level Incentive (DLI)	0.069** (2.65)	0.836** (2.70)	0.702** (2.45)	2.430*** (7.50)	1.314*** (8.90)	1.139*** (5.75)	0.950*** (5.35)
2009-2010	0.110*** (6.85)	1.324*** (6.04)	1.370*** (6.51)	0.359** (2.57)	0.378*** (3.54)	0.414*** (3.84)	0.553*** (4.67)
Hosp. 1	0.009 (0.33)	0.273 (0.65)	0.0736 (0.17)	-0.0166 (-0.06)	-0.123 (-1.05)	-0.195 (-1.62)	-0.149 (-0.93)
Hosp. 2	-0.0371 (-1.24)	-0.455 (-1.15)	-0.708* (-1.75)	-0.293 (-1.12)	-0.0784 (-0.58)	-0.231 (-1.61)	-0.252* (-1.71)
Hosp. 4	-0.139** (-3.34)	-1.381** (-2.60)	-1.390** (-2.45)	-0.927** (-2.41)	-0.564*** (-5.27)	-0.645*** (-4.78)	-0.655** (-2.87)
Lag performance				0.566*** (14.16)	0.366*** (15.14)	0.353*** (7.93)	0.261*** (6.02)
Constant	0.788*** (37.08)	2.196*** (7.96)	2.426*** (8.40)	1.400*** (6.34)	1.003*** (10.96)	1.029*** (9.71)	1.102*** (8.44)
APE (DLI)	0.069***	0.053**	0.041**	0.15***	0.075***	0.055***	0.047***
R^2	0.191	0.143	0.142	0.334	0.448		
Adj. R^2	0.188	0.139		0.331	0.444		
AIC	-829.4	5680.1		4471.1	.	7848.7	
ρ			0.412				
N	1233	1233	1233	1042	759	1042	1042

Note: t/z statistics in parentheses. Estimation on department-quarter data with department cluster robust standard errors. Dep. variable logit transformed if nothing else stated. APE (DLI) is the average partial effect of the department level incentive. For model 3, a Hausman test did not reject the use of a random effects model, and the overall R^2 is presented.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$

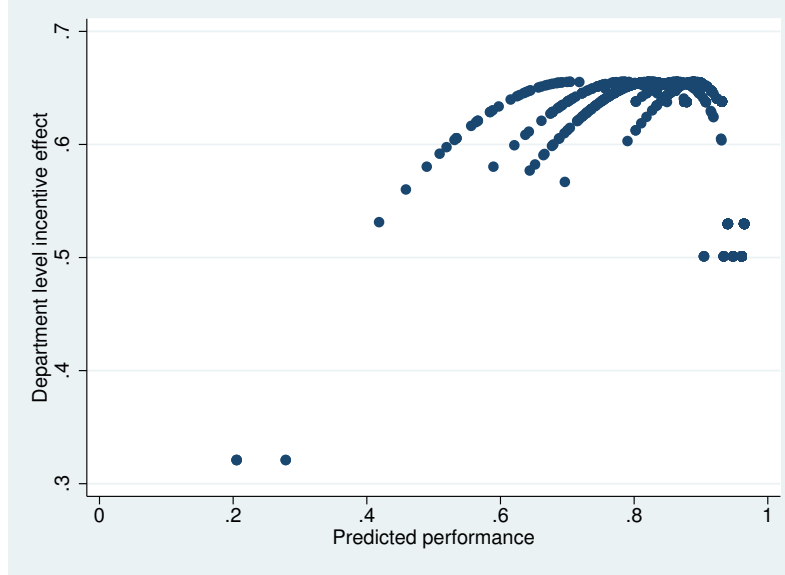
6.2. Heterogeneous effect estimates

We now follow the presentational approach suggested by Ai and Norton² and plot the interaction effect (Figure 2), the interaction effect divided by predicted performance (Figure 3) and the z-value against the predicted performance (Figure 4). All graphs are based on model 7 in Table 3, but we have verified the results in the other models, which produce similar results.

Figure 2 displays the interaction effect as a function of predicted performance. It can be seen that for hospital departments with low predicted performance, redistributing performance payments to the department level has the lowest absolute impact although it is still positive. The impact increases as predicted performance increases and is the largest for departments with predicted performance between roughly 70 and 90 percent. The four "arms" of the prediction arises from our inclusion of hospital fixed effects and a time dummy. For hospital departments with a predicted performance between 90 and 100 percent the absolute effect of targeting the financial

²The additional suggestions for analysing interaction effects in non-linear models in Greene (2010) is more relevant in models with continuous variables and we ignore it in this paper.

Figure 2: Department level incentive effect against predicted probability
 Estimates from a population averaged GEE model. Department level effect calculated as suggested by Norton et al. (2004). See the text for details.



incentive at the department level again decreases.

Figure 3 shows the effect of a direct department-level incentive relative to predicted performance. This figure shows that relative to predicted performance, a direct, department-level incentive has the largest relative effect for departments with low predicted performance and the relative effect decreases with predicted performance.

Figure 4 displays the z-statistics against predicted performance and demonstrates that although statistically significant for all departments, the effect of the department-level incentive is strongest for a hospital departments with average performance. Again, the multiple series arises from the hospital and time dummies included in the model.

In summary, the examination of the effect of a department-level incentive across departments confirms that the effect of targeting incentives directly at the department level is indeed positive and highly statistically significant for all departments in all model specifications with the exception of the OLS model with lagged performance which predicts some statistically insignificant effects of redistributing performance payments to the department level.

Figure 3: Department level incentive effect relative to performance
 Estimates from a population averaged GEE model. Department level effect calculated as suggested by Norton et al. (2004). See the text for details.

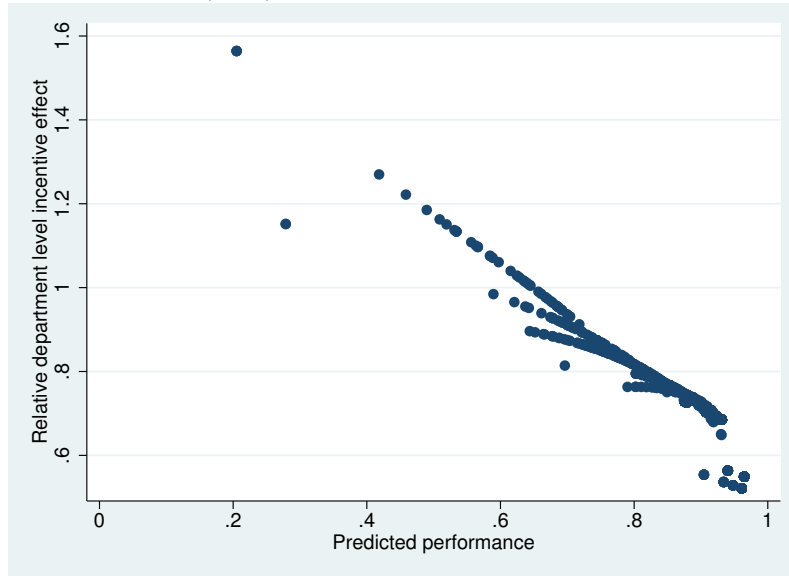
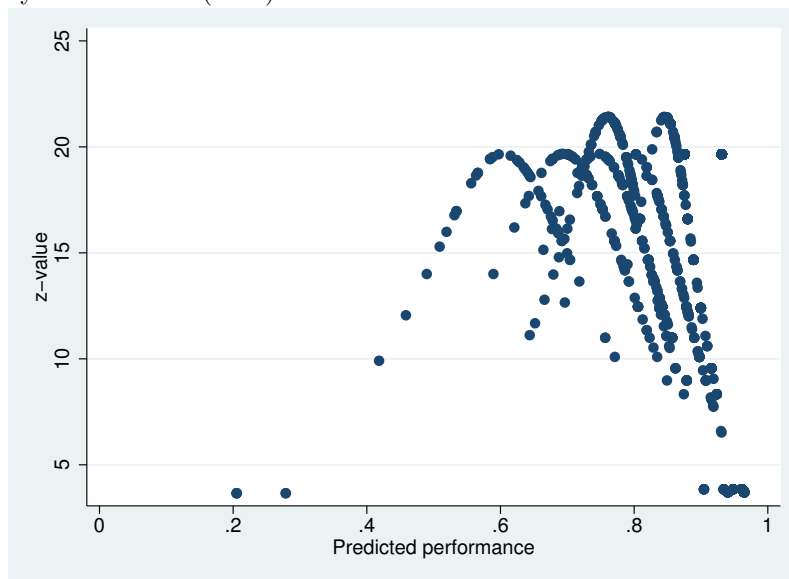


Figure 4: z-statistics of the Department level incentive effect
 Estimates from a population averaged GEE model. Department level effect calculated as suggested by Norton et al. (2004). See the text for details.



6.3. Sensitivity analysis

From Figure 1 it seemed that the difference between the two groups of hospital departments was especially large in the first 6 months. To investigate whether this difference affected our results we ran all of the models estimated without data from the first 6 months of 2007. We also tried excluding all of the observations from 2007. This test did not significantly alter our conclusions. Therefore, we do not report the results here. The results are available from the authors on request.

7. Discussion

In the words of Oliver and Brown (2011, p. 59), health care systems currently face pressure to deliver the highest possible "bang" for the "buck". Crafting incentive schemes for quality enhancement is one way to ensure that this goal is fulfilled.

Internal redistribution of incentives targeted to the hospital level has been suggested as a possible way of getting the most out of P4P (Ryan, 2009). This potential solution may be desirable because much of health care is a team effort and therefore incentivising individuals may be impossible or prohibitively costly, due to the difficulties in distinguishing individual contributions to the outcome, or undesirable due to the potential unintended consequences for team collaboration.

The analysis showed that hospital departments with a direct performance incentives increased HRP by approximately 5 percentage points compared to departments at hospitals that were also subject to a P4P scheme but did not redistribute performance payment to the department level. The estimated effect varies in magnitude when departments are examined individually, but the result of a positive and statistically significant effect was stable across the various model specifications.

Our results are in line with the theoretical expectations set out at the start of the paper. However, we wish to emphasise two limitations to the generalisability of our results.

Firstly, the choice of redistributing performance payments to the department level was not made at random, but by the general hospital management. The decision to redistribute was not taken by department management and applies to all departments. The decision is thus exogenous to the department level which is our unit of analysis. However, if departments at hospitals that

chose to redistribute performance payments differs systematically from departments at hospitals that did not redistribute payments (for example are more competitive), our results should not be given a causal interpretation. However, performance at hospitals that chose to redistribute performance payments to the department level was from the onset below the performance of hospitals that did not chose to redistribute payments. On that background we it unlikely that our results are subject to endogeneity bias although the risk cannot be ruled out. We attempt to adjust for potential endogeneity related to previous performance by including the lagged value of HRP in our regressions and found that this did not alter our conclusion.

Secondly, the performance indicator we studied measures hospitals' self reported performance on the DCMS. The indicator essentially measures a hospital department's effort in recording case managers' names in patients' medical records. This indicator can potentially be manipulated by hospitals without much effort and does not necessarily translate to an increase in these patients' experience of having a case manager. In that sense our indicator is different from indicators that incentivise processes such as brain imaging or outcomes such as reductions in mortality rates. Whether similar performance differences related to payment level would materialise for processes that require more effort is thus uncertain, but the theoretical framework put forward in this paper does indicate that similar effects are likely to occur for process and outcome indicators as well.

References

- Adams, Christopher. Does size really matter? empirical evidence on group incentives. *FTC Bureau of Economics Working Paper*, (52), October 2002.
- Ai, Chunrong and Norton, Edward C. Interaction terms in logit and probit models. *Economics Letters*, 80(1):123–129, July 2003.
- Alchian, Armen A. and Demsetz, Harold. Production, information costs, and economic organization. *The American Economic Review*, 62(5):777–795, 1972.
- Backes-Gellner, Uschi; Werner, Arndt, and Mohnen, Alwine. Team size and effort in start-up-teams - another consequence of free-riding and peer pressure in partnerships. March 2004.

- Barua, Anitesh; Lee, C.-H. Sophie, and Whinston, Andrew B. Incentives and computing systems for team-based organizations. *Organization Science*, 6 (4):487–504, July 1995.
- Conrad, D. A. and Christianson, J. B. Penetrating the” black box”: financial incentives for enhancing the quality of physician services. *Medical Care Research and Review*, 61(3):37–68, 2004.
- Duckett, Stephen. Design of price incentives for adjunct policy goals in formula funding for hospitals and health services. *BMC Health Services Research*, 8(1):72, 2008.
- Eijkenaar, Frank. Key issues in the design of pay for performance programs. *The European Journal of Health Economics*, 14(1):117–131, 2013.
- Eijkenaar, Frank; Emmert, Martin; Scheppach, Manfred, and Schffski, Oliver. Effects of pay for performance in health care: A systematic review of systematic reviews. *Health Policy*.
- Epstein, Arnold M. Will pay for performance improve quality of care? the answer is in the details. *New England Journal of Medicine*, 367(19):1852–1853, 2012.
- Gaynor, Martin and Gertler, Paul. Moral hazard and risk spreading in partnerships. *The RAND Journal of Economics*, 26(4):591–613, 1995.
- Greene, William. Testing hypotheses about interaction terms in nonlinear models. *Economics Letters*, 107(2):291–296, May 2010.
- Greene, William H. *Econometric analysis*, volume 4. Prentice Hall, Upper Saddle River, N.J., 2000.
- Greene, William H. *Econometric analysis*. Prentice Hall, Upper Saddle River, N.J., 7th edition, 2011.
- Hardin, J.W. and Hilbe, J. M. *Generalized linear models and extensions*. Stata Corp, 2007.
- Hasnain-Wynia, Romana and Jean-Jacques, Muriel. Filling the gaps between performance incentive programs and health care quality improvement. *Health Services Research*, 44(3):777–783, 2009.

- Holmstrom, Bengt. Moral hazard in teams. *The Bell Journal of Economics*, 13(2):324–340, 1982.
- Kandel, Eugene and Lazear, Edward P. Peer pressure and partnerships. *The Journal of Political Economy*, 100(4):801–817, 1992.
- Liang, Kung-Yee and Zeger, Scott L. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, April 1986.
- Maynard, Alan. The powers and pitfalls of payment for performance. *Health Economics*, 21(1):3–12, January 2002.
- McCullagh, P. and Nelder, J.A. *Generalized linear models*. Chapman and Hall Ltd, London, 2nd ed. edition, 1989.
- Mehrotra, A.; Damberg, C. L.; Sorbero, M. E. S., and Teleki, S. S. Pay for performance in the hospital setting: What is the state of the evidence? *American Journal of Medical Quality*, 24(1):19, 2009.
- Ministry of Finance, . [finansministeriet] agreements on the economy for local governments 2005 [aftaler om den kommunale økonomi for 2005]. Technical report, Finansministeriet, København, 2004.
- Nelder, J. A. and Wedderburn, R. W. M. Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3):370–384, January 1972.
- Newhouse, Joseph P. The economics of group practice. *The Journal of Human Resources*, 8(1):37–56, January 1973.
- Norton, E. C.; Wang, H., and Ai, C. Computing interaction effects and standard errors in logit and probit models. *Stata Journal*, 4(2):154–167, 2004.
- Oliver, Adam and Brown, Lawrence D. Incentivizing professionals and patients: A consideration in the context of the united kingdom and the united states. *Journal of Health Politics, Policy and Law*, 36(1):59–87, January 2011.
- Papke, L. E and Wooldridge, J. M. Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics*, 145(1):121–133, 2008.

- Papke, Leslie E. and Wooldridge, Jeffrey M. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *National Bureau of Economic Research Technical Working Paper Series*, No. 147, November 1993.
- Papke, Leslie E and Wooldridge, Jeffrey M. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics*, 11(6):619–32, 1996.
- Prendergast, Canice. The provision of incentives in firms. *Journal of Economic Literature*, 37(1):7–63, 1999.
- Roland, Martin. Pay-for-performance: Not a magic bullet. *Annals of Internal Medicine*, 157(12):912–913, December 2012.
- Rosenthal, M. B. and Frank, R. G. What is the empirical basis for paying for quality in health care? *Medical Care Research and Review*, 63(2):135, 2006.
- Ryan, Andrew. Hospital-based pay-for-performance in the united states. *Health Economics*, 18(10):1109–1113, 2009.
- Sutton, Matt; Elder, Ross; Guthrie, Bruce, and Watt, Graham. Record rewards: the effects of targeted quality incentives on the recording of risk factors by primary care providers. *Health Economics*, 19(1):113, 2010.
- The Danish Institute for Quality and Accreditation in Healthcare, . The danish healthcare quality programme - accreditation standards for hospitals(DDKM) 1st ver. 2nd ed. [den danske kvalitetsmodel - akkrediteringsstandarder for sygehuse].
- The Unit of Patient-Perceived Quality, . [enheden for brugerundersgelser] - the national danish survey of patient experiences [den landsdkkende undersgelser af patientoplevelser (LUP)], 2009.
- Town, R.; Wholey, D. R.; Kralewski, J., and Dowd, B. Assessing the influence of incentives on physicians and medical groups. *Medical care research and review*, 61(3):80–118, 2004.
- Van Herck, Pieter; De Smedt, Delphine; Annemans, Lieven; Remmen, Roy; Rosenthal, Meredith B, and Sermeus, Walter. Systematic review: Effects,

design choices, and context of pay-for-performance in health care. *BMC Health Services Research*, 10(1):247, 2010.

Van Herck, Pieter; Annemans, Lieven; De Smedt, Delphine; Remmen, Roy, and Sermeus, Walter. Pay-for-performance step-by-step: Introduction to the MIMIQ model. *Health Policy*, 102(1):8–17, September 2011.

Wooldridge, Jeffrey M. *Econometric analysis of cross section and panel data*. MIT Press, Cambridge, Mass., 2002. ISBN 0262232197 (cloth).