Pay-for-performance schemes: Should optimal prices vary across system and clinical quality indicators?  

SVERRE GREPPERUD ¹  

¹ Department of Health Management and Health Economics, University of Oslo  

Abstract: Quality indicators are classified into system or clinical quality indicators. Typically, different levels of an organization steer each of the two types of indicators. Decentralized levels control clinical indicators (blood pressure, blood sugar etc.) while centralized levels control system indicators (waiting time, electronic health records etc.). In this paper we examine optimal pay-for-performance schemes for the two indicators by considering a model consisting of hierarchy of principal-agent interactions where pay-for-performance rewards are distributed to the centralized level (unit of accountability). We find that the optimal pay-for-performance price depends on factors such as the degree and distribution of altruistic preferences, quality costs, the marginal cost of public funds, and the interdependence between the quality variables. The optimal price should differ for system and clinical indicators both when an internal incentive system is in place and when this is not the case. The optimal price for clinical indicators is to reflect the centralized levels’ ability to steer the decentralized level - the type of internal contract that exists between the two levels of the organization. The optimal price for system indicators is independent of the type of internal contract since such indicators are under the control of the unit of accountability. Finally, it is shown that rewarding organizations on the basis of clinical quality indicators can be optimal also when such incentives are not transmitted to the decentralized level of the organization. This conclusion is the result of the indirect effects that non-incentivized variables (system indicators) might have on the incentivized ones (clinical indicators).

JEL classification: D82, I11, I18, L51

Key words: pay for performance, multi-level organizations; system and clinical quality

1 Introduction

Pay-for-performance (P4P) is the provision of incentives to physicians and provider organizations to achieve improved quality performance (Cromwell et al., 2011). The last two decades a variety of P4P programs, targeted both at primary and hospital care, have been implemented. Such programs are most common in US and England, but lately several countries have adopted, are in the process of adopting, or are discussing such
programs (Milstein and Schreyoegg, 2016). Examples are Canada, New Zealand, Taiwan, Israel, Germany and Scandinavia (Eikjenaar, 2012; Eijkenaar et al., 2013). Main impressions from the health care literature are: (i) variations in design and institutional context (Meacock et al., 2014; Van Herck et al., 2010), (ii) programs have not lived up to original expectations since the reported effects are modest, mixed and short-lived (Felt-Lisk et al., 2007; Eijkenaar et al., 2013), (iii) the magnitude of the incentives has typically been moderate (Eggleston, 2005; Rosenthal et al., 2004) and, (iv) there are gaps in the literature, including a lack of theoretical models, thus little available information to support policymakers (Felt-Lisk et al., 2007; Meacock et al., 2014).

Over time, significant resources have been invested into improving the validity and reliability of quality indicators as well as increasing their number. Such improvements in observable quality dimensions might be one important factor behind the growth of such programs. The quality indicators typically fall into three categories; structure, process, and outcome, where structures typically pertain to aspects of the facility (equipment and technology), processes use specific evidence-based elements of an encounter or episode of patient care, while outcomes are the ultimate measures of healthcare including patient satisfaction (Donebedian, 1966; Lazar et al., 2012). The number and types of quality indicators vary across programs. The Quality and Outcomes Framework (QOF) in the UK, covering 10,000 general practices, contains 146 quality indicators of which 76 are covering clinical care for 10 chronic diseases while the remaining 70 cover organization of care and patient experiences (Doran et al., 2006; Gravelle et al., 2008), of these again, 31 are classified as structural measures (Cromwell et al., 2011). Eijkenaar (2012) describes 12 P4P programs initiated in 9 different countries, United States not included, and finds that the average number of quality indicators for such programs is about 30.

Two important characteristics of P4P schemes are; (i) to whom are the performance awards distributed (the identity of the receivers; units of accountability), and, (ii) what particular quality dimensions are incentivized. In hospital programs, payments for improved performance are typically paid to the institutions (Cromwell et al., 2011). A natural question then becomes; to what extent, and in what way, does the centralized level of an organization control the incentivized decisions? Decisions made by individual physicians, clinical teams, or hospital departments, are typically measured by process and outcome indicators (clinical quality indicators). This means, if the recipient of the rewards is the centralized level of an organization, that rewards and clinical incentives are disconnected unless there are mechanisms for transferring the same incentives to lower levels. In other cases, quality is a function of centralized decisions (system quality decisions) such as the adoption of information technology (electronic health records, medical records and computerized physician order entry), waiting time, use of patient navigators, organizational culture and organization of care. Now, there is a more direct relationship between rewards and the relevant decision-making unit.

---


2 There is an extensive number of reviews (systematic reviews and meta-analysis) on P4P programs (see e.g. Christianson et al., 2008; Van Herck et al., 2010; Flodgren et al., 2011; Scott and Ouakrim, 2011; Cromwell et al., 2011; Emmert et al., 2012; Eijkenaar et al., 2013).

3 UK performance payments are directed at practices, and not to individual physicians (Smith and York, 2014).

4 In Sweden a national P4P scheme incentivized performance on waiting time targets. The Norwegian national P4P scheme, implemented in 2014, contains indicators such as the; (i) number of cancellations, (ii) number of violations of epicrisis deadlines, and, (iii) number of corridor patients (Helsedirektoratet, 2014).
The literature that evaluates pay-for-performance schemes frequently report about third-party payers rewarding organizations on the basis of clinical quality indicators while the same indicators remain non-incentivized by the organization (see e.g. Bokheur et al., 2006; Kristensen et al., 2016b; Sinofa 2015; Eikjenaar, 2013). As to the potential role different types of quality indicators may have on optimal design, this is a question not given much attention in the literature. This is somewhat surprising since various levels of an organization are responsible for different quality decisions. System variables are typically under the control of managerial levels and such decisions may have important implications for the day-to-day decisions taken at lower levels (e.g. clinical decisions).

In the following we are concerned with two questions. First, is there any point in implementing P4P schemes when the organization as unit of accountability does not have incentives in place to steer clinical decision-making? Second, do different types of incentivized indicators (system or clinical) and different internal contracts have implications for optimal design? In order to address these questions we consider a model assuming a hierarchy of principal-agent interactions (three-tier model) consisting of a funder (top level), a centralized level (intermediate level) and a decentralized level (bottom level). The funder acts as a principal and the intermediate level (“the hospital”) as her agent, while the intermediate level acts a as a principal towards the lowest level (“the clinician”).

Our model is related to the hierarchy literature. This literature is concerned with the costs of hierarchies (private information and monitoring costs), collusion incentives, and the implications from hierarchies relative to centralized contracting. An early work in this field is Tirole (1986) that includes a third layer (a supervisor) into a standard principal-agent model where the principal cannot observe agreements and payments among subordinates. The introduction of a supervisor, however, creates new problems (incentives for coalition formation) if the interest of the supervisor is in conflict with the ones of the principal. Mookerherjee (2006) provides an overview of theoretical research on delegated decision making in hierarchical organizations based on an application of mechanism design theory. According to this review most of the literature is concerned with agents having private information about their costs (adverse selection) and a typical model is one with a single principal, one or two producing suppliers, and a manager (monitor) that plays no role in production. Important conclusions are that the main trade-off in designing contracts (delegation) is between productive efficiency and paying information rents to agents and that centralized decision-making cannot be dominated by any delegation arrangement.

The theoretical literature on P4P health care schemes is limited. Eggleston (2005) presents a model with an altruistic provider that makes two quality decisions (one verifiable and one non-verifiable) and shows that a performance program may increase the verifiable one and decrease the non-verifiable one. Kaarboe and Siciliani (2011) extend

5 Demski and Sappington (1987) study a three-tier hierarchy in which the principal designs all contracts. The intermediate agent (regulator) is a self-interested strategic player being able to gather improved information about the agents’ productivity. The principal must motivate the intermediate agent to acquire the information. Mookerherjee (2006) presents various organizational alternatives of centralization and delegation. The alternative that is closest to our model is termed “Delegation of Suppliers”. Here the principal only contracts with the agent at the intermediate level and delegates authority over contracting with the agent at the bottom level. Our model differs from this literature since applying an incomplete contract framework (moral hazard). Furthermore, our model differs from the “Delegation of Suppliers” since the productive decision of the bottom level can be non-contractible for the intermediate level and/or the top-level and since the productive decision of the intermediate level can be non-contractible for the top-level.

6 This conclusion follows since not considering possible delegation benefits.
the analysis of Eggleston (2005) by solving explicitly for the optimal P4P scheme. Both works relate to the multitasking literature since studying the difficulties of selective payment for performance when there are many tasks expected of a supplier (see Holmstrom and Milgrom, 1991; Dewatripoint et al., 1999). Other works are Kristensen et al., (2016a), comparing hospitals that distribute performance payments with those that do not, and Kristensen et al., (2016b), concerned with the actual P4P prices offered to hospitals relatively to the theoretically optimal ones. The above works study multitasking problems in an incomplete contracting setting where the provider has altruistic concerns but assume a non-hierarchical structure.\(^8\) \(^9\)

We find that optimal P4P schemes may improve social welfare also when the hospital is unable to steer clinical decision-making (absent within-hospital incentive schemes). For incentivized system indicators this follows because the hospital both is the unit of accountability and decides on system quality. For incentivized clinical indicators the hospital may induce changes in clinical decisions via the indirect effect that non-incentivized variables (system indicators) have on the incentivized ones (clinical quality). We find that optimal P4P prices differ across system and quality indicators both when an internal incentive system is in place and when this is not the case (the type of internal contract). The optimal prices relate differently to quality costs and because the interdependencies between the quality variables (the responsiveness) are different. Finally, it is shown that optimal prices, under some conditions, can be negative (“penalizing-for-performance”). The remainder of the paper is as follows. In section 2, we present the model and the first-best solution. In section 3, we derive the optimal P4P-prices for system and clinical variables. Section 4 concludes.

2 The model and the first-best solution
Consider the centralized level of a provider organization (“the hospital”) that decides on system quality, \(h\), and delegates clinical quality decisions, \(c\), to a lower level of the organization (“clinician”). The hospital may, or may not, have an internal incentive system in place. Both decision-makers (the hospital and the clinician) are semi-altruistic. A third-party payer (“the funder”) reimburses the organization on the basis of a contractible quality variable (a system quality indicator or a clinical quality indicator). The hospital is the recipient of the P4P rewards (unit of accountability) and all three decision-makers are concerned with patient benefits but their objective functions differ. We abstract from risk considerations and assume passive patients.

2.1 Patient benefits
The patients’ benefits from treatment, \(U(h,c)\), are increasing and concave in each of the two quality variables while the cross partial derivative of the patient benefit function can be positive, negative or zero; \(U_{hc}(c,h) = U_{h,c}(c,h) > (\leq) 0\). \(^{10}\)

\(^8\) Several works include patients’ utility as part of provider utility functions (Ellis and McGuire, 1986; 1990; Danzon, 1994; Chalkley and Malcomson, 1998; Jack, 2005). See also the literature on pro-social behaviour and intrinsic motivation and implications for performance incentives (see Benabou and Tirole, 2003 and 2006).

\(^9\) A recent trend is to promote health care quality by using disincentives to encourage desired behavior. Examples are non-payments for “never-happen-events” and medical errors (see Grepperud, 2005 and 2007; Rosenthal, 2007; Meacock et al., 2014; Eager et al., 2013).

\(^{10}\) For simplicity third derivatives are assumed to be zero.
2.2 The clinician (the bottom level)

The clinician (deciding on \(c\)) is partly altruistic since the objective function is an additive function of a share, \(\alpha\), of patient benefits, \(U(c,h)\), where \(0 < \alpha < 1\), and net income, \(y\). Net income is defined as the income that follows from an internal contract, consisting of a lump-sum payment (a fixed wage component; \(A\)) and, possibly, a part that makes income contingent upon clinical quality (\(b\)), subtracted clinical quality costs, \(kc\), where \(k > 0\).\(^{11}\) Thus the benefit function of the clinician, \(V^c(c,h)\), becomes:

\[
V^c(c,h) = \alpha U(c,h) + y = \alpha U(c,h) + A + bc - kc
\]

The clinician maximizes (1) with respect to \(c\), given \(V^c(c,h) \geq V^c\), where \(V^c\) is the reservation benefit.

2.3 The hospital (intermediate level)

The hospital decides on system quality, \(h\), and the internal contract (\(A\) and \(b\)). The objective function, \(V^h(c,h)\), depends on a share, \(\beta\), of patients’ benefits, \(U(c,h)\), and profits. Hospital costs are the sum of the expenses that follow from the internal contract and system quality costs, \(Kh\), where \(k > 0\) is the system quality unit cost.\(^{12}\) The hospital is reimbursed by the funder by means of a fixed budget, \(R\), and a linear payment scheme, where \(P\) is the income per unit of quality.\(^{13}\) \(P\) is either reimbursed on the basis of a system quality indicator, \(P^s\), or a clinical quality indicator, \(P^c\), thus rewarding on the basis of absolute performance.\(^{14}\) The hospital objective function now becomes:

\[
V^h(c,h) = \beta U(c,h) + \Pi = \beta U(c,h) - Kh - (A + bc) + R + [\gamma P^s h + (1 - \gamma) P^c].
\]

From the hospital objective function it follows that for \(\gamma = 1\), the hospital is reimbursed on the basis of a system quality indicator while for \(\gamma = 0\) it is reimbursed on the basis of a clinical quality indicator. By inserting a binding version of the participation constraint of the clinician, \(V^c(c,h) = V^c\), into the hospital objective function, we get:

\[
V^h(c,h) = (\alpha + \beta) U(c,h) - Kh - kc + [\gamma P^s h + (1 - \gamma) P^c] + R - V^c
\]

We observe that the preferences and costs of the clinician, \(\alpha U(c,h)\) and \(kc\), are parts of (2), thus both elements are internalized by the hospital. This property follows since, for a binding participation constraint, the clinician must be compensated financially by the hospital. In the following it is assumed that \(\alpha + \beta < 1\) meaning that patients’ benefits are not fully internalized by the hospital.

---

\(^{11}\)The concept of paternalistic altruism is often used to describe preferences for which specific elements of other utilities are included (Archibald and Donaldson, 1976).

\(^{12}\)We assume linear quality costs for the hospital and the clinician, however, in section 3 we also discuss optimal prices given non-linear quality costs.

\(^{13}\)Both the P4P price, \(P\), and within-organization payments, \(b\), are contingent upon quality performance. In the following \(P\) is termed the “price” while \(b\) is termed the “bonus”.

\(^{14}\)Alternative payment structures are performance relative to; (i) past performance (improvements), and, (ii) the performance of other provider organizations.
2.4 The funder (the top level)

The funder decides on the size of the global budget, $R$, and the price, $P^s$ or $P^c$, and maximizes an objective function, $W$, equal to the difference between the patients’ benefits from treatment, $U(c,h)$, and the treatment costs adjusted for the cost of public funds, $(1+\delta)(K_h+k_c)$, where $\delta$ is the marginal deadweight loss from taxation;\(^{15}\)

$$W(c,h)=U(c,h)-(1+\delta)(K_h+k_c) \tag{3}$$

The objective function in (3) is a social welfare function where the social benefits equal patients’ own valuation of treatment benefits while the social costs are the resource consumption associated with treating patients adjusted for the cost of public funds.\(^{16}\) A strictly utilitarian approach implies a summation of the benefits of the patient, the hospital, and the clinician, which again produces the following objective function; $(1+\alpha+\beta)U(c,h)-(1+\delta)(K_h+k_c)$. Such an approach, however, implies a “triple” counting of patients’ benefits which is not particularly satisfying since the social value of treatment increases with the number of actors involved in treatment (see Chalkley and Malcomson (1998) for a discussion). A fully informed funder maximizes (3) w.r.t. $c$ and $h$, which produces the following first-order conditions;\(^{17}\)

$$U_c(c^*,h^*)=(1+\delta)k \quad \text{and} \quad U_h(c^*,h^*)=(1+\delta)K \tag{4}$$

Former assumptions combined with strictly positive values of $K$ and $k$ ensure interior solutions. It follows from (4) that the first-best is achieved by equating the marginal benefit from each quality variable with the marginal unit cost (adjusted for the marginal cost of public funds), this solution, however, is unattainable since one of the two quality variables is non-contractible.\(^{18}\)

3 Model analysis

In the following two models are considered that differ with respect to the type of indicator being incentivized. For $\gamma=1$ the funder is incentivizing a system indicator (Model S), thus $h$ is observable and contractible for the funder and the hospital. The clinical variable, $c$, is unobservable for the funder while for the hospital it is; (i) contractible, or, (ii) non-

\(^{15}\) The objective function in (3) implies that the funders’ maximization problem becomes independent of the hospital participation constraint, however, the funder can still set $R$ by fulfilling the participation constraint of the hospital. The size of $R$ will depend on whether assuming a non-financial $(V^{nh}(c,h)\geq V^n)$ or a financial $(\Pi(c,h)\geq \Pi)$ participation constraint.

\(^{16}\) A different approach would be to assume a funder that is concerned with the difference between patient benefits, $U(c,h)$ and budget expenses (adjusted for the cost of public funds), $(1+\delta)[R+(\gamma P^h+1-\gamma)P^c)]$. Now the objective function becomes a function of the chosen participation constraints for the hospital and the clinician. Given non-financial participation constraints, $(V^{nh}(c,h)\geq V^n$ and $V^c(c,h)\geq V^c$), the objective function becomes (ignoring constants): $W(c,h)=U(c,h)\theta-(1+\delta)(K_h+k_c)$ where $\theta=1+(1+\delta)(\alpha+\beta)$. For a non-financial hospital constraint and a financial clinical constraint, $(\gamma(c,h)\geq \gamma)$. we get $\theta=1+(1+\delta)$. For a financial hospital constraint, $(\Pi(c,h)\geq \Pi)$ and a non-financial clinical constraint, and when both constraints are financial, we get $\theta=1$, thus the funder objective function in the last two cases coincide with (3).

\(^{17}\) Superscript * denotes first-best levels. The first-best is defined when taking the costs of public funds into account. An alternative would be to define the first-best when ignoring such costs. For expositional simplicity, we do not denote equilibrium values.

\(^{18}\) App. a provides the second-order condition for problem (3).
contractible. A contractible clinical variable means that an internal incentive contract contingent upon the same variable can be established \((b \neq 0)\) while a non-contractible clinical variable means that the internal contract becomes a fixed wage contract \((b = 0)\). For \(\gamma = 0\), the funder is incentivizing a clinical indicator (Model \(C\)), thus \(c\) is observable to all parties and contractible for the funder while \(h\) is non-contractible. Furthermore, the clinical variable can be both contractible \((b \neq 0)\) or non-contractible \((b = 0)\) for the hospital.

Two models and two internal contracts imply four cases; \(SI\) (incentivized system quality and an internal incentive contract), \(SF\) (incentivized system quality and fixed wages), \(CI\) (incentivized clinical quality and an internal incentive contract) and \(CF\) (incentivized clinical quality and fixed wages). The various cases are solved as three-stage games where the funder first decides on the global budget, \(R\), and the price, \(P\). In the second stage, the hospital, for given levels of \(R\) and \(P\), decides on system quality and the internal contract \((A\) and \(b\)). Finally, the clinician, for a given internal contract and a given system quality level, decides on clinical quality. We apply backward induction to solve the problem and the equilibria are subgame-perfect Nash equilibria.

### 3.1 The problem of the clinician

The clinician maximizes (1) with respect to clinical quality, which yields (the condition for an interior solution in available from App. \(b\));

\[
H_c(c,h) = \alpha U_c(c,h) + b - k = 0 \Rightarrow \alpha U_c(c,h) = k - b
\]

(5)

Optimal clinical quality is determined by equality between the marginal utility that derives from altruistic concerns and the difference between clinical quality unit costs and the bonus. We observe that the price has no direct impact on clinical quality since the rewards are distributed to the hospital. Condition (5) also defines the response function of the clinician;

\[
c = c(h,b)
\]

(6)

The partial effects become;

\[
c'_c(h,0) = c'_c(h,b) = -\frac{U_{\alpha}(c,h)}{U_{\alpha}(c,h)} \geq (\leq) 0
\]

(7)

\[
c'_c(h,b) = -[1/\alpha U_{\alpha}(c,h)] > 0
\]

(8)

From above we see that clinical quality is strictly increasing with \(b\), while it may increase, decrease, or is independent of \(h\). From (7) it follows that the cross partial derivative of the patient benefit function defines the direction and the degree of the responsiveness between the quality variables (“the responsiveness effect”). The two variables \((c\) and \(h\)) can be strategic substitutes \((U_{\alpha,h} < 0 \Rightarrow c_h < 0)\), strategic complements \((U_{\alpha,h} > 0 \Rightarrow c_h > 0)\) or strategic independent variables \((U_{\alpha,h} = 0 \Rightarrow c_h = 0)\).
3.2 The problem of the hospital

In the second stage the hospital decides on system quality and the internal contract by maximizing (2), taking the response function of the clinician into account (see 6), which yields:

\[
\max_{h,b} V^H(c,h) = (\alpha + \beta)U(c(h,b),h) + R + P[\gamma h + (1-\gamma)c(h,b)] - Kh - kc(h,b) \tag{9}
\]

By maximizing (9) we get the following expression for the optimal bonus (see App. c for a complete description of first and second order conditions);

\[
b = \frac{\beta}{\alpha + \beta}k + (1-\gamma)\frac{\alpha}{\alpha + \beta}P \tag{10}
\]

We observe that the optimal bonus is independent of the price when system quality is incentivized, \(\gamma = 1\), while being strictly increasing when clinical quality is incentivized, \(\gamma = 0\). The separation between the bonus and the price for incentivized system quality follows because the incentivized decision, \(h\), is not taken by the clinician. Given incentivized clinical quality, on the other hand, the hospital has an interest in encouraging the clinician to invest more resources into quality improvements.

From (9), given an internal incentive contract, the first-order condition for system quality becomes (conditions for interior solutions are available from App. c);

\[
(\alpha + \beta)U_s(c(h,b),h)) = K - \gamma P \tag{11}
\]

Given incentivized system quality, \(\gamma = 1\), we see that the hospital takes the price into account when deciding on \(h\). Given incentivized clinical quality, \(\gamma = 0\), however, the price has no direct effect on \(h\). The first-order condition for system quality, given fixed wages, becomes;

\[
(\alpha + \beta)U_s(c(h,0),h)) = K - \gamma P + \left(1 - \gamma\right)P - \frac{\beta}{\alpha}k c_s(h,0) \tag{12}
\]

The price now enters the first-order condition both for incentivized system quality and incentivized clinical quality. For incentivized clinical quality the effect is indirect since the price works through the responsiveness effect \((c_s)\), implying that the last term in (12) acts as an imperfect substitute to an internal incentive contract.

The partial effects from a higher price on the quality variables are available from App. d. A higher price increases system quality when system quality is incentivized \((SI\) and \(SF)) and increases clinical quality when clinical quality is incentivized \((CI\) and \(CF))

The sign of the spill-over effects (the effect from a higher price on the non-incentivized quality variable) depends on the sign of the responsiveness-effect. Both for incentivized system quality and incentivized clinical quality a higher price will downgrade the non-incentivized quality variable (“teaching to the test”) given strategic substitutes \((c_s < 0)\) and

---

19 Problem (9) defines the response functions of the hospital; \(h(P)\) and \(b(P)\) [see (c1) and (c2) in App. c]. By inserting \(h(P)\) and \(b(P)\), into (6), we get the following response function for the clinician; \(c = \hat{c}(h(P),b(P))\).
upgrade the non-incentivized quality variable (“attention reinforcement”) given strategic complements \( (c_h > 0) \).

3.3 The problem of the funder

In the third stage the funder decides on the price by maximizing social welfare (see 3) taking into account the response-functions of the hospital and the clinician, which yields:

\[
\text{Max}_p \quad W(c, h) = U(\bar{c}(h(P), b(P)), h(P)) - (1 + \delta)[Kh(P) + k\bar{c}(h(P), b(P))] \tag{13}
\]

The maximization problem yields the following first-order condition (the second-order condition is presented in App. e):\(^{20}\)

\[
[U_h - (1 + \delta)K]h_p + [U_c - (1 + \delta)k]\varepsilon = 0 \quad \text{where} \quad \varepsilon = c_h h_p + c_b b_p \tag{14}
\]

In what follows we present the optimal price for each of the four cases (see App. f for further details on the identification of the optimal prices). The price for SI becomes;

\[
P^{SI} = [1 - (1 + \delta)(\alpha + \beta)](K + kc_h) \tag{15}
\]

Besides quality costs and the responsiveness-effect, the optimal price is a function of a term that reflects two externalities; \( \Psi = [1 - (1 + \delta)(\alpha + \beta)] > 0 \). The first one relates to social benefits since all patient benefits are not internalized; \( \alpha + \beta < 1 \). The second one relates to social costs since the costs of public funds are not internalized (in the forthcoming discussion it is assumed that \( \Psi > 0 \)). We observe that the optimal price decreases, the higher the share of internalized benefits, \( \alpha + \beta \), and the higher the marginal cost of public funds, \( \delta \).

The optimal price is strictly positive (“reward”) if the quality variables are complements or independent. For substitutes, the optimal price becomes zero if the responsiveness-effect is equal to the unit cost ratio, \( c_h = -k / K \), while being negative if the responsiveness-effect is lower than the unit cost ratio, \( c_h < -k / K < 0 \), thus introducing a “penalty - for-performance”.\(^{21}\)

The optimal price for SF becomes;

\[
P^{SF} = [1 - (1 + \delta)(\alpha + \beta)](K + kc_h) \tag{16}
\]

The optimal price in (16) is identical to the one derived for SI thus for incentivized system quality the presence of an internal incentive system has no bearing on the optimal price.\(^{22}\)

This conclusion follows, since for SI and SF, the incentivized decision is under the direct control of the unit of accountability (the hospital).

\(^{20}\)To simplify notation, arguments are in the following omitted.

\(^{21}\)An alternative formulation of (15) is (arguments omitted); \( P^{SI} = \Psi[U_c + U_c c_h] \), saying that the optimal price is to equal a share, \( \Psi = [1 - (1 + \delta)(\alpha + \beta)] \), of “the marginal patient benefit of higher system quality adjusted for the responsiveness of the two quality variables times the marginal benefit of clinical quality.”

\(^{22}\)This conclusion follows because all third-derivatives are assumed to be zero. In general, however, the equilibrium values \((h, c)\) will differ for different internal contracts.
Figure 1: The optimal price for SI and SF and the responsiveness-effect.

Figure 1 describes the relationship between the optimal price and the responsiveness-effect (both for SI and SF). As already shown, the optimal price is strictly positive for complements while for substitutes it can be both positive and negative. Social welfare for SI is always higher than social welfare for SF since the incentive contract of SI ensures that hospital altruistic concerns are internalized by the clinician.\(^{(23)}\)

Next we analyze incentivized clinical quality indicators. The optimal price for CI becomes:

\[
P_{CI} = [1 - (1 + \delta)(\alpha + \beta)] \left[ k + K \frac{h_p}{c_i h_p + c_i b_p} \right]
\]

A first observation is that the responsiveness-effect for CI now differs from the same effect for SI and SF. The added complexity follows because the effect on clinical quality now works through both decisions made by the hospital (system quality and bonus). Under

\(^{(23)}\) For simplicity consider independent variables. The first-order conditions given SI (see 15) now become; 
\[U_i(c^i, h^i) = (1 + \delta)K\] and \((\alpha + \beta)U_i(c^i, h^i) = k\), while the first-order conditions given SF (see 16) become; 
\[U_i(c^{SF}, h^{SF}) = (1 + \delta)K\] and \(\alpha U_i(c^{SF}, h^{SF}) = k\). From this we get; 
\(h^i = h^{SI} = h^{SF}\) and \(c^i > c^{SI} > c^{SF}\), thus 
\(W(c^i, h^i) > W(c^{SI}, h^{SI}) > W(c^{SF}, h^{SF})\).
SI and SF, in contrast, clinical quality is only influenced by adjustments in system quality. It can be shown that (see App. g);  
\[
\frac{h_p}{c_p h_p + c_p b_p} = -\frac{U_{c_h}}{U_{h_b}} = h_c
\]  
(18)

From (18) it follows that the responsiveness-effect in (17) can be re-expressed as the responsiveness of \(h\) with respect to \(c\). By inserting (18) into (17) we get;
\[
P^{CI} = \left[1 - (1 + \delta)(\alpha + \beta)\right](k + kh_c)
\]
(19)

The optimal pricing rule for \(CI\) has a structure similar to the ones of SI and SF but differs in two important ways. First, because \(h_c = U_{c_h}/U_{h_b}\) differs from \(c_c = U_{c_c}/U_{c_c}\). The responsiveness-effect for all three cases refer to how the incentivized variable affects the non-contractible one, but since the incentivized variable in \(CI\) differs from the one in SI and SF, the effects differ. Second, the optimal price relates differently to the quality unit costs. For \(CI\), the optimal price increases with a higher clinical unit cost while it may decrease, increase or be independent of the system quality unit cost.  

The relationship between the price and the responsiveness-effect for \(CI\) is described in Figure 2.

**Figure 2:** The optimal price for CI and the responsiveness-effect.

---

24 \(-\left(\frac{U_{c_h}}{U_{h_b}}\right)\) follows from differentiating the first-order condition, \((\alpha + \beta)U_c(c,h) - K = 0\), with respect to \(h\) and \(c\).

25 Given independent quality variables the optimal prices become as follows; \(P^{SI} = P^{SF} = \Psi K\) and \(P^{CI} = \Psi k\).
The final case to consider is CF. We arrive at the following expression;

\[
p^{\text{CF}} = \frac{[1 - (1 + \delta)(\alpha + \beta)]}{c_h} (K + kc_h) = [1 - (1 + \delta)(\alpha + \beta)] \left[ k + \frac{K}{c_h} \right] \quad \text{for } c_h \neq 0
\]

(20)

It follows from (20) that no solution exists for independent quality variables; \( U_{ch} = 0 \Rightarrow c_h = 0 \). The only way the funder now can steer the clinician is by influencing system quality. It follows that;

\[
\lim_{c_h \to -\infty} P^{\text{CF}} [1 - (1 + \delta)(\alpha + \beta)]k > 0 \quad \lim_{c_h \to +\infty} P^{\text{CF}} [1 - (1 + \delta)(\alpha + \beta)]k > 0 \quad \lim_{c_h \to -\infty} P^{\text{CF}} \rightarrow -\infty
\]

The relationship between the optimal price and the responsiveness-effect is now non-linear. As the responsiveness-effect approaches zero, the optimal price increases (or decreases) at an increasing rate. Thus a modest or insignificant responsibility-effect makes it necessary to set the optimal price high (complements) or low (substitutes) to achieve the desired changes in clinical quality.

**Figure 3:** The optimal price for CF and the responsiveness-effect.

The relationship between the responsiveness-effect and the optimal price for CF is described in Figure 3. Given “attention reinforcement” (complementarity), a lower spill-over effect increases the optimal price, while for “teaching-to-the-test” (substitutability) a

---

26 By comparing (20) with (15) and (16) we observe that \( c_hP^{\text{CF}} = P^{SI} = P^{SP} \).
lower spill-over effect (in absolute terms) induces a more negative price ("penalizing-for-performance"). Furthermore, a higher responsiveness effect (in absolute terms) produces a price that approaches \(1-(1+\delta)(\alpha+\beta)\) for both complements and substitutes.

Social welfare for \(CI\) is higher than for \(CF\) because the internal incentive contract ensures that hospital altruistic concerns are internalized by the clinician.\(^{27}\) An interesting observation is that the price is strictly positive ("reward") when the responsiveness-effect is significant while a negative price ("penalty") is optimal when the same effect is moderate or weak. These findings differ from the conclusions arrived at for the three preceding cases. For substitutes ("teaching-to-the test"), all cases yield a price equal to zero when the input cost ratio (in absolute terms) equals the magnitude of the responsiveness-effect. Furthermore, in all cases the price depends on the share of patients’ benefits being internalized by the hospital; \((\alpha+\beta)\). This finding follows since the targeted decision-maker is the same in all cases (the hospital as unit of accountability). The absence of hospital altruism (\(\beta=0\)) implies a higher optimal P4P price in all four cases. This means that that optimal price vary across pure profit hospitals and non-profit hospitals. It is somewhat surprising that the responsiveness-effect for \(CF\) (\(c_h\)) has an interpretation opposite of the ones identified for \(SI, SF\) and \(CI\). For \(CF\) this effect refers to how the non-incentivized quality variable affects the incentivized one and follows because the incentivized variable is not under the direct control of the hospital. For \(CI\), in contrast, the hospital steers clinical quality via the internal incentive contract. Finally, it is worth mentioning that the significance of the responsiveness-effects for \(CF\) and \(CI\) typically differ for a similar technology. This again may have important implications for optimal design.

An extension of our analysis is to replace the non-linear quality costs, \(kc\) and \(Kh\), with the following cost functions; \(k(c;h)\) and \(K(h;c)\) where \(k_c(c;h) \geq 0, k_h(c;h) \leq (>0), K_k(h;c) \geq 0\) and \(K_h(h;c) \leq (>0)\). Non-linear cost functions allow for the quality costs of the clinician (the hospital) to vary with the quality decisions of the hospital (the clinician). If so, the welfare function now equals \(W(c,h)=U(c,h)-(1+\delta)[K(h;c)+k(c;h)]\), and the optimal price for \(SI\) becomes,

\[
\overline{P}^\delta = [1-(1+\delta)(\alpha+\beta)]\left[K_h + k_c(K_h+k_c)c_h\right] \text{ where } c_h = -\left(aU_{c,c}-k_c\right)/Z \text{ and } Z = aU_{c,c}-k_c < 0. \tag{21}
\]

From (21) we see that the optimal price takes into account how; (i) the quality decision of the hospital affects clinician quality costs \((k_c)\), and, (ii) the quality decision of the clinician affects hospital quality costs \((K_h)\). We notice from the expression for \(c_h\) that the responsiveness-effect now is more complex since depending on the cross partial derivative of the clinician cost-function. It follows that the differences identified across the optimal prices for linear costs are valid for non-linear costs.\(^{28,29}\)

\(^{27}\) To see this consider independent quality variables. The first-order conditions for \(CI\) (see 17) become;
\[(\alpha+\beta)U_c(c^*,h^*) = K \text{ and } U_c(c^*,h^*) = (1+\delta)k\), the first-order conditions for \(CF\) (see 20) become;
\[(\alpha+\beta)U_c(c^*,h^*) = K \text{ and } aU(c^*,h^*) = (1+\delta)k\). From this we get; \(h^* > h^C = h^C\) and \(c^* = c^C > c^C\), thus \(W(c^*,h^*) > W(c^C,h^C) > W(c^C,h^C)\).

\(^{28}\)\(^{29}\) \(P^\delta\) has a structure similar to (21). The expressions for the remaining two prices are;

\[
\overline{P}^\delta = [1-(1+\delta)(\alpha+\beta)]\left[K_h + k_c(K_h+k_c)h_h\right] \text{ and } P^\delta = \overline{P}^\delta / c_h.
\]

\(^{29}\) Implications on optimal prices from non-linear cost functions are discussed in Kaarboe and Siciliani (2011).
The quality incentives directed at the hospital for CF are not transmitted to the clinician meaning that clinical quality is contracted upon by the funder and the hospital but not by the hospital and the clinician. Such a situation is somewhat surprising since one would expect the hospital to implement an incentive contract. On the other hand there are frequent observations reported in the literature (see the references in the introduction) about hospitals being rewarded on the basis of clinical quality indicators where the same incentives are not transmitted to the clinical level. Our analysis shows that CF, in the presence of a responsiveness-effect, is welfare-improving relatively to laissez faire, thus our analysis can be said to provide a rationale for the existence of case CF.

CF observations suggest a reluctance among hospitals to use internal incentive contracts to improve quality. Possible explanations, lying outside of our model, are; (i) resistance among organized groups (trade unions) to institute rewards at decentralized levels (institutional barriers), (ii) events that are eligible for performance measurement occur too infrequently to allow for statistically reliable and valid measurements of performance at decentralized levels (sample size issues), (iii) several clinical teams and departments are involved in each treatment episode (measurement problems), (iv) the need to foster cooperation among employees (team incentives), and (v) risk considerations. An additional possibility is that clinical decision-making is effectively influenced by non-financial incentives. For example, pay-for-performance schemes may convey a message to clinicians about the organizational goals been given priority that again triggers a change in behaviour.

An alternative to CF would be for the funder to contract with the clinical level. Such a strategy, however, may incur high contract costs since such costs, in comparison to contracts with the hospital level, involve a higher number of contractual partners. Contracts between funders and hospitals may also involve lower contract costs because the hierarchical distance is shorter. This is particularly likely in national health systems where funders and hospitals often are the same judicial entity which suggests relatively low costs in writing and enforcing contracts.

4 Conclusions

The use of pay-for-performance has become widespread in US and UK, and lately, additional health care systems have adopted such schemes. In this paper we extend former analysis by suggesting a model that pays attention to the hierarchical structure of organizations. A hierarchical model allows for the analysis of quality decisions taken at different levels and enables us to distinguish between the level that acts as unit of accountability and the level that does not. Such a framework is both relevant and necessary for analyzing the possible roles system and clinical quality indicators play. Our main findings are; (i) the presence of spill-over effects may explain the existence of performance incentives targeted at the centralized level of an organization when such incentives are not distributed to lower levels, and, (ii) the optimal structure of a pay-for-performance scheme depends on the type of decisions being incentivized (system or clinical quality) and the type of internal contract.

The take home message is that performance-pay-prices should differentiate between clinical and system indicators. Furthermore, the price for clinical indicators should pay attention to a hospitals' ability to steer clinical quality decisions. For system

30 Some literature discusses the desirability of the implementation of team-orientated versus individual financial incentives (see Che and Yoo, 2001; Kvaløy and Olsen, 2012; Lamantia and Pezzino, 2016).

31 Infinite risk aversion implies that agents do not respond to monetary incentives (Aghion and Tirole, 1997).
quality indicators, however, this is not the case, since the relevant decision is under the direct control of the unit of accountability (the hospital). We also find that internal incentive contracts are unresponsive to the implementation of pay-for performance when system quality is incentivized. For incentivized clinical quality indicators, however, such contracts become sensitive to a pay-for-performance scheme; the higher the price, the higher the power of the internal incentives directed at the clinical level.

In addition to the incentivized decision and the centralized levels’ ability to steer clinical decisions, optimal prices depend on factors such as the degree and distribution of altruistic preferences, quality costs, the cost of public funds, and the interdependence between the quality variables. For quality variables being substitutes, the optimal price may be negative (“penalizing-for-performance”) and for quality variables being complements, the optimal price is always strictly positive (“rewards”). Furthermore, social welfare is higher for internal incentive contracts relatively to fixed wages both for incentivized system quality and incentivized clinical quality.

The literature that evaluates pay-for-performance schemes refers to rather disappointing quality effects and points to modest prices as one possible explaining factor. Our analysis is not directly concerned with the identification of conditions that induce such outcomes, however, we show that modest or weak prices, under some circumstances, might be optimal. In sum our findings confirm that knowledge about management systems and the organizational structures within an organization becomes important for optimal incentive design.

Our analysis also identifies a particular situation where pay-for-performance schemes will not improve quality performance. This occurs for incentivized clinical quality indicators when hospitals are without an ability to steer clinical decisions and there is not interdependence between quality variables (absent spill-over effects). In such situations, alternative strategies would be to incentivize system quality decisions or to contract directly with clinical decision-makers. The relatively high number of quality indicators means that many quality dimensions are observable and contractible. In view of this it is somewhat surprising that hospitals, in response to the introduction of performance schemes, may choose not to transmit the quality incentives to the clinical level. This is a question that deserves more attention in future research.

References


© 2019 by the author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).
Appendices

Appendix a: The second-order condition for problem (3)

The second-order condition for problem (3) is:
\[ D^W = W_{hh} W_{cc} - W_{hc} W_{ch} = U_{hh} U_{cc} - (U_{hc} U_{ch}) < 0 \]  (a1)

where \( D^W \) is the determinant of the Hesse matrix. From A2 it follows that (a1) is fulfilled.

Appendix b: The condition for an interior solution to problem (5)

The patients’ benefits from treatment increase at a decreasing rate with each of the two quality variables and are concave in \((c, h)\); \( U'_c(c, h) > 0, U''_c(c, h) < 0, U'_h(c, h) > 0, U''_h(c, h) < 0 \) and \( U_{hc}(c, h) U_{hh}(c, h) - U_{ch}(c, h) U_{hh}(c, h) > 0 \). To ensure a strictly positive quality level, we assume that:
\[ \lim_{c \to 0+} U_c(c, h) < \infty, \quad \lim_{c \to 0+} U'_c(c, h) = 0, \quad \lim_{c \to 0+} U''_c(c, h) > 0 \] and
\[ \lim_{h \to 0+} U_h(c, h) < \infty, \quad \lim_{h \to 0+} U'_h(c, h) = 0, \quad \lim_{h \to 0+} U''_h(c, h) > 0. \]

For fixed wages the first-order condition yields an interior solution for \( k > 0 \). For incentive contracts the first-order condition yields an interior solution for \( b > k \).

Appendix c: The first-order and second-order conditions for problem (9)

The first-order conditions for problem (9) become:
\[ V^H(h, b) = c'_h(h, b)[(\alpha + \beta) U'_c(c(h, b), h) + P(1 - \gamma) - k] = 0 \]  (c1)
\[ V^H(h, b) = (\alpha + \beta)[U'_c(c(h, b), h) + U'_c(c(h, b), h) c'_h(h, b)] - k c'_h(h, b) + P[\gamma + (1 - \gamma) c'_h(h, b)] = 0 \]  (c2)

From former assumptions we get an interior solution for (c1) if \( \gamma = 1 \). For \( \gamma = 0 \), (c1) has an interior solution if \( k > P \). By inserting (5) into (c2) we arrive at (10). It follows from (10) that for \( \gamma = 1 \), the requirement for an interior solution, \( k > b \) is fulfilled. Using (c1) and (c2) we arrive at the first-order condition for system quality given an incentive contract (see 11). Given former assumptions, (11) has an interior solution for \( \gamma = 0 \) since \( K > 0 \). For \( \gamma = 1 \), (11) has an interior solution if \( K > P \). Using (c1) and (c2), the first-order condition for system quality, given a fixed wage contract, we arrive at (12). Given former assumptions, (12) has an interior solution for \( \gamma = 1 \) if \( k - P > 0 \) while for \( \gamma = 0 \), (12) has an interior solution if \( K + (P - (\beta / \alpha) k) c_h > 0 \). The condition for an interior solution for system quality given \( SI \), \( k - P SI > 0 \), is fulfilled for \( [(1 + \delta)(\alpha + \beta)K] / |\Psi| > c_h \). The condition for an interior solution system quality given \( SF \) is \( K - P SF - (\beta / \alpha) k c_h > 0 \) which is fulfilled for \( |\Psi a K| / (\beta + (1 - \gamma) \alpha) c_h > 0 \). For \( CF \), the condition for an interior solution for system quality is \( |(K + (1 - \Psi) a K)| / (\beta k - (1 - \Psi) a K) > c_h \).

\[ D^H = V^H_{hh} V^H_{cc} - V^H_{hc} V^H_{ch} > 0. \]  (c3)
Given third-derivatives equal to zero, we get:

\[ V_{hh}^N = (\alpha + \beta)[U_{hh} + c_nU_{hn}] < 0 \]  
\[ V_{hc}^N = (\alpha + \beta) c_n U_{hn} = 0 \]  
\[ V_{cc}^N = (\alpha + \beta)(c_n)^2 U_{cc} < 0 \]  

Inserting (c4) - (c6) into (c3) yields:  
\[ D^N = U_{cc} (\alpha + \beta)(c_n)^2 U_{cc} - (U_{hh})^2 \]  
From (c4) - (c6) it follows that \( D^N > 0 \) (see c7).

**Appendix d: Partial effects (the two quality variables and the bonus)**

Case SI:  
\[ h_p = -\frac{1}{(\alpha + \beta)Y} > 0 \quad \text{and} \quad b_p = 0 \quad \Rightarrow \quad h = h(P), \quad \text{and} \quad c = \bar{c}(h(P),0) \]  
(Case SF):  
\[ h_p = -\frac{1}{(\alpha + \beta)Y} > 0 \quad \text{and} \quad b_p = 0 \quad \Rightarrow \quad h = h(P) \quad \text{and} \quad c = \bar{c}(h(P),0) \]  
(Case CI):  
\[ h_p = -\frac{c_n}{(\alpha + \beta)Y} > (\leq) 0 \quad \text{and} \quad b_p = -\frac{\alpha}{\alpha + \beta} > 0 \quad \Rightarrow \quad h = h(P), b = b(P) \quad \text{and} \quad c = \bar{c}(h(P),b(P)) \]  
(Case CF):  
\[ h_p = -\frac{c_n}{(\alpha + \beta)Y} > (\leq) 0 \quad \text{and} \quad b_p = 0 \quad \Rightarrow \quad h = h(P) \quad \text{and} \quad c = \bar{c}(h(P),0) \]

where  
\[ Y = U_{hh} + U_{hn}c_n < 0 \]

The effects on clinical quality from a higher price are as follows;

(Case SI and Case SF):  
\[ c_p^{SI} = c_p^{SF} = \bar{c}_p h_p = -\frac{c_n}{(\alpha + \beta)Y} > (\leq) 0 \]  
(d6)

From (d1) we have that \( h_p > 0 \) thus the effect from a higher price will depend on the sign of the responsiveness effect (\( \bar{c}_p \)). Given complements (substitutes), a higher price will increase (decrease) clinical quality.

(Case CI):  
\[ c_p^{CI} = \bar{c}_p h_p + \bar{c}_b b_p = -\frac{(c_n)^2}{(\alpha + \beta)Y} + \bar{c}_b b_p > 0 \]  
(d7)

The first term is positive and the second term is positive since \( \bar{c}_b > 0 \) (see 8) and \( b_p > 0 \) (see 10).

(Case CF):  
\[ c_p^{CF} = \bar{c}_p h_p = -\frac{(c_n)^2}{(\alpha + \beta)Y} > 0 \]  
(d8)

**Appendix e: The second-order condition for problem (13):**

The second-order condition, given all third derivatives being zero, is:

\[ W_{pp} = [U_{hh}h_p + U_{hn}\bar{c}_p] h_p + [U_{hh}h_p + U_{hn}\bar{c}_p] \bar{c}_p = U_{hh}(h_p)^3 + U_{hh}(\bar{c}_p)^3 + 2U_{hh}h_p\bar{c}_p < 0 \]  
(e1)
For SI, SF and CF, \( c_P = c_h h_p < 0 \), hence condition (e1) is fulfilled for these three cases. A sufficient condition for (e1) being fulfilled for CI is \( c_P = c_h h_p + c_j b_p < 0 \Rightarrow c_j b_p < -c_h h_p \).

**Appendix f: Solving for the optimal prices**

The solution to the social problem follows from (14). By inserting the first-order conditions (of the clinician and the hospital) into the parenthesis of (14), we arrive at the following expressions for the parenthesis in the four cases:

**Case SI**: \( U_h c_k h_p + U_k c_k b c_h = \frac{1}{\alpha + \beta} \) \( K \) \( - \frac{1}{\alpha + \beta} P^c h c_h = \frac{1}{\alpha + \beta} \) \( K \) \( - \frac{1}{\alpha + \beta} P^c h c_h \) and \( U_i - (1 + \delta) k = \frac{1}{\alpha + \beta} \) \( K \) \( - \frac{1}{\alpha + \beta} P^c h c_h \) \( (f1) \)

**Case SF**: \( U_k c_k h_p + U_k c_k b c_h = \frac{1}{\alpha + \beta} \) \( K \) \( - \frac{1}{\alpha + \beta} P^c h c_h \) \( (f2) \)

**Case CI**: \( U_h c_k h_p + U_k c_k b c_h = \frac{1}{\alpha + \beta} \) \( K \) \( - \frac{1}{\alpha + \beta} P^c h c_h \) \( (f3) \)

**Case CF**: \( U_k c_k h_p + U_k c_k b c_h = \frac{1}{\alpha + \beta} \) \( K \) \( - \frac{1}{\alpha + \beta} P^c h c_h \) \( (f4) \)

The optimal prices follow from inserting the relevant expressions in (d1-d4) and (f1-f4) into (14) and solving for the price.

**Appendix g: Proof for \( \frac{h_p}{c_p} = h_i \) (case CI; see 18)**

We will show that \( \frac{h_p}{c_p} = \frac{h_i}{c_i} \) where \( \frac{h_p}{c_p} = \frac{h_i}{c_i} \)

By inserting (8) into \( h_p \) (as defined in 18), we arrive at the following expression;

\[
\frac{h_p}{c_p} = -\frac{U_{ch}}{(\alpha + \beta) [U_{ch} + U_{ch} c_h] U_{cv}} = \frac{U_{ch}}{(\alpha + \beta) [U_{ch} U_{cv} - (U_{cv})^2]}
\]  

\( (g2) \)

By inserting (g2) into (g1), and rearranging, we get; \( \frac{h_p}{c_p} = \frac{U_{ch}}{U_{ch}} \)