Faculty Research Working Papers Series

Effort as Investment:
Analyzing the Response to Incentives

John N. Friedman
Department of Economics – Harvard University

Steven Kelman
John F. Kennedy School of Government - Harvard University

May 2007

RWP07-024

Effort as Investment:
Analyzing the Response to Incentives*

John N. Friedman[†] Steven Kelman[‡]

April 18, 2007

Abstract

We analyze a model in which incentives in one period on one task can affect output more broadly through learning. If agents can invest in human or organizational capital, then output will increase both before and after short-term incentives. We develop a model of these effects, and then we evaluate its predictions using data from hospitals in Britain during a series of limited-time performance incentives offered by the government. We find empirically that these policies increase performance not only during the incentivized periods but also before and after, matching the preditctions of our model. We also examine performance along non-incentivized dimensions of quality of care and find little evidence of classical effort substitution.

Keywords: Incentives, public sector incentives, multi-tasking.

JEL Codes: D20, H51, I12, J33.

---

*We would like to thank David Cutler, Edward Glaeser, Larry Katz, Andrei Shleifer, and Richard Zeckhauser for helpful suggestions and valuable discussion, as well as participants in seminars at Harvard University and the Univeristy of California at Berkeley. Friedman acknowledges financial support from the NSF and from NIA Grant T32-AG00186 to the National Bureau of Economic Research. Kelman acknowledges financial support from the The Roy and Lila Ash Institute for Democratic Governance and Innovation at the Kennedy School of Government at Harvard University.

[†] Department of Economics, Harvard University, email: jnfriedm@fas.harvard.edu

[‡] Kennedy School of Government, Harvard University, email: steve_kelman@harvard.edu

1 Introduction

What are the effects of a performance incentive beyond the immediate task and time period that are rewarded? Much of the existing literature on incentive theory focuses on the direct effects of incentives on output, looking at the case where effort today affects only output today, thus limiting the effect of a single period of incentives (for a summary, see Bolton and Dewatripont (2003)). There are many contexts, though, where the broader effects of incentives may be important. For instance, Healy (1985), Oyer (1998), and McNichols (2000), among others, show that, when agents can shift effort or output across periods, output can fall just before or after an incentivized period. Deci (1971) and Gneezy and Rustichini (2000) argue that a post-incentive depression in performance may be permanent, if incentives crowd out intrinsic motivation. In classical effort substitution, as in Holmstrom and Milgrom (1991), agents may redirect effort across tasks.

In other cases, though, incentives can have broad effects that reach beyond the immediate task or time period as a result of learning and investment. Consider, for instance, a basketball coach who, at the beginning of a 30 game season, offers rewards to his players for their free-throw percentages during games 11-20. Players may respond to these incentives by simply trying harder during only those incentivized games, thereby "hustling" more. But many players might also try to improve their free throw shooting skills in advance, by "learning," in which case output will be higher than usual in the first ten games of the season too. If this better technique stays with the players, their free-throw percentage will stay high even after the incentives have passed.[1]

In this example, as in many other contexts, effort serves not simply to increase output in one period but also to affect the relevant stock of capital, especially human or organizational capital. The effect may be positive or negative. For instance, agents might work to develop new procedures or technologies for production in response to incentives, so that the capital stock (as well as output) is higher in future periods. Alternatively, efforts to increase output in the near term could draw down the stock of capital or cause agents to forego good investments that otherwise would have happened, as in Stein (1989). In either of these cases, incentives in even a single period, or on just one task out of many, could affect performance in a very general way across tasks or time periods. This effect is analogous to the idea in public finance that a change in taxes in one year can affect output in other years through a change in human capital accumulation. This paper analyzes the consequences of such a model and provides empirical evidence that incentives can indeed have such broad-ranging effects on performance.

To explore this broader concept of effort, we first develop a simple model. Our framework nests the standard model, but it also allows effort to affect the capital stock in either a positive or a negative way. To match the model most closely to the data we use here, we focus on the case where

---

[1] We thank Richard Zeckhauser for suggesting to us this wonderful example.

the spillovers are across time rather than across tasks. From this setup we then derive a number of testable implications. Most importantly, if effort increases the capital stock, then a single-period performance bonus boosts output not only in the incentivized period but also before and after; if effort draws down the capital stock, incentives may depress output around the incentivized period.

To evaluate our model empirically, we then turn to data from hospitals in the National Health Service (NHS) in England. Prime Minister Tony Blair campaigned for office in 1997 and 2001 in large part on a promise to improve public services. Upon entering office, the Blair government established a set of performance targets for public services, including schools and hospitals. In this paper, we focus on a series of incentivized targets, implemented by the government beginning in 2003, for hospital emergency rooms. The goal was to meet a performance hurdle for an established statistic measuring a single, precisely defined aspect of quality: the fraction of patients treated within four hours in an emergency room. The incentives tied to this goal lasted for only two short periods of time - one for a week and another for 13 months. Furthermore, the incentives were extremely simple, rewarding hospitals simply for performing above a threshold that was uniform across hospitals. The discontinuous nature of these incentives provide an ideal setting in which to test our hypothesis that effort can affect the capital stock; since the incentives began and ended sharply, we can easily separate the broader impact from the direct effect of the incentives.

Our results strongly support the hypothesis that incentives can induce capital improvements. Hospitals increase performance well in advance of the incentivized periods (though not before their announcement), and the performance improvements continue at a high level long after the incentives expire. The magnitude of these effects is economically large. We use discontinuities in the rate of performance improvement at the announcement or ending of incentives to show that these effects are not solely the result of secular changes in available resources or other omitted variables. Thus, the effort-as-investment story provides an explanation for the broad increases in performance, both before and after these short-lived incentives, which would otherwise be a puzzling result.

We can also see a number of other effects from these incentives in the data. As might be expected from a threshold contract, performance increases are most pronounced in hospitals that begin as poor performers, though the differences across hospitals are not statistically significant, and economically significant effects appear in all hospitals covered by the incentives. We also examine performance in a number of non-incentivized aspects of quality of care and find little evidence that performance decreases along these alternative dimensions as a result of the incentives. These results contrast with much of the pay-for-performance literature in the private sector (Prendergast, 1999) or the public sector (Dixit, 2002), and especially results in healthcare contexts (Rosenthal and Frank, 2006).

We first turn to a theoretical exploration of these concepts, with an eye towards generating testable hypotheses, in Section 2. Section 3 then provides some background on the example we use here for empirical testing, giving information on the National Health Service in England and the specific system of incentives implemented there. Section 4 describes the data and presents the main empirical results concerning the time series of output, after which Section 5 investigates the extent of effort substitution and gaming. Section 6 discusses the implications of our results for policy, and Section 7 concludes.

2 Theory

Before proceeding to the empirical evidence, we pause in this section to examine the consequences of interpreting effort as capital investment. We first develop the concept in a simple model, after which we generate empirical hypotheses which we can test in the data.

2.1 The Model

Consider a risk-neutral agent employed by a benevolent government to produce output y_t for $t \in [0,\infty)$.[2] We refer to $t = -1$ as the "pre-period." Each period t, the agent puts forth a base level of effort $\underline{e}$ in return for a fixed wage a. We normalize $\underline{e} = 0$, and we assume that merely providing $\underline{e}$ does not affect the capital stock. In addition, the agent can choose extra effort e_t at additional cost $c(e_t)$, where we assume that $c(\cdot)$ is an increasing and convex function with $c'(0) = 0$.[3] In each period, the agent thus has utility function $u_t = v_t - c(e_t)$ where v_t is monetary compensation in period t. The agent discounts the future with discount factor β and acts to maximize the future discounted sum of expected utility.

The agent produces output by

$$y_t = \gamma e_t + K_t + \varepsilon_t$$

where the capital stock K_t evolves as

$$K_t = \delta K_{t-1} + (1-\delta)\bar{K} + \delta\phi e_{t-1}$$

and $\varepsilon_t \sim F(\varepsilon)$ (i.i.d draws) with $E(\varepsilon_t) = 0$, where $\gamma \geq 0$ and $\beta, \delta \in [0,1]$. Note that this is a special case of the more general specification where $K_t = K_t\left(e_{t-1}, e_{t-2}, \ldots, K_{t-1}, \bar{K}\right)$. Such a framework would allow, for instance, for effort in one period to increase the capital stock more as time passed, perhaps as workers become accustomed to a new procedure, or through "learning-by-doing" (as in Arrow (1962), Schmenner (1993: 462-466), and Argote (1999), for instance). But since these effects

---

[2] We model the government service provider as a unitary actor, whether an individual or an entire organization (as in our empirical analysis). The differences between the operation of incentives in these two cases is an interesting subject for future work.

[3] Since the base level of effort is not actually 0, but has simply been normalized to 0, we could allow for $e_t < 0$. Intuitively, workers might provide less than the minimal level of effort, but they do so at increasing marginal risk of being caught and punished.

would not substantively change the predictions of the model, we focus on this simpler case for the sake of tractability. To ensure the concavity of the problem, we make the technical assumption that f is everywhere non-decreasing.[4]

This production function allows for effort in one period to affect output not just in that period, as in the standard model, but also in future periods, either positively or negatively. This is because effort in one period affects the capital stock in future periods. Note that, in our model, capital cannot be bought or sold but instead must be accumulated through effort by the agent. Thus, the capital stock here is more likely to represent human capital, such as know-how, and organizational capital, such as the management structure or set of procedures (as in Nelson and Winter (1982)), than traditional machines.[5] This setup nests several interesting special cases. For instance, the standard production technology in the incentive theory literature, where $y_t = e_t + \varepsilon_t$, corresponds to the case where $\gamma = 1$ and $\phi = 0$. Effort as a pure capital investment would imply that $\gamma = 0$, $\phi > 0$. Effort that is partially investment might be captured as $\gamma > \phi > 0$. If effort draws down future capital stock, as it might if workers overuse machines or forsake positive net-present-value investments to boost current output, then $\phi < 0 < \gamma$. Note that this effect is distinct from shifting effort across periods, as in Oyer (1998). In this different story, diverted effort or output is stored until actively drawn down, much like a saving account. In our model, capital, once accumulated (or destroyed), decays back to the steady-state level exogenously over time.

The government compensates the agent in two ways. First, the agent receives a flat wage a for showing up to work and putting forth a minimal level of effort $\underline{e} = 0$. Second, the government offers an incentive contract conditional only on output produced in period T. The government can observe and contract upon output, but not effort.

In this paper, we do not solve for the optimal government incentive scheme, but instead consider a two-part incentive scheme in which the government offers a prize of value W paid if and only if output $y_T > \chi$ in period T , for some threshold level χ, as well as a piece rate w for output in period T. We assume, however, that the agent signs this contract in period 0, and thus knows its terms from then forward. The agent has no expectation of this contract at $t = -1$. Thus, in the "pre-period," which represents the non-incentivized steady state, the agent will put forth effort $e_{-1} = 0$, and output is $Ey_{-1} = \bar{K}$. This form of contract nests both the threshold contract actually used in hospitals in England and commonly employed piece rates.

---

[4] In our model, we allow only for one type of effort, which affects both current output and the future capital stock. A more general model would allow for two forms of effort: "hustle" and "learning." The empirical predictions from such a model would be no different from those here, though, so we present only this case for the sake of simplicity.

[5] The reader may note that most of the literature following Rosen (1973) on these non-physical forms of capital assume no depreciation. While the model certainly goes through just the same if we were to simply assume that $\delta = 1$, there may be reasons that capital in this problem does depreciate. For instance, the skills and procedures agents develop may be somewhat situation specific. If the particular problems or issues a public service unit faces change gradually over time, then upgrades designed become less applicable as time passes. Similarly, expertise developed with a given set of equipment or hospital procedures becomes less important over time as those machines or regulations change.

From this model we derive Proposition 1, which characterizes the variation of effort across time.[6]

Proposition 1 *If $\phi > 0$, then effort in periods $t \geq 0$ before the incentive period will continually increase in periods $t < T$. If $\gamma > \beta\delta\phi$, effort will peak in period T, else it peaks in period $T-1$. If the marginal disutility of effort increases linearly, so that $c''' = 0$, then effort increases exponentially up to a peak at $t = T$. If the marginal disutility of effort is convex (concave), then effort increases at slower (faster) rate. If $\phi \leq 0$, then $e > 0$ only in period T.*

The intuition for the first part of Proposition 1 is straightforward; if effort positively affects the future capital stock, so that $\phi > 0$, then the agent is motivated to exert positive effort even before the incentive period. In those earlier periods, however, the incentive to act decreases exponentially with the time to the incentive period, since the payment is further in the future and since the positive impact of effort on the capital stock will have depreciated more by the time $t = T$. When $\phi \leq 0$, effort is set to 0 in all periods $t < T$ since there are no positive spillovers onto production in the incentivized period.[7]

The shape of the $c(\cdot)$ function also has an impact on the rate at which effort increases as the incentive period grows closer. (Suppose for the moment that $\phi > 0$). Intuitively, the more sharply the marginal disutility of effort increases, the less impact the changing incentives have on effort, therefore slowing down the rate of increase. In the extreme, when for some level of effort $\bar{e}$ cost is $c(\bar{e}) = 0$ but $c(\bar{e} + de) = \infty$, the agent exerts $\bar{e}$ in all periods. When the marginal disutility of effort increases linearly, then effort increases proportionally as $\frac{e_t}{e_{t-1}} = \frac{\phi\beta\delta}{\gamma}$ per period, exactly matching the changing impact of e_t on Ey_t. But when the marginal disutility of effort increases faster than linearly, effort decreases by less as one moves away from the incentive period.

Unlike in the time leading up to the incentive period, ϕ has no effect on effort at $t = T$, since ϕ impacts only how the capital stock is changed once the incentives have passed. Period T is also the only time at which $e > 0$ when $\phi \leq 0$. Since current effort decreases (or does not affect) future output, the agent has no incentive to work harder than the minimum.

We can also characterize the time series of output, as in Proposition 2.

Proposition 2 *When $\phi > 0$, output grows above $y_{-1} = \bar{K}$ until the incentive period. If $\gamma > \phi$ or $\gamma = 0$, expected output is highest in period T, else it peaks in period $T+1$. After $t = T$ output decays at rate δ, during which time $Ey_t > 0$. When $c''' = 0$ and $T \to \infty$ (so that the incentives are announced infinitely far in advance), then output grows at precisely rate $\beta\delta$ above $\bar{K}$ for $t < T$.*

---

[6] The proof for Proposition 1, as well as all other proofs, appear in the Appendix.

[7] If we allow for $e_t < 0$, the opposite pattern is true when $\phi < 0$; agents will try to shirk excessively in order to aritifically inflate the capital stock above the steady state in advance of the incentive period.

When $\phi \leq 0$, then $Ey_t = 0$ for $t < T$, $Ey_T > 0$, after which output falls to $Ey_t \leq 0$ and converges at rate δ back up to $y_{-1} = \bar{K}$.

Figure 1 depicts the time series of expected output, as characterized by Proposition 2, for several different values of ϕ. (We assume for these figures that $\gamma > \phi$). When $\phi > 0$, since the agent exerts positive effort in the time before the incentive periods, output is positive. As the effort increases towards the peak effort in the incentive period, so does the output. Furthermore, after the incentive period has passed, output only gradually decays back to the no-effort steady state of $Ey = \bar{K}$. When $\phi \leq 0$, however, this pattern will not obtain. In this case, output remains at $\bar{K}$ until the incentive period, at which point it spikes sharply upwards. In the standard model, when $\phi = 0$, output will return to $\bar{K}$ immediately. But if effort destroys some of the capital stock, then output will fall below $\bar{K}$ and gradually converge back up to $\bar{K}$.

We can also derive comparative statics that speak to the cross-sectional distribution of performance.

Proposition 3 *An increase in the steady state capital level $\bar{K}$ (or a decrease in the output threshold χ) will increase Ey_t for all t, but by less than then increase in $\bar{K}$, so that $1 > \frac{\partial Ey_t}{\partial \bar{K}} > 0$.*

To ensure concavity of the problem, we assumed that f is everywhere increasing. This implies that moving above the threshold has decreasing marginal returns to the probability that the agent clears the hurdle. Therefore as one increases $\bar{K}$, the effective monetary returns to increasing output fall, along with effort. Of course, output is still higher than before; but the gain above $\bar{K}$ falls.

To further capture the realistic differences between agents with heterogeneity in steady state capital levels, we offer the following corollary to Proposition 3.

Corollary 1 *Suppose that $\phi > 0$. Suppose further that there is some fixed cost ψ of exerting positive effort in each period. Then there exists some τ such that $e_t = 0$ if and only if $t < \tau$. Furthermore, τ is increasing in $\bar{K}$.*

To see the intuition behind this proposition, first note that the fixed cost of exerting positive effort does not alter the marginal incentives once the agent begins to act. Therefore, the agent either acts as she would without the cost, or she chooses not to act at all. Effort (in the simpler model) increases monotonically throughout the periods leading up to the incentive contract, as does the marginal impact of effort on performance in period T. Therefore, the total returns to exerting positive effort increase as the incentive period grows closer; so long as the fixed cost is not too high, at some point the agent finds it worthwhile to begin exerting effort.

We now formulate the results from our model into testable predictions that we can take to the data.

2.2 Generating Testable Hypotheses

Propositions 1 through 3 yield clear empirical predictions for agents facing short-term incentives. The first two concern the time-series of output surrounding the incentive period, and reflect the sign and magnitude of the effort spillover parameter ϕ.

Hypothesis 1: If $\phi > 0$, then performance will increase in advance of the incentive period. After peaking in the incentive period, it will remain above the steady state as it decays back down.

Hypothesis 2: If $\phi \leq 0$, then performance will not increase before the incentive period. After spiking upwards while the incentives are active, performance will drop sharply, and if $\phi < 0$ then output will fall below the initial level, recovering only slowly to the steady state.

These predictions are a direct result of Proposition 2. In contrast to the standard model, in which the agent puts forth effort only in the incentivized period, the agent may exert positive effort in all periods up to and including the incentivized period. Thus, if effort positively affects the capital stock, output should gradually climb to a peak before the incentive period, rather than spiking up in only that period, as in the standard model. Furthermore, the effort put forth before and during the incentive period has lasting positive effects that remain even after the incentive period has passed.

Of course, all this will occur only if effort has positive spillovers into output in future periods. If this is not the case, then performance will not increase in advance of the incentive period. The agent will obviously wish to work during the incentive period, during which output will become positive, but output will drop back to the steady state (or even below it) once the incentives have passed.

The empirical hypotheses thus far have been based on the time series of output for a representative agent. We can look across agents as well. Proposition 3, along with its Corollary, yields a prediction about the cross-sectional differences in agents' response to the incentives.

Hypothesis 3: Better performing agents (ex ante) will demonstrate less improvement but will maintain a higher level of performance through all periods. Furthermore, such agents will wait until closer to the beginning of the incentive period to begin exerting abnormal effort.

Since the government compensates agents using a threshold performance contract, the incentives to exert effort are not uniform across the agents. For agents who are better performers ex ante, and thus have higher steady-state levels of capital, it would take a very bad shock to perform below the threshold. Since the density of shocks is lower for more negative values, the incentive to perform may be quite small. Furthermore, in the presence of fixed costs for positive effort, these high-performing agents will wait until closer before the incentive period - perhaps even until the incentive period itself - to exert effort.

2.2.1 Other Hypotheses

The first two hypotheses each relate primarily to measuring the sign and magnitude of the effort spillover parameter ϕ. But if current effort does increase future output, we can also use the results in Proposition 2 to recover the parameter β from the data.

Hypothesis 4: The difference between the rate of output increase up to the incentive period and the rate of decay away from the incentive period will increase with the discount rate β. Both the rate of increase and decrease will increase with the depreciation rate δ. If the depreciation rate of these investments is zero, as in classical human capital theories, then the rate of performance improvement leading up to the incentive period is greater for lower discount rates, while performance is flat after the incentive period.

Output grows in the periods leading up to the threshold contract since effort is increasing. Effort, in turn, increases for two reasons: First, the as the incentive period grows closer, current effort has a mechanically greater impact on y_T since the capital stock will have depreciated less. Effort also increases because the potential payment is closer in time, and thus less discounted. Effort, and therefore output, thus grow at roughly rate $\frac{1}{\beta\delta}$ approaching the incentive period. Once the incentive period passes, though, only one of these forces - depreciation - works in the opposite direction to decrease output. If we observe that performance increases up to and decreases away from the incentive period at roughly the same rates, then we know that discounting is low. If, on the other extreme, we observe a rapid increase before the incentive period but little or no decline afterwards, then we know that depreciation is low and discounting is high.

Though lying outside the particular model which we specify above, we offer a few more predictions that would follow from a slightly more general model. Suppose that agents can choose between multiple modes of effort, some of which primarily increase the capital stock and others of which mostly increase current output (perhaps even at the cost of decreasing the capital stock). Intuitively, the agent can choose between effort spent "investing" and effort spent "hustling." Furthermore, suppose that the threshold incentive need not be focused only on a single period, but may be spread over a number of consecutive periods (while keeping the sum of prizes the same, in present value terms). This slightly richer model yields a further prediction.

Hypothesis 5: When the incentive periods are prolonged, the drop-off in effort after the incentive period ends will be smaller.

When an agent has a choice between modes of effort, she tends towards more lasting forms of effort when the incentives are stretched over more periods. When the incentives are more concentrated, then the agent would rather increase current output without reference to future periods. But if the agent chooses to "hustle" to meet the incentives rather than "invest," then performance will drop off more after the incentives conclude. Thus, we should expect a larger falloff in output after shorter incentive periods.

In our model, effort and capital enter into the production function in an additively separable way, so that the level of capital does not affect the marginal product of effort; it may be that, in many settings, effort and capital are complements. For instance workers may be better able to treat patients if a hospital is better organized and its procedures more efficient. In this case, our model yields an additional hypothesis.

Hypothesis 6: If effort and capital are complements in the production function, the post-incentive decline in output is smaller as worker's other incentives, whether intrinsic or external, increase.

The intuition here is most easily seen in the extreme case when the production function is just effort times the capital stock. Once the incentive period has passed, the workers have a much larger capital stock than in the "pre-period" steady state. But because of the production function, the degree to which these improved procedures, say, increase output depends on the worker's level of effort. If the agents care strongly about production, for instance because they intrinsically value the output (as doctors and nurses might in a hospital, for instance), then agents will work a great deal to take advantage of the greater marginal product of effort. Output remains quite high even after the incentives have passed. If, on the other hand, agents do not care at all about output, then they will put forth zero effort; despite the larger capital stock, output returns immediately to zero. Though this dichotomy is not as stark when the production function takes a more general form, the effect applies so long as effort and capital are complements. Thus, the extent of the post-incentive falloff in production depends on the agent's other incentives for output.

2.3 Alternative Explanations

There are a number of potential confounds in testing each of these hypotheses. One is that other models of production may be consistent with the hypothesized patterns of output, even if effort in one period did not affect output in future period. For instance, the agent may experiment with ever more efficient methods of temporarily increasing effort leading up to the incentive period. In this model of learning, though, output would revert to the steady state immediately after the incentive period passes, in contrast to the slower decay as predicted by our model. This would be the case with any model of convex costly upward adjustment for effort. But also to generate output that remains above steady-state after the incentives have passed required convex costly *downward* adjustment costs as well, a peculiar assumption.

Another serious concern is that other inputs into production - perhaps the resources or technology available - increase throughout the relevant period. This would lead to a pattern of uniformly increasing effort, which could be confused with the impact of investment before the incentive period and the lingering effect of the capital stock after the threshold contract has passed. Ideally, one could control for the production process with similar agents not affected by the incentive contracts; alas, our empirical design does not allow for this fix.

In the absence of a control group, we instead rely on the discontinuous nature of the incentives provided to rule out secular increases in production. The technology and resources available in period $T+1$, immediately after the end of the incentive contract, should not be much different from those in use during the incentive period. The monetary returns to effort change drastically between one period and the next, however; we should be able to attribute any large jump in the time series of output to the incentive effects. Furthermore, we should see an abrupt shift in the slope of improvement as the incentives period passes. For this reason, Proposition 5 tells us that shorter incentive periods will be less prone to the confounding effects of increasing resources. The announcement of the incentives also provides a discontinuity we can analyze. The rate of increase of output should increase when the incentives are announced, especially if the incentivized period is not too far into the future. Finally, if $\phi > 0$, the rate of increase in output before the incentive period should still be greater than that after the contract has passed, even in the presence of secularly increasing production.

Another potential confound stems from the multi-task nature of hospital performance. Since the government implemented incentives on one of many dimensions of hospital performance, it may be that these organizations simply moved along a “production possibilities frontier” away from other tasks towards decreasing emergency room wait times. If hospitals undergoing such shifts in focus incur convex adjustment costs, then the change in focus would begin before the actual incentive period and linger long after. To investigate this possibility, Section 5 investigates contemporaneous changes in other measures of hospital output.

Another alternative hypothesis is that agents increase effort not in response to the explicit incentives but as a result of other factors related to the incentives. For example, perhaps agents are uncertain about the true preferences of the government, and those in power may have difficulty conveying these preferences without costly signaling. The incentives may then serve as this costly signal of the high value that the government places on performance in emergency rooms, or hospitals, or public services. The agents, now aware of the government’s preferences, might then increase effort in response to the expectation of some future government action, as might be the case if agents thought the government would invest in well-performing hospitals. Or, agents might derive intrinsic value from high quality healthcare and interpret the incentive programs as the government’s statement that A&E departments could use improvement; agents might also wish to respect the commitment to follow direction implicit in an employment relationship (as in Barnard, 1938; Williamson, 1975) Any of these stories would predict a broad increase in performance that would begin before and continue beyond the end of the formal incentives, as in Hypothesis 1.

The essential difference between the empirical predictions of these stories and those of our theory is that, under these alternative theories, agents are primarily responding to a general desire to increase output and not the particular threshold or goal. Thus, agents' responses should not be sensitive to the specifics of the incentive scheme. For instance, the specific timing of the incentives should not matter, since agents care no less about improving output in the period after incentives than in the period before. To the extent that either the level of output or the rate of improvement of performance varies with timing of incentives, this alternative theory cannot alone account for the data. The exact thresholds laid out in the incentives also should not matter; this theory thus makes the opposite prediction from Hypothesis 3: initially worse firms or poorer hospitals (for whom the incentive payments are a larger fraction of the budget) need not respond more to the incentives than those firms already well above the threshold.

3 Background: Hospital Reform and Incentives in England

England's National Health Service was established in 1948 as the nation's primary healthcare system.[8] The system includes 155 local "hospital trusts" (in our period), each of which manages the local hospital(s) and associated care centers with funding almost entirely from the national government. Reform of the NHS, and more generally of public service provision, was a key element of Tony Blair's election platform in 1997. In contrast to the movement towards privatization under Margaret Thatcher and John Major, Blair proposed implementing government-managed measurement of performance and standards, or "targets" (Kelman, 2006).

In 2000 the Department of Health formulated Blair's plan for reform in "The NHS Plan" (Department of Health, 2000). The report proposed a 33% real increase in funding for the NHS over the next five years. The report also criticized the NHS as a "1940s era system operating in a 21st century world," operating with a "lack of national standards ... and clear incentives and levers to improve performance." Patients complained most vocally about the length of waiting times, both in "Accident & Emergency" facilities (A&Es) and in surgical wards.[9]

In response to these management problems, the Report proposed a system of measurement and comparative rating across hospitals that included a number of different areas of hospital responsibilities. This system came to be known as the "star rating" system, since hospitals receive overall

---

[8] Regional governments separately manage the NHS in other areas of Great Britain: Scotland, Wales, and Northern Ireland; and neither the performance targets described here nor the accompanying incentive regimes we analyae existed outside of England. Therefore, our data cover only England.

[9] Accident & Emergencies are equivalent to emergency rooms in the United States.

ratings between zero and three stars (as well as less publicized ratings on performance on individual measures). The star ratings were one of a number of "league tables" that measured comparative performance of public organizations; others were set up for schools and local governments. The first such ratings for hospitals were prepared and publicized in 2001, and required by law in 2002.

In A&E departments, the 2000 NHS Report targeted the number of patients waiting more than four hours from arrival to discharge or admission (to impatient wards), setting an interim target of 90% compliance by March 2003 and an eventual goal of 100%. At the beginning of January 2003, the government announced that A&E waiting times relative to the interim target would be included, for the first time, in the star ratings. Hospital performance would be measured for this purpose in the final week of March 2003.[10] In some sense, this week functioned as a "sweeps week" for the hospitals.[11]

It is useful at this point to discuss exactly how the star ratings may have incentivized the hospitals. There was no direct monetary payment associated with a high rating. Furthermore, when the ratings were proposed, patients had little choice of hospitals, so a high star rating could have only a small impact on the number of patients presenting, and thus the income a hospital received. The 2000 NHS Plan stated that high scoring hospitals would be given automatic access to two contemplated sources of incentive funding, but this never occurred.

The star rating did have some bite, though. According to the 2000 NHS Plan, persistently low scoring hospitals would be subject to replacement of their leadership or takeover by another hospital. Between 2001 and 2003 the government replaced the leadership of six hospital trusts due to low star ratings, and in 2006 one low-performing hospital was taken over by a more successful one (Department of Health, 2005a: 65; BBC, 2006). In addition, low-performing trusts were subject to special measures short of leadership replacement, such as heightened monitoring from NHS headquarters. Also, in 2003 the government began promoting high performing hospital trusts (judged largely on the basis of the star ratings) to the level of "Foundation Trust," after which hospital leadership received a greater level of independence from national oversight. Finally, the threat of humiliation from low performance ("name and shame," in UK parlance) - either in the public eye or within the medical profession - probably contributed to the incentive effects from these ratings.

Nine months after the "sweeps week," the Prime Minister's Delivery Unit, an organization Blair created to work on government performance targets, released the "5-Point Plan" for meeting the

---

[10] Other measured areas in the star ratings included the waiting time for elective surgery, the death rate following major operations, survey feedback from patients and doctors, as well as a number of softer criteria such as a consultant appraisal and the quality of hospital food.

[11] Four times per year, a week of television programming is designated a "sweeps week." Program ratings from viewer diaries recorded during this week are then used to set advertising rates until the next measurement week. It is widely acknowledged that networks attempt to schedule the best, or most outrageous, or most eye-catching episodes of a given series during these weeks.

A&E targets. In particular, this document announced monetary incentives for hospitals that met A&E targets. In each of the next five fiscal quarters, hospitals would receive a lump-sum grant of £100,000 if the percent of patients treated within four hours across the entire quarter rose above a threshold. The thresholds began at 94% for March 2004.[12] The targets increased by a single percentage point each quarter thereafter until reaching 98% for January through March 2005.[13] The report explicitly stated that the incentive payments would end after March 2005. Officially, the funds could only be used for capital expenditures, not operating budgets; to the extent that hospital trusts could not reallocate money in response, this may have reduced the value of these payments. We refer to this second phase of incentives as the "cash" period.

Table 1 presents statistics on the number of hospitals clearing each of the thresholds between March 2004 and March 2005. At no point did more than 52.6% of trusts manage to clear a hurdle. As the threshold levels increased, fewer hospitals achieved the milestones. Also, clearing one hurdle by no means implies that a hospital cleared all of the earlier, lower hurdles. Fewer than 10% of trusts managed to clear all the hurdles; for comparison, if performance increases were perfectly correlated across hospitals, then all 27.9% that made the final threshold should have cleared all five hurdles. If performance within one quarter were entirely independent of that in other quarters, then only 1.1% of hospitals would have won all bonuses. Similarly, only 20.8% of hospitals cleared no hurdles.

In all, between 2003 and 2005, A&E departments faced two temporary incentives: the "sweeps week" in March 2003, followed by the "cash period" in 2004 and early 2005. These episodes provide a excellent experiment to test the hypotheses generated from our model. Anecdotally, the reforms hospitals implemented (and also those on which the government focused) looked to improve A&E procedures. Starting around the announcement of the Star Ratings incentive and continuing through the rest of 2003 and 2004, various central government agencies publicized and organized training around various organizational process changes they recommended as ways for A&E departments to improve their waiting time performance. One particularly prominent reform around this time was the "See and Treat" reform, which suggested redesigns to hospitals' triage procedures to more immediately treat minor injuries. In previous systems, a low-priority patient (for instance, someone a minor wound requiring stitches) might wait for many hours until all more serious patients were treated; under the reform, the hospitals tried to assign nurse-practitioners to deal with such injuries more quickly. Other A&E managers speak of procedural changes to better coordinate different stages of treatment. For example, some nurses tried to bring test results back to doctors more quickly after the tests were completed, while others made sure that patients were waiting for a treatment room while treatment rooms were available. Of course, these stories mean little unless

---

[12] Since the plan was only announced in January 2004, the first incentive payment covered only March 2004 rather than the entire fiscal quarter.

[13] The 5-Point Plan changed the final target from 100% to 98% of patients handled within four hours, based on consultations with doctors about justifiable exceptions to the four-hour treatment standard.

we can verify the hypotheses developed above in the data, a task to which we now turn.

4 Analyzing the Time Series of Performance

4.1 Data and Summary Statistics

Our data come from the British Department of Health. The primary variable of interest is the percent of patients treated within four hours of arrival in each A&E department, recorded weekly in each of 155 trusts across England. Our data begin in January 2003 and run through the beginning of September 2006.[14] We refer to this variable as "performance" throughout the rest of this paper.

A change in the patient allocation procedures means that there are actually two such series: those for "Type 1" patients, and those for all patients. Patients who are not "Type 1" refer to patients seen in newly established alternative treatment facilities for low-grade ailments known as "walk-in centers." Beginning in October 2003, the Department of Health included these patients in the local hospital trust's attendance figures. In practice, the walk-in centers almost never make a patient wait more than four hours, and so the effect is to increase the denominator of the ratio which forms the dependent variable. The correlation between the two series is greater than 0.999, and so the choice of series does not substantively affect the results below. We thus use the "Type 1" patient series to make comparison across different periods easier.

Figure 2 plots mean performance within each week between April 2002 and September 2006. Though the table is annotated, it is not difficult to discern, at a glance, the timing of the two performance incentives. Compliance rates spike up by more than 10 percentage points to 93.0% over the few months before the "sweeps week" in March 2003. Though performance falls by nearly 5 percentage points in the following week, it remains nearly level over the next nine months at a level far above pre-sweeps performance. This spike during sweeps week matches some anecdotal evidence that hospitals used a number of explicitly short-run measures, such as cancelling vacations and putting in overtime, to augment their efforts during the one week of incentives. After the government announces the cash incentives in January 2004, performance once again begins to increase, though now the increase is more steady across time than around the "sweeps" week. Performance climbs from an average clearance rate of 90.9% to 97.9% across the "cash period." Once the incentives end, compliance falls a bit but remains relatively high, hovering between 97.2% and 99.0% throughout the seventeen remaining months of data. It is worth noting the sheer size of these increases ; average performance increases by more than 15 percentage points over the 45 months of this sample, more than 1.5 times the standard deviation of performance across hospitals in the pre-incentive period. Put another way, the 10th percentile of performance in the final period lies above the 90th percentile

---

[14] We have data from some hospitals back to April 2002, but we only have complete data beginning in 2003. Our data extend through the week ending Sunday, September 10, 2006.

of performance in the first period.

Table 2 provides summary statistics for performance, as well as the number of total patients seen (attendance) and the number waiting more than four hours (breaches), within each week over the entire sample. Type 1 patients are treated somewhat less efficiently in the full sample both because the measure tends to be lower than the "All Patients" statistic in a given period and because the observations for the broader measure come disproportionately from later years of the sample. We provide standard deviations not only within the pooled data but also conditional on performance predicted from a trust-specific cubic spline; this "within trust" estimate of the standard deviation of output provides a better sense of the inherent noise in the performance generating process from a given trust. Breaches represent patients not seen within four hours and so make up only a small fraction of total attendance.

Panel B of Table 2 breaks out the performance statistics into the five periods in our sample. The "pre-sweeps" period begins in January 2003 and runs up until the sweeps week at the end of March 2003. The "sweeps week" period comprises only the one incentivized week that ended Sunday, March 30, 2003. The "middle" period runs for the next nine months until the start of the "cash" period in early January 2004. The "post-incentive" period begins in April 2005 and runs to September 2006. As suggested by Figure 1, mean performance increases sharply during the sweeps week before dropping to an intermediate level in the middle period. Performance increases once again during the cash incentives, so that the final period has the highest mean rate of compliance. Attendance slowly increases across time (though is higher during sweeps week, even controlling for seasonality). The residual standard deviation shrinks as performance edges higher (and closer the 100% upper-bound) in the later periods.

Figures 3 and 4 show the variation across hospitals. Figure 3 plots five quantiles of the performance distribution over time.[15] The increase in performance leading up to the sweeps week is clearly driven by the bottom half of the distribution; the 90th percentile of the distribution actually falls in this period. The entire distribution shifts up during sweeps week and then down afterwards, though these movements too are driven primarily by the bottom of the distribution. The distribution converges slightly during the nine months before the cash incentive, after which it shifts up and converges more rapidly throughout the second incentive period. The bottom of the distribution drops a bit after the cash incentive period passes. It is important to note that the lines in Figure 3 are not cohorts, but rather display the distribution of hospitals within a given week.

Figure 4 shows the fate of five cohorts of hospitals, as ranked by initial performance. Each line represents the average performance, during each week, of hospitals within a given quintile of initial performance. Since we are missing a fair number of observations for some hospitals between April and December 2002, we omit the first nine months of data in this plot. To remove some of the noise

---

[15] These figures are for Type 1 patients.

in the data, we average over the first month in the sample (now January 2003) to determine initial ranking. Though the lines stay roughly ordered throughout, convergence in this plot is much faster than in Figure 3, as might be expected..

4.2 Results

We now turn to our formal analysis of the time series of hospital performance. We identify our model using the discontinuities in the average slope of performance increase within each period. To do so, we define five sub-periods as above and allow an independent intercept and linear slope in time for performance within each period. (Of course, Sweeps Week has no slope coefficient). The functional form is thus

$$y_{hpt} = \alpha_0 + \beta_0 t + \alpha_2 I_2 + \sum_{p=3}^{5} \left(\alpha_p + \beta_p t\right) I_p + \nu_h + \varepsilon_{hpt} \tag{1}$$

where y_{hpt} denotes performance, t a linear function in time, I_p an indicator function that equals 1 in period p, and ν_h a trust-specific fixed effect for hospital h in week t in period p. To facilitate the interpretation of coefficients, we rearrange the parameters so that our intercept estimates correspond to the size of discontinuous jumps between each period; if $\bar{t}_p$ is the final week in period p, then we estimate the parameter

$$d_p = \alpha_{p+1} - \alpha_p - \beta_p \left(\bar{t}_p - \bar{t}_{p-1}\right)$$

as the discontinuity d_p occurring between period p and $p+1$. (The parameter d_3 measures the increase relative to the trend of performance in the pre-sweeps period). Similarly, for periods 3 through 5 we estimate the change in the slope of performance improvement, which is β_3 in period 3 but is $\beta_p - \beta_{p-1}$ for periods $p = 4, 5$. To control for potential autocorrelation in performance within trusts, we cluster by trust.

Table 3 presents these regressions. The dependent variable in columns (1) through (3) is the percent of Type 1 patients treated within four hours of arrival, scaled such that the variable ranges from 0 to 100. Column (1) presents the basic specification. Performance increases by an incredible 7.2 percentage points, eliminating more than 40% of breaches, in the sweeps week, after which it falls to 1.91 percentage points above the ending pre-sweeps level. Performance trends change significantly between all periods; the slope of the performance increases becomes markedly flatter just after the Sweeps week, kicks back up during the cash incentive period, and then shifts down again in the post-incentive period. Especially following sweeps week, the trend break is economically large; had the pre-sweeps trend continued, performance would be 6.5 percentage points higher just one quarter into the middle period.

The data overwhelmingly support Hypothesis 1, which states that performance should increase prior to and only decay slowly after the incentivized periods when effort positively impacts the stock of capital. The data soundly reject the hypothesis that effort has a negative impact on the capital stock. Thus, a key result is that, even if agents seek only to improve performance during a single incentive period, the investment element of their efforts can increase performance in other periods. As predicted by our model, performance begins to increase well in advance of the actual incentive period, once the government announces the incentives. In 2003 since sweeps week was known before our data begin, we see strong improvement in pre-sweeps period. In 2004 the announcement effect is particularly prominent in the sharp change in slope in January, when the government suddenly reveals the cash incentives. The rate of performance increase also falls sharply after each incentive period, though the level of performance remains high.

We now perform a number of robustness check on the results in column 1 of Table 1. Column (2) replaces the background linear function in time with a quintic and so we estimate the equation

$$y_{hpt} = \alpha_0 + f(t) + \alpha_2 + \sum_{p=3}^{5} \left(\alpha_p + \beta_p t\right) I_p + \nu_h + \varepsilon_{hpt} \tag{2}$$

where $f(t)$ is a five-degree polynomial. This helps to better control for England-wide shifts over time, such as the gradual increase in funding or staff available. The changes in slope between each of the periods, as well as the discontinuities, are estimated as above. With this more flexible control for overall effects on hospital performance, the estimates here are substantively similar to those in column (1). Performance increases significantly in the sweeps week, and also in the week after, relative to the pre-sweeps level. The gradual increase in performance decreases by 0.45 percentage points per week after sweep week, and increases now by somewhat more than before entering into the cash incentive period. This specification estimates a positive discontinuity following the end of the cash incentive period, which would not be consistent with the model.

Column (3) adds month-specific fixed effects to the quintic specification is column (2). Treatment rates are somewhat lower in the winter, due to the increase in disease and accidents. Performance is especially low in the weeks following New Year's Day, since regular GP offices are closed for the inter-holiday period and the hospitals must work extra until the backlog is cleared. Performance also tends to dip slightly in February and August when a new crop of hospital residents phase into rotations. By including month fixed effects, we go some way toward eliminating these potentially confounding sources of variation, especially since this analysis is identified entirely from the time series variation. The estimates here are almost identical to those in column (2).

One potential problem with our specification in equation (1) and (2) is that we use a linear equation to model a dependent variable that is bounded above by 100. What is more, we might not want to equate a two percentage point increase from 96% to 98% with that from 76% to 78%.

To capture these effects, we rescale performance as

$$\hat{y}_t = -\ln(1 - y_t).$$

Like actual performance, $\hat{y}_t > 0$, and $\hat{y}_t = 0$ when $y_t = 0$. This new dependent variable equally values *proportional decreases* in the breach rate; improving from 96% to 98% is now equivalent to improving from 76% to 88%.

The results from this specification appear in Table 3 in columns (4) through (6). The estimated rate of performance increase during the cash incentive period, at 1.95, is higher than before relative to the rate of increase in the pre-sweeps period (3.19). Since the new performance measure values small increases at high levels more than before, this change is not surprising. However, the coefficients are substantively very similar to those in columns (1) through (3). As in columns (2) and (3), the specification in column (4) finds a positive jump at the end of the incentive period.

Columns (5) and (6) add a quintic control in time and month effects, mirroring the specifications in columns (2) and (3). Very little changes in column (5) from column (4), though the slope no longer decreases significantly following the cash incentive period. When we include month effects, though, the entire increase between the "middle" period and the "cash" period is picked up by the "Cash Period" jump; this estimate is quite discordant with the coefficients in other specifications, resulting from the particular combination of the month dummies, the log specification, and the quintic function in time. (Without each of these three, the odd estimates disappear). Thus we place little weight on this particular specification relative to the many others which are consistent with each other.

All of the formal analysis thus far has focused on the time series of output from all hospitals; Hypothesis 3, which states that hospitals further below the threshold will put forth more effort, and thus increase output more, in the lead-up to the incentive periods, deals instead with differences across hospitals. To explore this hypothesis, we run regressions similar to that in equation (2) but also allowing for an interaction between the period-specific slope coefficients and a hospital's level of performance. In essence, we are testing to see if the abrupt changes in the rate of performance for well performing hospitals are smaller than that for hospitals who must improve to reach the milestones. Such a finding - a negative coefficient on the interaction between trend growth and lagged performance in non-incentive periods - would confirm Hypothesis 4 that convergence between the hospitals is faster during incentive periods (or when incentives loom in the future) than during post-incentive periods.

Table 4 displays results that allow for convergence across hospitals. The first two columns use simple performance as the dependent variable and include interactions between the slope coefficients and lagged performance, labelled as "convergence" variables. (Lagged performance itself is also included as a control, as are interactions between this variable and the jumps between periods). The "Pre-Sweeps Convergence" coefficient of -0.709 implies that hospitals are converging in performance in the pre-sweeps period, since those hospitals with higher lagged performance are improving less. Similarly, the positive post-sweeps convergence coefficient of 0.476 implies that convergence slows when incentives are no longer present, as in Hypothesis 3. Though the signs of the Δ-convergence coefficients are largely consistent with Hypothesis 3, none is statistically significant in either column (1) or (2). Columns (3) and (4) investigate the same hypothesis using the log-adjusted measure of performance. This transformation of the dependent variable values increases in performance more when a hospital is closer to the upper limit of 100%; as one might expect with this new measure, the regressions indicate much less convergence, and even divergence. None of these coefficients are statistically different from 0 either, though. Hypothesis 3 remains largely unconfirmed.

These data also speak, in a more limited way, to our other hypotheses. Hypothesis 4 posits that an increase in the depreciation rate of capital makes performance fall off faster as one moves away from the incentivized period (either before or after), while the discount rate affects only the rate of effort buildup beforehand. Since we cannot measure discount or depreciation rates independently across trusts, we cannot test this hypothesis directly. But we can use its theoretical insight to learn about the relevant parameters in this example. In these data, performance increases quite rapidly up to the incentive period, but does not decay rapidly afterwards. Though the estimated rate of change in the final period is negative, the magnitude is very small, and the slope in between incentive periods is ever so slightly positive. From one perspective, this is not surprising, since researchers have not traditionally emphasized depreciation when analyzing the non-physical forms of capital we model here. Since capital improvements seem so long-lasting, trusts then wait until quite near to the sweeps week to begin boosting performance, perhaps leaving high-return investments on the table. This phenomenon could be the result of a high discount rate or of high fixed costs to exerting positive effort.

The data are also consistent with Hypothesis 5, which predicts that the drop in output after a prolonged incentive period will be less pronounced than that after a short one. We only have two incentive periods, and thus we cannot make serious statistical inference about the validity of this hypothesis, but the drop in output after the sweeps week is significantly greater than the drop following the cash period in every specification. One potential alternative explanation for this finding is that the government continued to closely monitor hospitals performing poorly following the cash incentive, while they did not do so as intensely following the sweeps week. There is some anecdotal evidence that this was the case. In any case, this hypothesis awaits more thorough testing.

We can do even less to test Hypothesis 6 within our data, as the level of intrinsic and basic compensation is the same throughout our example. Workers in the A&E departments in these hospitals are likely to have a high degree of intrinsic motivation, though; thus, our finding that output does not fall much following the incentive periods is consistent with this prediction of our model. We hope that future work will evaluate this model across a greater range of settings to substantiate or qualify the suggestive evidence here.

5 Assessing Classical Effort Substitution and Gaming

The regressions in Table 3 formally confirm what Figure 2 depict: both the Sweeps Week and the cash incentives led to lasting gains in performance. One potentially powerful alternative hypothesis, though, is that hospitals simply reallocated production towards dimensions rewarded by the incentives. As discussed in Section 2.3, hospitals facing convex costly adjustment along its production possibilities frontier would produce a pattern of decreasing waiting times similar to that seen in the results above, but the response to the incentives would not be “investment” in any real sense but instead non-reversed effort substitution (akin to the result of policy persistence in Coate and Morris (1999)). Though the inertial nature of the performance increases is interesting theoretically, the lessons for public policy would be very different if these gains resulted from such institutionalized effort substitution rather than true improvements in the quality of care. And since overall performance may be difficult to quantify appropriately in this setting, agents have ample room to "game" the system (see, for instance, Dixit 1996). This section investigates these possibilities by examining changes in other output measures that were not incentivized.[16]

5.1 Data and Summary Statistics

In addition to our primary series, we also have a number of alternative measures of quality of care, albeit collected only every quarter, extending from the fourth quarter of 2002 through 2005. Our non-incentivized measure of quality is the distribution of patient waiting times within one-hour bins up to four hours. Most obviously, the four-hour cutoff gives A&Es an incentive to reduce the longest waits by increasing the shortest waits. Since wait time in an emergency room is often determined by the severity of a patient’s ailment, this redistribution would entail an increase in wait times for more serious patients in exchange for a decrease for less serious patients. Unless the marginal cost to patients of waiting is rapidly increasing, such a shift, all else equal, would lower the quality of care, especially when serious patients have life-threatening conditions requiring rapid attention.

We also have data tracking the number of patients that return to the A&E department within a short time after a related first visit. This statistic should give us another crude measure of quality, since most of the return visits are in response to ineffective treatments.

---

[16] We present additional empirical tests, and discuss the empirical results in the context of theories of effort subsitution and "gaming" in response to performance measures, in Friedman and Kelman (2007).

Another more institution-specific method for artificially boosting the measured statistic is to admit extra patients to inpatient wards. When patients requiring serious treatment come through the A&E, they clearly cannot be fully treated within four hours, so the target specifies that they must be admitted to inpatient wards (where further treatment will take place) within four hours. Therefore, faced with the prospect that a patient might break the four-hour target, hospitals have an incentive to admit that patient to the hospital rather than treat them in the A&E department; such an unnecessary diversion is costly because the standard of care in inpatient units is higher, and because free beds in the inpatient wards are often in short supply. Our data record the number of patients admitted to the hospital in each quarter.[17]

Outright falsification is the most blatant channel available to trusts to avoid real reform. This possibility concerned the officials administering these incentives as well. To combat this possibility, the Department of Health conducted regular audits on the record-keeping in each trust. Though the audits reveal a non-trivial error rate in the paperwork of 11%, mistakes were overwhelmingly of a administrative nature. The audits revealed a few instances of errors in time records and no evidence of systematic administrative fraud (Department of Health, 2005b).[18]

We first look at the distribution of waiting times within hospitals by quarter. The data sort patient visits into hour-sized buckets up to four hours, as well as labelling breaches. If hospitals are "gaming" the system in this way, then the fraction of patients in both short-wait and long-wait bins should decrease over time. Figure 5A displays the distribution of patients across these buckets in four distinct periods, each roughly one year apart. These periods occur before the sweeps week, after the sweeps week, during the cash incentive, and after the cash incentive, respectively. The procession of four bars within each cluster represents the fraction of patients within in given one-hour bin during each of four time periods. Though the fraction of breaches shrink over time, the fraction of patients in each lower bin grows, suggesting that hospitals are actually becoming more efficient.

Figure 5B represents these data in a slightly different way. The cumulative distribution of patient waiting times is displayed vertically for any given quarter; as the height of each colored bar changed, we can follow the cumulative distribution function evolving over time. This chart shows that the distribution of patient waiting times in later periods nearly always first-order stochastically dominates the distribution in any earlier period. This distribution shifts slightly up in the last quarter plotted, mirroring the slight increase in breaches during this period shown in Figure 2. But overall, hospitals do not appear to be sacrificing shorter stays to reach the milestone. We further

---

[17] During the period covered by our data, many hospitals also introduced "clinical observation wards," which were inpatient facilities with lower standards than those in traditional hospital room (but higher than A&E departments), for A&E patients requiring tests that would take more than four hours. Patients admitted into these wards within four hours were considered to have met the four-hour target. However, such patients were coded as having been admitted to standard inpatient wards; thus, the extent that hospitals used these new wards to meet the standards is reflected in the inpatient admission figures.

[18] In Friedman and Kelman (2007b), we present other empirical tests of effort substitution and gaming, along with a larger discussion of these issues.

investigate this hypothesis in the regressions below using the fraction of patients waiting fewer than two hours as a representative statistic for the size of the left tail of the wait time distribution.

Though the data only come in hour-sized bins, we can also calculate an approximation to the mean waiting time for patients within a given trust and quarter. To do so, we assume that patients in the 0 to 1 hour bucket, on average, waited 30 minutes, and so on upwards. Figure 6 displays the evolution of this estimated mean waiting time using three different assumptions about the average waiting time for those patients not treated within four hours. Even in the most conservative specification, in which we assume that such patients were treated in an average of only four hours, "mean" wait times improve by nearly 40 minutes, or more than 25%, over the sample period. The changes are even starker for more realistic assumptions about the average waiting time for breached patients. Figure 7 plots the distribution of mean waiting times across hospitals assuming a 5 hour average for extreme waiting times. As with the measured statistic, the distribution converges over time, and the convergence is once again most rapid in the periods with announced incentives. The slight decrease in performance appears in the final quarter in Figure 6 as well. We further investigate this calculation of "mean" effort in the regressions below.

Table 5 presents summary statistics for the four alternative measures of quality of care. As shown above, the "mean" wait time decreases over time, while the percent of patients treated within two hours increases. The fraction of patients who come into the A&E on a follow-up visit falls a bit over our sample period; the percent of patients admitted to the general hospital increases slightly.

5.2 Results

To test for effort substitution, we estimate regressions of the general form

$$z_{ht} = \alpha + \beta y_{ht} + \nu_h + \varphi_t + \varepsilon_{ht} \tag{3}$$

where z_{ht} represents a non-incentivized measure of performance, y_{ht} is the fraction of patients treated in under four hours, ν_t and φ_h are quarter- and hospital-specific fixed effects, for hospital h in quarter t. We also replace y_{ht} with $\Delta y_{ht} = y_{ht} - y_{h,t-1}$ in some specifications, and we experiment with the alternatively scaled measure of effort $\hat{y}_t$, as discussed above. In this specification, the parameter β estimates the extent to which increases in incentivized performance are associated with changes in in the alternative performance measures; since our measures of alternative performance are bad, an estimate of $\beta > 0$ imply significant levels of effort substitution.[19]

---

[19] So that lower measures of z are always better, we use the fraction of patients treated in *more* than two hours as our fourth statistic.

Using each of these independent variables, we also run regressions using the specification

$$z_{ht} = \alpha + \beta y_{ht} + \gamma y_{ht} * I_t + \nu_h + \varphi_t + \varepsilon_{ht} \tag{4}$$

where I_t is a dummy variable equal to one in the pre-sweeps or the cash incentive periods - that is, in those periods where present or future incentives loomed. The parameter γ estimates effort substitution by looking at the differential response rate to increases in the measured statistic between incentivized and non-incentivized periods; if we estimate $\gamma < 0$, there is evidence of effort substitution. As before, we cluster standard errors by trust.

Table 6 reports results from this set of 32 regressions; there are 4 dependent variables (z_{ht}), 4 versions of the measured statistic (y_{ht}), and 2 specifications (those in equations (3) and (4)). A vertical column displays the four regressions that share a dependent variable and a specification. For instance, $\hat{\beta} = 0.050$ when the dependent variable is "% Waiting < 2 Hours," the independent variable is the log-scaled version of the measured performance ($\hat{y}_{ht}$), and the specification is that in equation (3).

In general, statistically significant coefficients appear in the regressions using "mean" wait time and the sub-two hour wait share as the dependent variable. But the increases in the incentivized task here seem primarily to help these alternative measures, contrary to the effect predicted by effort substitution. For instance, seven of eight of the statistically significant coefficients in the regressions using "mean" wait time are negative, so that, as fewer patients wait more than four hours, mean wait time decreases.[20] Similarly, most of the significant coefficients in regressions using the sub-two hour fraction are positive.

There are a few significant coefficients that go in the other (bad) direction, such as in the second panel of columns (6) and (7). For instance, relative to a similar move in periods without incentives, increasing performance by 1 percentage point increases the fraction of patients waiting more that two hours by 0.275 percentage points. This effect is less than 2.5% of one standard deviation of this variable, though. Similarly, a one percentage point increase in performance increases mean wait time by 0.01 hours more in incentive periods than in other periods. The significant coefficients in regressions using follow-up visits are even smaller in economic magnitude. There is no compelling evidence of any effort substitution.

6 Discussion

While the private sector has traditionally turned to performance-based incentives to manage workforces, governments have only recently begun to do so for their many employees. For example, the

---

[20] This result is robust to instead making the extremely conservative assumption that breached patients wait only four hours, on average.

US government implemented performance-related pay for job-counselors and managers in the Job Training and Partnership Act in 1983. New Zealand's Reserve Bank Act of 1989 details a scheme of incentive-based compensation for the central bank's Chairman. Most recently, school systems in Houston and Denver adopted government-funded incentive pay for teachers.

The success of public sector incentive schemes has been mixed; indeed, theory suggests that the public sector may be especially susceptible to many of the familiar pitfalls of formal incentives. Standard political economy problems may generate crude and inefficient contracts, while the multiple objectives often involved in public sector activities make performance difficult to quantify and thus give agents ample room to "game" the system. Due to the multifaceted nature of public services, Holmström and Milgrom (1991) and Baker (1992) suggest that the optimal contract may avoid incentives entirely. Organizations overseeing service provision in the public sector may be more inefficient than parallel structures in the private sector, perhaps even optimally so (Prendergast, 2003), but the monitoring required for proper incentives may suffer as a result.

But if effort has a positive impact on the capital stock, as in these results, even weak incentives may be able to produce greater effects than typically imagined. Investment effects would also be expected to reduce problems with effort substitution across periods and tasks. Performance-based incentives may thus have substantially greater scope in the public sector than previously believed. That public sector employees tend to be intrinsically motivated, which they may be endogenously (Prendergast, 2007), makes it more likely that gains from increases in human or organizational capital in government should last for some time; in the public management literature, this is often referred to as "public service motivation" (Perry and Wise, 1990, and Crewson, 1997). However, if incentives were to encourage public sector employees to develop experience in unproductive activities, the long-term efforts predicted by our model might make the situation far worse.

Another factor tending towards improvements through investments in the public sector is a political constraint on government statements. In many cases, it is far less politically acceptable for governments to recommend that public sector employees simply "try harder" to improve performance; instead, governments are more likely to suggest various procedural or managerial improvements. One can see this emphasis in general government communications to employees in the context of increasing performance; the British government focused on various process improvements (such as a reform of triage procedures) in the lead-up to the incentives studied here, and the U.S. government has focused almost exclusively on similar changes when dealing with schools (as in the reforms following the No Child Left Behind reforms). If agents find it easiest to "look under the lamp post" and follow government directions, the effects of incentives through investment we study here will be especially relevant.

The theory of effort-as-investment may also suggest that the small incentives we typically see in the public sector might nonetheless produce relatively large performance responses. In order for investment to be an important aspect of agents' responses, there must be relatively high-return investments that have not yet been implemented, since the return to the investment must occur only during the incentive period and not after. Empirically, we see in our data that a quite modest time-limited incentive produced a large improvement in performance. But in this case, one must naturally ask: why were these investments not already undertaken? In the private sector, principals or other residual claimants on output will likely keep relevant capital stocks close to the optimal levels. For instance, a company will reform its procedures before they fall too far out of date. But in the public sector, the natural corrective mechanisms are far weaker; it is more likely that there exist high-return investments that have not been made. If low-powered incentives, announced appropriately early, can convince employees to undertake these procedural changes, then in the public sector, weak incentives may be incentive enough.

How do our results relate to specific circumstances where people have considered incentives in the public sector? Schools are one area in which policymakers frequently discuss the possibility of incentives. Our model predicts that the success of such incentives depends on the extent to which they encourage teachers to learn how to teach better. If they do so, the benefits will likely last, especially since teachers tend to be strongly intrinsically motivated. One concern raised by many opponents of teacher incentives is that teachers will not learn to teach better, but rather they will "teach to the test," that is engage in various forms of effort substitution, gaming, or even outright cheating. Our model suggests that if this occurs the problem may be even worse than commonly thought; the negative effects of effort substitution and gaming will appear not only when tests are given but also before and after, as teachers learn to alter their teaching styles.

7 Conclusion

This paper develops a model in which effort can have an impact on capital stocks, such as the human capital or the organizational structure of a service provider. Though theoretically the effect can be either positive or negative, our empirical evaluation shows that the impact here is large and positive. As evidence of this we have shown that performance increases in advance of incentivized periods and remains high after the rewards have passed. We also examine performance along non-incentivized dimensions of quality of care and find little evidence of classical effort substitution or "gaming."

8 Bibliography

1. Argote, Linda (1999). *Organizational Learning: Creating, Retaining and Transferring Knowledge.* New York: Springer.

2. Arrow, Kenneth. (1962). "The Economic Implications of Learning by Doing." *Review of*

Economic Studies, Vol 29(3), pp. 155-73.

3. Baker, George. (1992). “Incentive Contracts and Performance Measurement.” *Journal of Political Economics*, Vol 100, pp. 598-614.
4. Barnard, Chester I. (1938). *The Functions of the Executive.* Cambridge, MA: Harvard University Press.
5. Bolton, Patrick, and Mathias Dewatripont. (2005). *Contract Theory.* Cambridge, MA: MIT Press.
6. British Broadcasting Corporation. (2006). “Trust Plans Hospital Takeover.” Available at *http://news.bbc.co.uk/1/hi/england/west_midlands/5399130.stm*
7. Chart, Elizabeth. (1996). “The Performance Measurement Agenda in the NHS.” University of Southampton Unpublished Masters Thesis.
8. Crewson, Philip E. (1997). “Public Service Motivation; Building Empirical Evidence of Incidence and Effect.” *Journal of Public Adminstration Research and Theory*, Vol. 4, pp. 499-518.
9. Coate, Steven, and Steven Morris. (1999). “Policy Persistence.” *American Economic Review*, Vol 89(5), p. 1327-36.
10. Deci, Edward. (1971). “Effects of Externally Mediated Rewards on Intrinsic Motivation.” *Journal of Personality and Social Psychology*, Vol 18, p. 105-15.
11. Department of Health. (2000). “The NHS Plan: A Plan for Investment, A Plan for Reform.” Available at *http://www.nhsia.nhs.uk/nhsplan/nhsplan.htm*
12. Department of Health. (2001). “NHS Reform Bill Published.” Available at *http://www.dh.gov.uk/PublicationandSt*
13. Department of Health. (2005a). “Department Report 2005.” Available at *http://www.dh.gov.uk/assetRoot/04/11/*
14. Department of Health. (2005b). “A&E Update, 10/12.” Section 5.1.1.
15. Dixit, Avinash K. (1996). *The Making of Economic Policy: A Transaction-Cost Politics Perspective.* Cambridge, MA: MIT Press.
16. Dixit, Avinash K. (2002). “Incentives and Organizations in the Public Sector; An Interpretive Review.” *Journal of Human Resources*, Vol 37(4), pp. 696-727.
17. Dranove, David, Daniel Kessler, Mark McCellan, and Mark Satterthwaite. (2003). “Is More Information Better? The Effects of “Report Cards” on Healthcare Providers.” *Journal of Political Economy*, Vol 111(3), pp. 555-88.

18. Friedman, John. (2007). "What Do Government Interventions Signal? A 'Lucas Critique' of Program Evaluation." Harvard University Working Paper.

19. Friedman, John and Steve Kelman. (2007a). "Incentives as Signals." Harvard University Working Paper.

20. Friedman, John and Steve Kelman. (2007b) "Effort Substitution, 'Gaming,' and the Response to Performance Targets: Evidence from Hospital Emergency Room Waiting Time Targets in England." Harvard University Working Paper.

21. Gneezy, Uri and Aldo Rustichini. (2000). "Pay Enough or Dont Pay At All." *Quarterly Journal of Economics*, Vol 115, p. 791-811.

22. Harris, Milton, and Bengt Holmström. (1982). "A Theory of Wage Dynamics." *Review of Economic Studies*, Vol 49, p. 315-53.

23. Healy, Paul M. (1985). "The effect of Bonus Schemes on Accounting Schemes." *Journal of Accounting and Economics*, Vol. 7, pp., 85-107.

24. Heckman, James, Carolyn Heinrich and Jeffrey Smith (1997). "Assessing the Performance of Performance Standards in Public Bureaucracies." *American Economic Review*, Vol 87, p. 389-95.

25. Holmström, Bengt, and Paul Milgrom. (1991). "Multitask Principal-Agent Analyses: Incentive Contract, Asset Ownership, and Job Design." *Journal of Law, Economics, and Organizations*, Vol 7, pp.24-53.

26. Kelman, Steve. (2006). "Improving Service Delivery Performance in the United Kingdom: Organization Theory Perspective on Central Intervention Strategies." *Journal of Comparative Policy Analysis*, Vol 8(4), p. 393-419.

27. McNichols, Maureen F. (2000). "Research Design Issues in Earnings Management Studies." *Journal of Accounting and Public Policy*, Vol 19, pp. 313-45.

28. Nelson, Richard and Sidney G. Winter. (1982) *An Evolutionary Theory of Economics Change.* Cambridge, MA: Belknap Press of Harvard University Press.

29. Oyer, Paul (1998). "Fiscal Year Ends and Nonlinear Incentive Contracts: The Effects on Business Seasonality." *Quarterly Journal of Economics*, Vol 113, pp. 149-85.

30. Perry, James L. and Lois R. Wise. (1990). "The Motivational Bases of Public Service." *Public Administration Review*, Vol 50, pp. 367-73.

31. Prendergast, Canice. (1999). "The Provision of Incentives in Firms." *Journal of Economic Literature*, Vol 3, pp. 7-63.

32. Prendergast, Canice. (2003). "The Limits of Bureaucratic Efficiency." *Journal of Political Economy*, Vol 111(5), pp. 929-58.

33. Prendergast, Canice. (2007). "The Motivation and Bias of Bureaucrats." *American Economic Review*, Vol. 97(1), pp. 180-96.

34. Rosenthal, Meredith, and Richard G. Frank. (2006). "What is the Empirical Basis for Paying for Quality in Healthcare?" M*edical Care Research and Review*, Vol 63(2), pp. 135-57.

35. Schmenner, Roger. (1993). *Production/Operations Management.* Fifth Edition, Englewood Cliffs: Prentice-Hall.

36. Stein, Jeremy. (1989). "Efficient Capital Markets, Inefficient Firms." *Quarterly Journal of Economics*, Vol 104(4), pp. 655-69.

37. Williamson, Oliver E. (1975). *Markets and Hierarchies: Analysis and Antitrust Implications.* New York: The Free Press.

9 Appendix

Proof of Proposition 1. The agent solves the problem

$$\max_{\{e_t\}_{t=0}^{\infty}} E\sum_{t=0}^{\infty}\beta^t\left[v_t - c\left(e_t\right)\right],$$

where $v_t = 0$ for $t \neq T$, and $v_t = wy_T + W * I\left\{y_t > \chi\right\}$. Substituting in for the production function, the capital stock, and the expected reward, we can rewrite the program as

$$\max_{\{e_t\}_{t=0}^{\infty}} \beta^T\left[W\left(1 - F\left(\chi - \gamma e_T - (1-\delta)\bar{K} - \phi\sum_{t=0}^{T-1}\delta^{T-t}e_t\right)\right) + w\left(\gamma e_T - (1-\delta)\bar{K} - \phi\sum_{t=0}^{T-1}\delta^{T-t}e_t\right)\right] - \sum_{t=0}^{\infty}\beta^t c\left(e_t\right).$$

The program seeks to maximize a concave function over a convex set, and so the first order conditions are necessary and sufficient conditions for maximization. Differentiation yields the set of FOCs

$$\begin{aligned}
\beta^T\phi\delta^{T-t}\left[Wf\left(\chi - Ey_T\right) + w\right] &= \beta^t c'\left(e_t\right) \quad \text{for} \quad t < T \\
\beta^T\gamma\left[Wf\left(\chi - Ey_T\right) + w\right] &= \beta^T c'\left(e_t\right) \quad \text{for} \quad t = T \\
0 &= \beta^t c'\left(e_t\right) \quad \text{for} \quad t > T.
\end{aligned}$$

It is clear that, for all parameter values, $e_t^* = 0$ for $t > T$. Furthermore, defining $z(\cdot) = [c']^{-1}$ and taking ratios implies that

$$\frac{e_t}{e_T} = \frac{z\left(\phi(\beta\delta)^{T-t}\left[Wf(\chi - Ey_T) + w\right]\right)}{z\left(\gamma\left[Wf(\chi - Ey_T) + w\right]\right)}. \tag{5}$$

Note that $z(0) = 0$. Thus, we know that

$$\phi(\beta\delta)^{T-t}\left[Wf(\chi - Ey_T) + w\right] = 0 \iff e_t = 0$$

and therefore that $\phi = 0$ or $[Wf(\chi - Ey_T) + w] = 0$ imply $e_t = 0$. Similarly, $\gamma = 0$ or $[Wf(\chi - Ey_T) + w] = 0$ imply $e_T = 0$.

We also know that $z(\cdot)$ is an increasing function, and therefore that, when $\phi[Wf(\chi - Ey_T) + w] > 0$, that $e_t < e_{t+1}$ for $t < T$. Furthermore, if $c''' = 0$ (which would imply that $z(x) = cx$ for somew constant c), then

$$\begin{aligned}\frac{e_t}{e_T} &= \frac{\phi}{\gamma}(\beta\delta)^{T-t} \\ \frac{e_{t-1}}{e_t} &= \beta\delta.\end{aligned}$$

This case, in which the marginal disutility of effort increases linearly, provides a useful benchmark. If this is the case, and if $\phi > 0$, effort will increase exponentially up to a peak at $t = T$. Suppose now that $c''' \neq 0$. If marginal disutility of effort is convex (concave), then $z(\cdot)$ is concave (convex). If $c''' > 0$, then we know that

$$\frac{e_t}{e_T} = \frac{z\left(\phi(\beta\delta)^{T-t}\left[Wf(\chi - Ey_T) + w\right]\right)}{z\left(\gamma\left[Wf(\chi - Ey_T) + w\right]\right)} > \frac{\phi}{\gamma}(\beta\delta)^{T-t}$$

where the opposite is true if $c''' < 0$. If $\phi \leq 0$, then the standard model applies, in which case $e_t = 0$ except at $t = T$. ■

Proof of Proposition 2. From equation (5) above, we know that, when $\phi > 0$, then

$$\begin{array}{ll} e_t = z\left(\phi(\beta\delta)^{T-t}\left[Wf(\chi - Ey_T) + w\right]\right) & \text{for } t < T \\ e_T = z\left(\gamma\left[Wf(\chi - Ey_T) + w\right]\right) & \text{for } t = T \end{array}.$$

Thus, without needing to calculate Ey_T explicitly, we can derive the time series ratios of output. We know that

$$Ey_t = \sum_{s=0}^{t-1}\delta^{t-s}\phi e_s + \gamma e_t + \bar{K} = \sum_{s=0}^{t-1}\phi\delta^{t-s}z\left(\phi(\beta\delta)^{T-s}\left[Wf(\chi - Ey_T) + w\right]\right) + \gamma z\left(\phi(\beta\delta)^{T-t}\left[Wf(\chi - Ey_T) + w\right]\right) + \bar{K}.$$

At this point, it is convenient to normalize $\bar{K} = 0$; this is equivalent to redefining a new variable $\hat{y}_t = y_t - y_{-1} = y_t - \bar{K}$. Note that we also replace $\chi = \chi - \bar{K}$. The ratio between output in different periods will then satisfy

$$\frac{Ey_{t-1}}{Ey_t} = \frac{\sum_{s=0}^{t-2} \phi\delta^{t-s-1} z\left(\phi\left(\beta\delta\right)^{T-s}\left[Wf\left(\chi - \bar{K} - Ey_T\right) + w\right]\right) + \gamma z\left(\phi\left(\beta\delta\right)^{T-t+1}\left[Wf\left(\chi - \bar{K} - Ey_T\right) + w\right]\right)}{\sum_{s=0}^{t-1} \phi\delta^{t-s} z\left(\phi\left(\beta\delta\right)^{T-s}\left[Wf\left(\chi - \bar{K} - Ey_T\right) + w\right]\right) + \gamma z\left(\phi\left(\beta\delta\right)^{T-t}\left[Wf\left(\chi - \bar{K} - Ey_T\right) + w\right]\right)}.$$

Since $z' > 0$, then for all s

$$\begin{aligned} z\left(\phi\left(\beta\delta\right)^{T-s-1}\left[Wf\left(\chi - \bar{K} - Ey_T\right) + w\right]\right) &> z\left(\phi\left(\beta\delta\right)^{T-s}\left[Wf\left(\chi - \bar{K} - Ey_T\right) + w\right]\right) \\ \gamma z\left(\phi\left(\beta\delta\right)^{T-t}\left[Wf\left(\chi - \bar{K} - Ey_T\right) + w\right]\right) &> \gamma z\left(\phi\left(\beta\delta\right)^{T-t+1}\left[Wf\left(\chi - \bar{K} - Ey_T\right) + w\right]\right) \end{aligned}$$

the set of terms in the denominator pairwise deminates those in the denominator, the extra positive term in the denominator ($\phi\delta^t z\left(\phi\left(\beta\delta\right)^{T-s}\left[Wf\left(\chi - \bar{K} - Ey_T\right) + w\right]\right)$ remains unpaired) notwithstanding. Therefore,

$$Ey_t > Ey_{t-1} \quad \forall t \leq T.$$

We can solve for the exact growth rate with two more assumptions.First, $c''' = 0$, then the ratio $\frac{Ey_t}{Ey_{t-1}}$ simplifies to

$$\frac{Ey_{t-1}}{Ey_t} = \beta\delta\left[\frac{\phi^2\frac{\beta\delta^2}{1-\beta\delta^2}\left(1 - \left(\beta\delta^2\right)^{t-1}\right) + \gamma\phi}{\phi^2\frac{\beta\delta^2}{1-\beta\delta^2}\left(1 - \left(\beta\delta^2\right)^{t}\right) + \gamma\phi}\right] < 1. \tag{6}$$

Now, keeping $T - t$ constant, let $t \to \infty$, so that the incentive is announced an infinite time in advance. The ratio limits to

$$\frac{Ey_{t-1}}{Ey_t} = \beta\delta < 1$$

The ratio involving the incentive period is slightly different, at

$$\frac{Ey_{T-1}}{Ey_T} = \beta\delta\left[\frac{\phi^2\frac{\beta\delta^2}{1-\beta\delta^2}\left(1 - \left(\beta\delta^2\right)^{T}\right) + \phi\gamma}{\phi^2\frac{\beta\delta^2}{1-\beta\delta^2}\left(1 - \left(\beta\delta^2\right)^{T}\right) + \gamma^2}\right] \tag{7}$$

This ratio is greater (less than) $\beta\delta$ when $\gamma > (<)\,\phi$, and does not tend towards $\beta\delta$ in general. Thus, if $\phi > 0$, output increases exponentially up until $t = T$.

For $t > T$, we know that

$$Ey_t = \phi\delta^{t-T}\sum_{s=0}^{T}\delta^{T-s}e_s = \phi\delta^{t-T}\left[\sum_{s=0}^{T-1}\delta^{T-s}z\left(\phi\left(\beta\delta\right)^{T-s}\left[Wf\left(\chi - \bar{K} - Ey_T\right) + w\right]\right) + z\left(\gamma\left[Wf\left(\chi - \bar{K} - Ey_T\right) + w\right]\right)\right]$$

and thus that

$$\frac{Ey_{t+1}}{Ey_t} = \delta \tag{8}$$

and

$$\frac{Ey_{T+1}}{Ey_T} = \delta \left[\frac{\phi^2 \frac{\beta\delta^2}{1-\beta\delta^2} \left(1 - \left(\beta\delta^2\right)^T\right) + \gamma\phi}{\phi^2 \frac{\beta\delta^2}{1-\beta\delta^2} \left(1 - \left(\beta\delta^2\right)^T\right) + \gamma^2} \right] \tag{9}$$

which is less (greater) than δ when γ is greater (less) than ϕ. Beginning at time $t = T+1$, output remains positive but decays exponentially as the capital stock depreciates back to $\bar{K}$.

When $\phi < 0$, then $Ey_T = e_T = z\left(\gamma\left[Wf\left(\chi - \bar{K} - Ey_T\right) + w\right]\right)$ as above. But instead of decaying down to 0, output will drop to

$$Ey_{T+1} = \delta\phi e_T < 0,$$

after which point the ratio $\frac{Ey_{t+1}}{Ey_t} = \delta$ from above will obtain, with all $Ey_t < 0$ where $t > T$.[21] ■

Proof of Proposition 3. The level of effort depends on the marginal distribution of ε_t around the value $\chi - \bar{K} - Ey_T$. We know that

$$Ey_T = \sum_{s=0}^{T-1} \phi\delta^{T-s} z\left(\phi\left(\beta\delta\right)^{T-s}\left[Wf\left(\chi - \bar{K} - Ey_T\right) + w\right]\right) + \gamma z\left(\gamma\left[Wf\left(\chi - \bar{K} - Ey_T\right) + w\right]\right).$$

Since $z' > 0$ and $f' > 0$ by assumption, implicitly differentiating this expression implies that

$$\frac{\partial Ey_T}{\partial \bar{K}} = \frac{\sum_{s=0}^{T-1} W\phi^2\left(\beta\delta^2\right)^{T-s} z'\left(\phi\left(\beta\delta\right)^{T-s}\left[Wf + w\right]\right) f' + \gamma^2 W z'\left(\gamma\left[Wf + w\right]\right) f'}{\sum_{s=0}^{T-1} W\phi^2\left(\beta\delta^2\right)^{T-s} z'\left(\phi\left(\beta\delta\right)^{T-s}\left[Wf + w\right]\right) f' + \gamma^2 W z'\left(\gamma\left[Wf + w\right]\right) f' + 1} \in (0,1)$$

It is a parallel calculation that, given this derivative, $\frac{\partial Ey_t}{\partial \bar{K}}$ for which $t \neq T$. ■

Proof of Corollary 1. If there is a fixed cost ψ to acting in each period, then, if agents act, they will act as they did before, since ψ does not alter the incentives on the margin. However, the agent will only act if

$$\Delta Ey_T\left(e_t\right)\left[Wf\left(\chi - \bar{K} - Ey_T\right) + w\right] > \psi.$$

We know from the production function and equation (5) that

$$\Delta Ey_T\left(e_t\right) = \phi\delta^{T-t} z\left(\phi\left(\beta\delta\right)^{T-t}\left[Wf\left(\chi - \bar{K} - Ey_T\right) + w\right]\right),$$

---

[21] In the case where $\phi < 0$. The ratio in equations (6) through (9) still hold. Furthermore, we know that $Ey_t > 0$. Note, however, that the fractional term in brackets may be negative if

$$\gamma < -\phi\frac{\beta\delta^2}{1-\beta\delta^2}\left(1 - \left(\beta\delta^2\right)^T\right).$$

In this case, output will actually decrease at the point when the incentives are announced, with the output depression growing exponentially until output jumps to $y_T > 0$ in period T. After period T, output is once again less than the reference level until it converges back up to 0.

which is clearly a monotonically increasing function in t. Thus there exists a unique τ such that

$$\Delta E y_T\left(e_{\tau-1}\right)\left[W f\left(\chi-\bar{K}-E y_T\right)+w\right]<\psi<\Delta E y_T\left(e_\tau\right)\left[W f\left(\chi-\bar{K}-E y_T\right)+w\right].$$

■

Figure 1: Predicted Time Series of Output for Different "Phi" Values

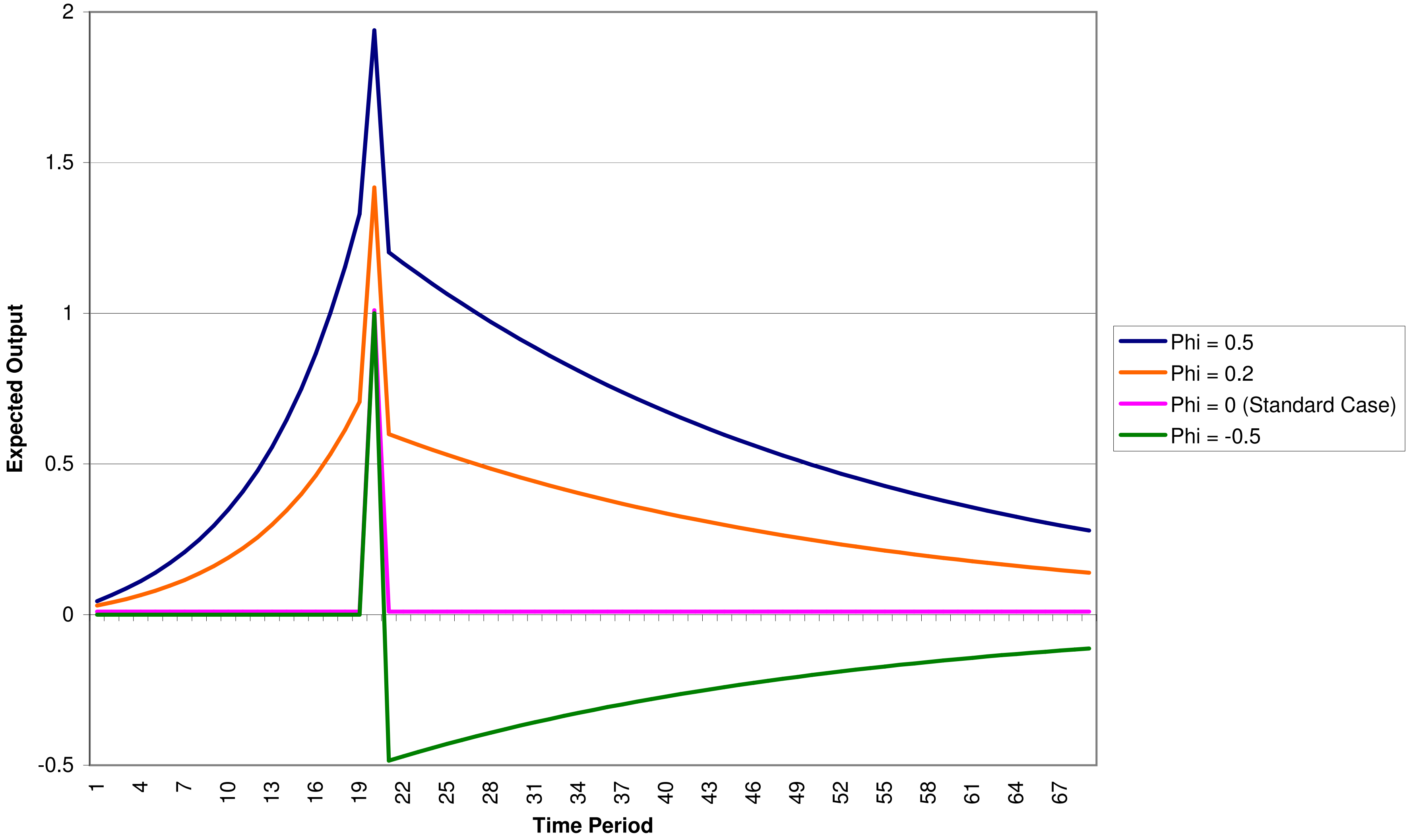

Figure 2: Mean Hospital Performance over Time

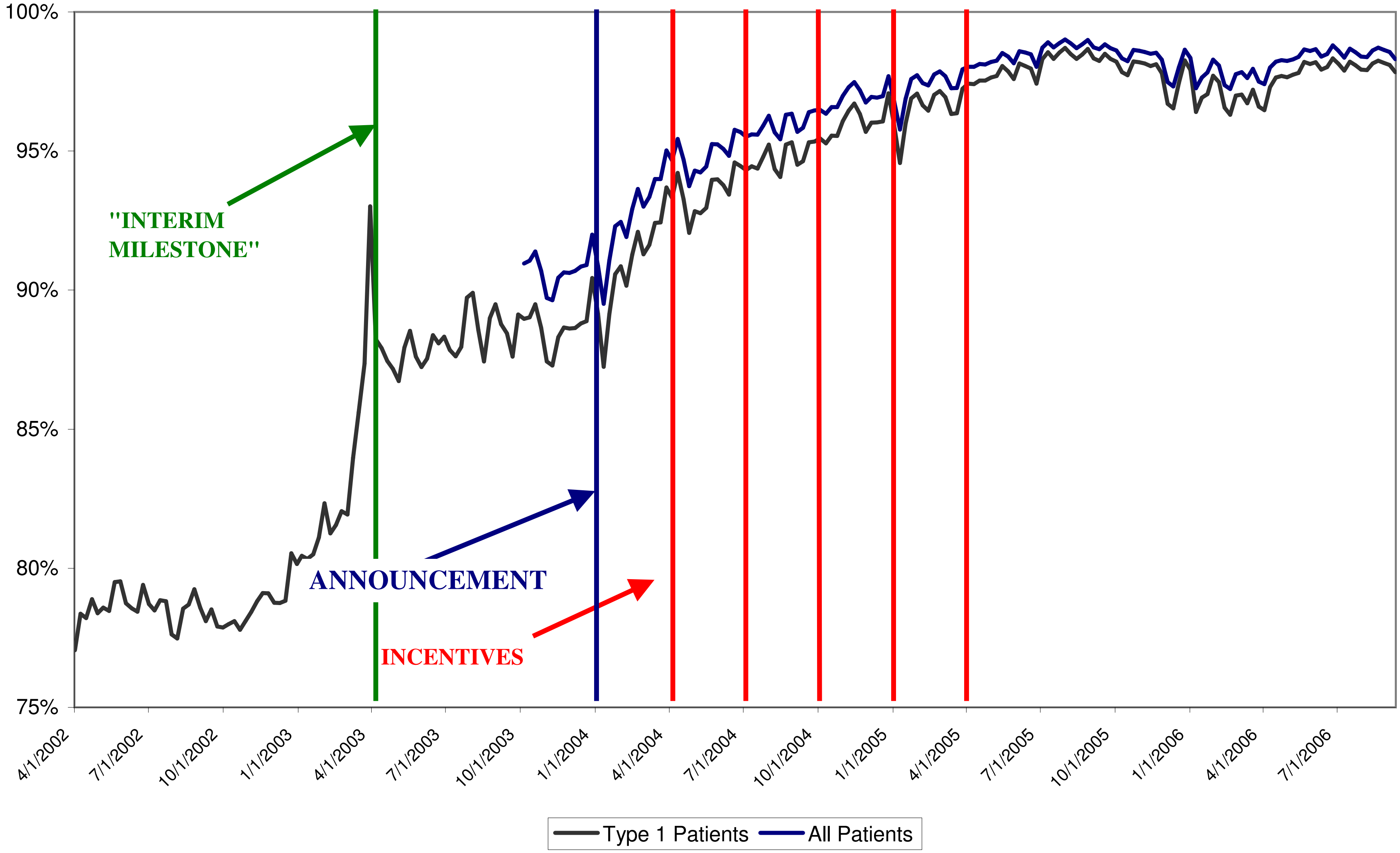

Figure 3: Quantiles of Hospital Performance over Time

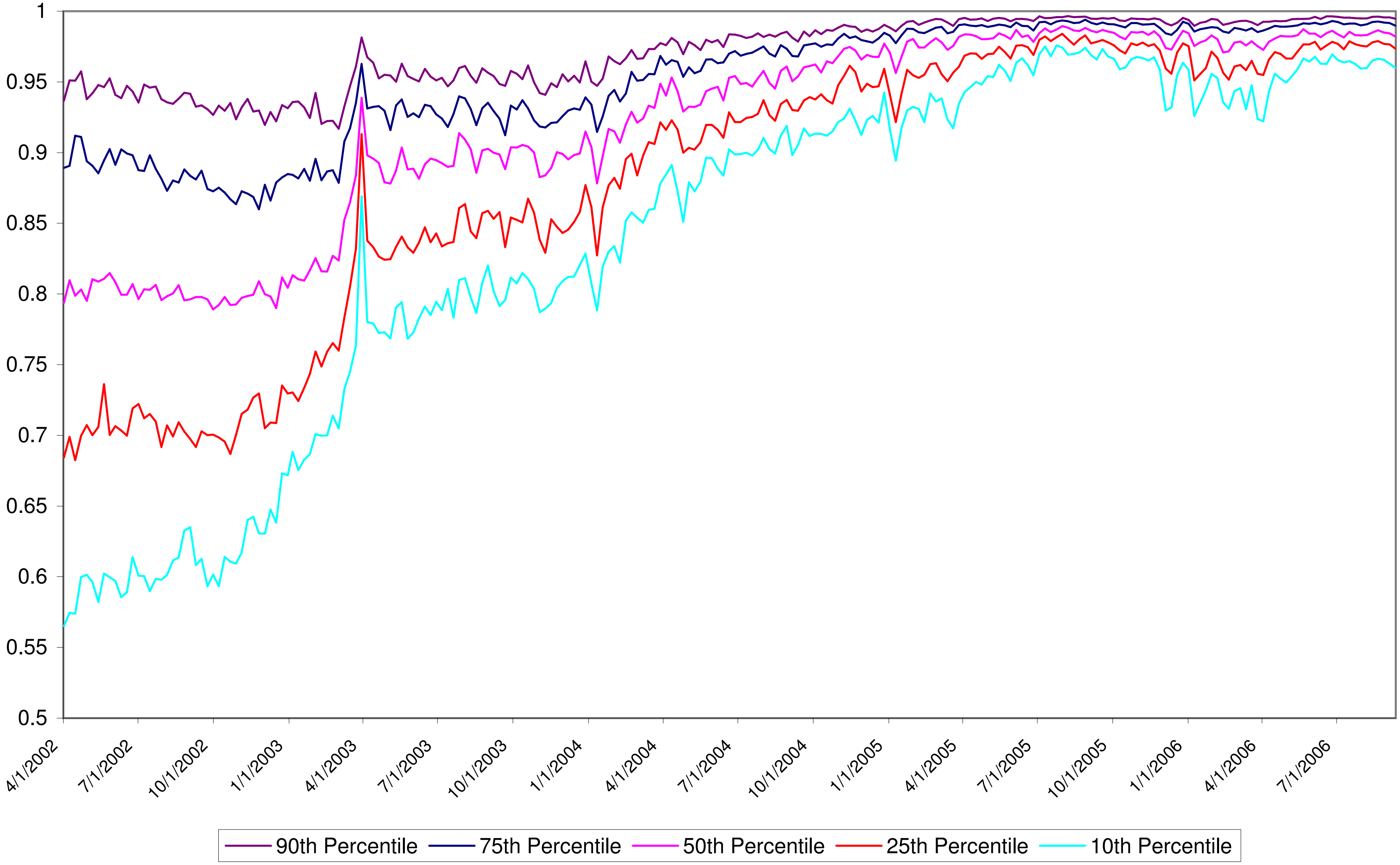

Figure 4: Hospital Performance over Time, by Initial Quality

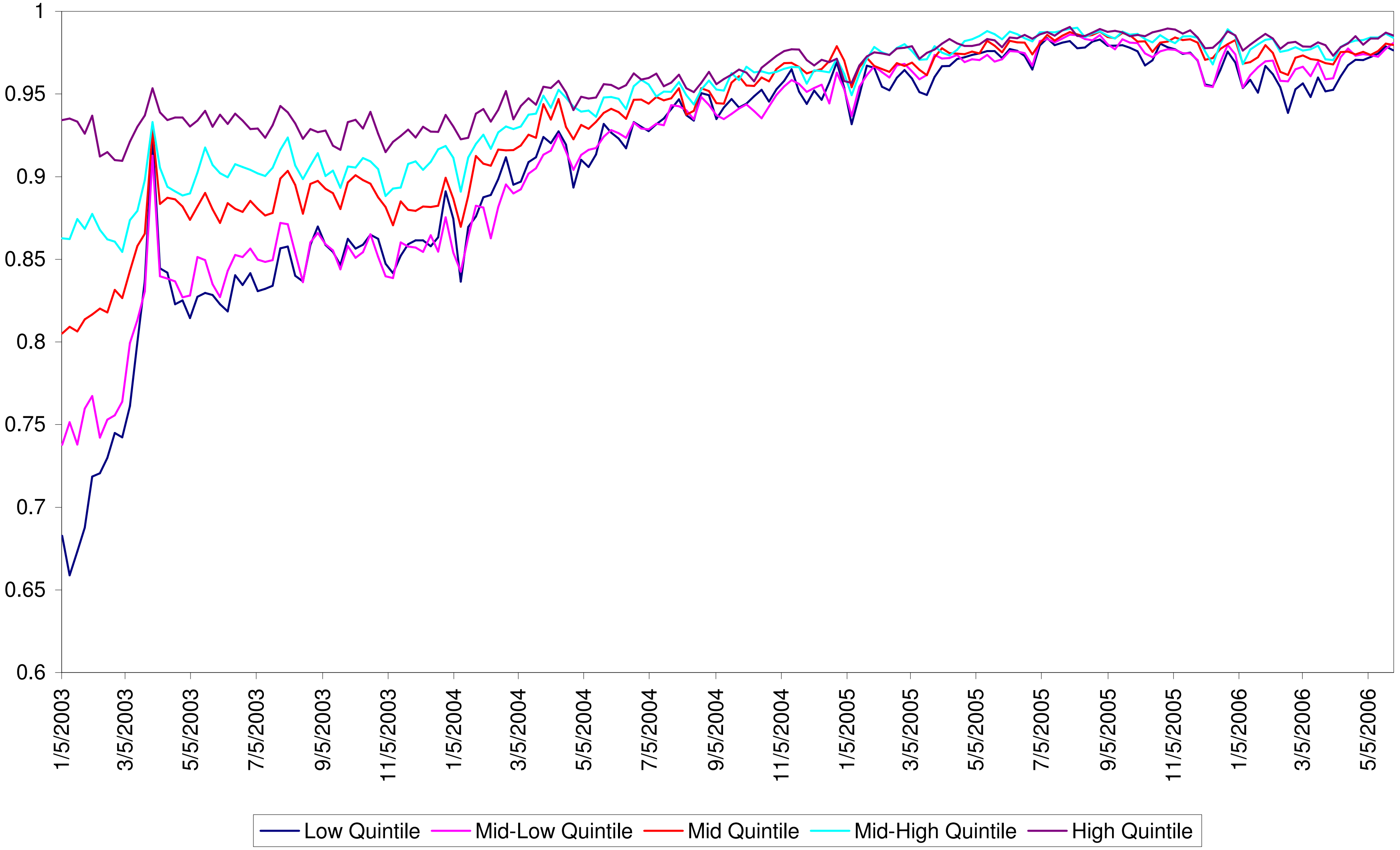

Table 5A: Distribution of Waiting Times

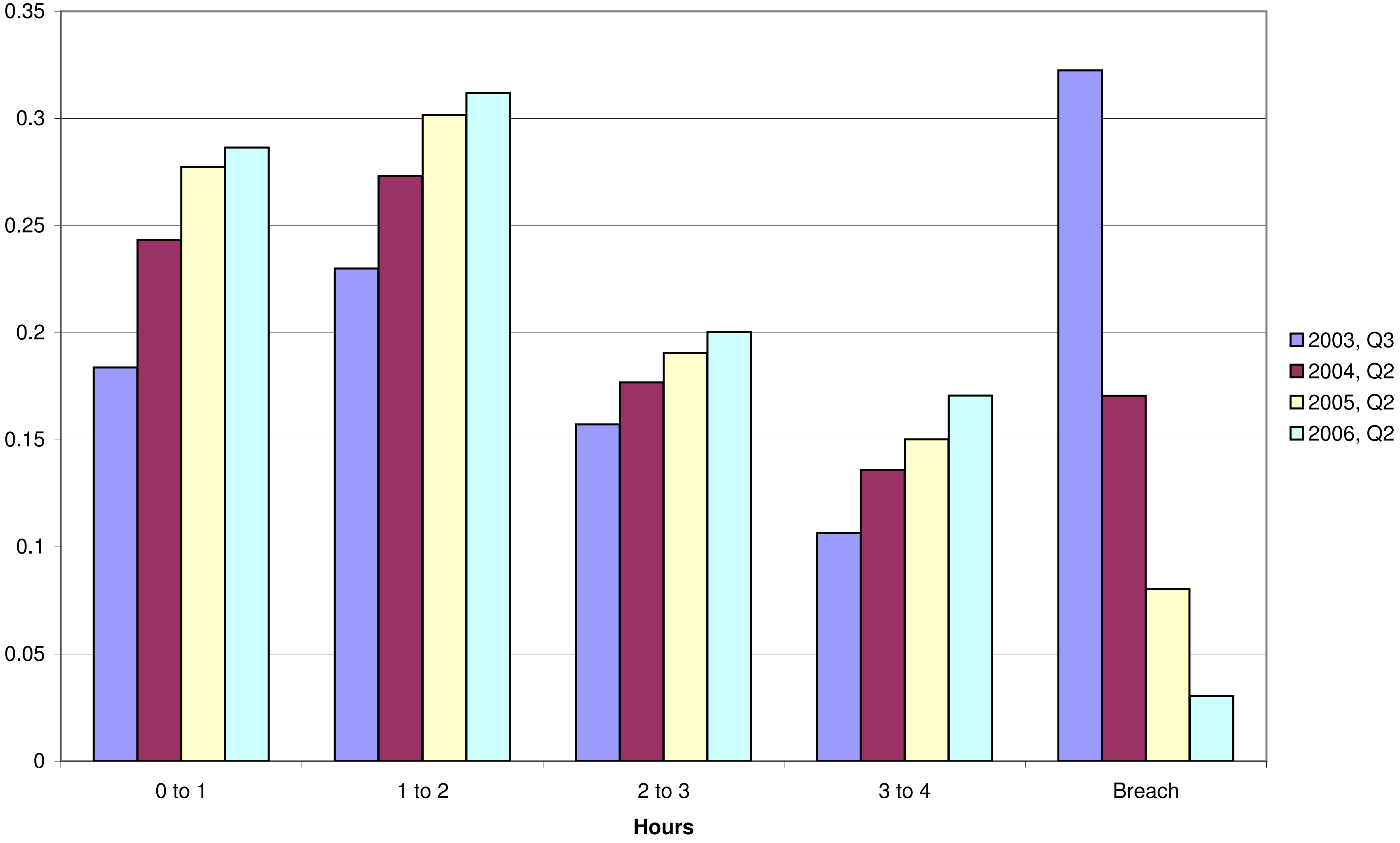

Table 5B: Evolution of the Waiting Time Distribution

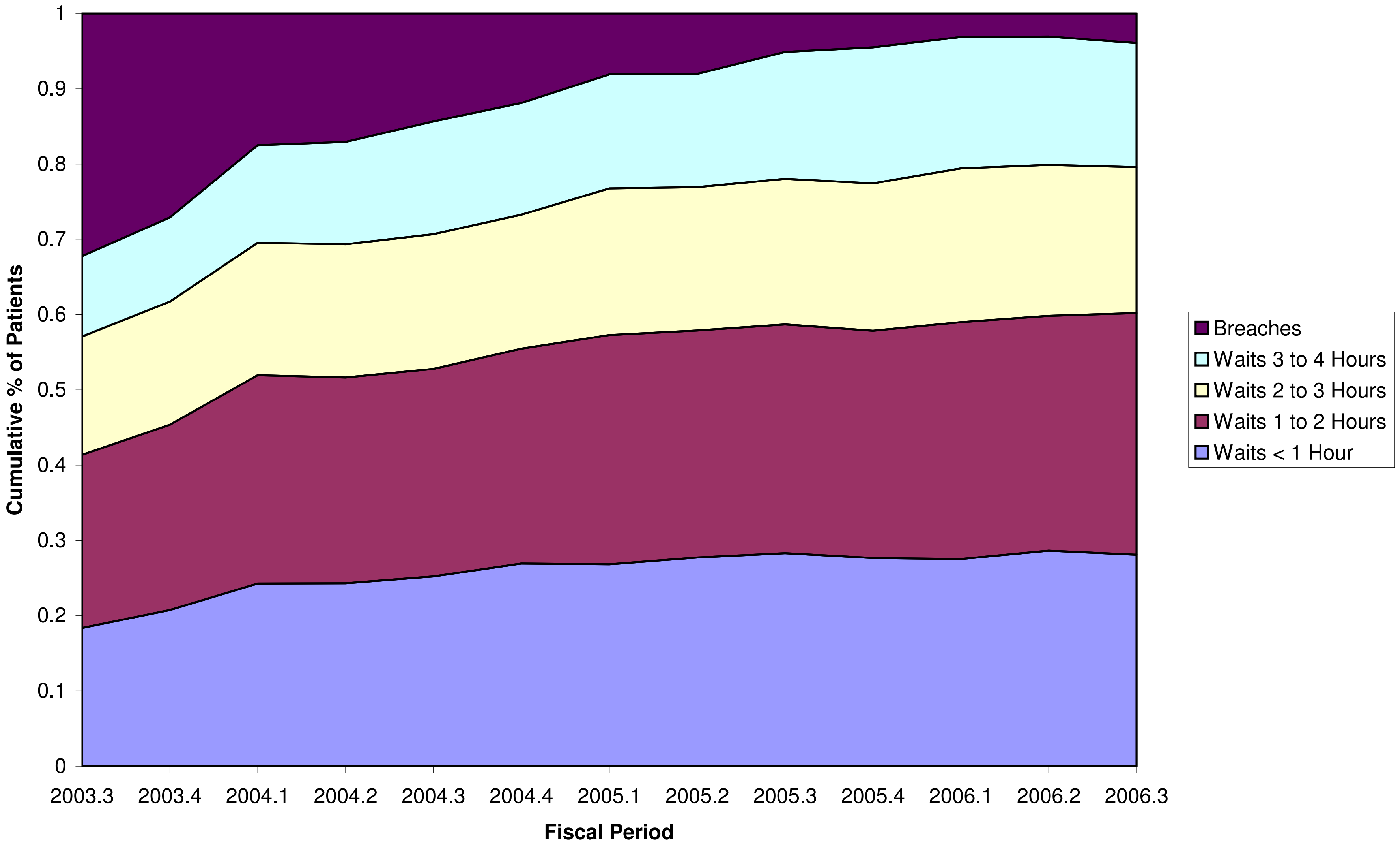

Figure 6: "Mean" Waiting Times

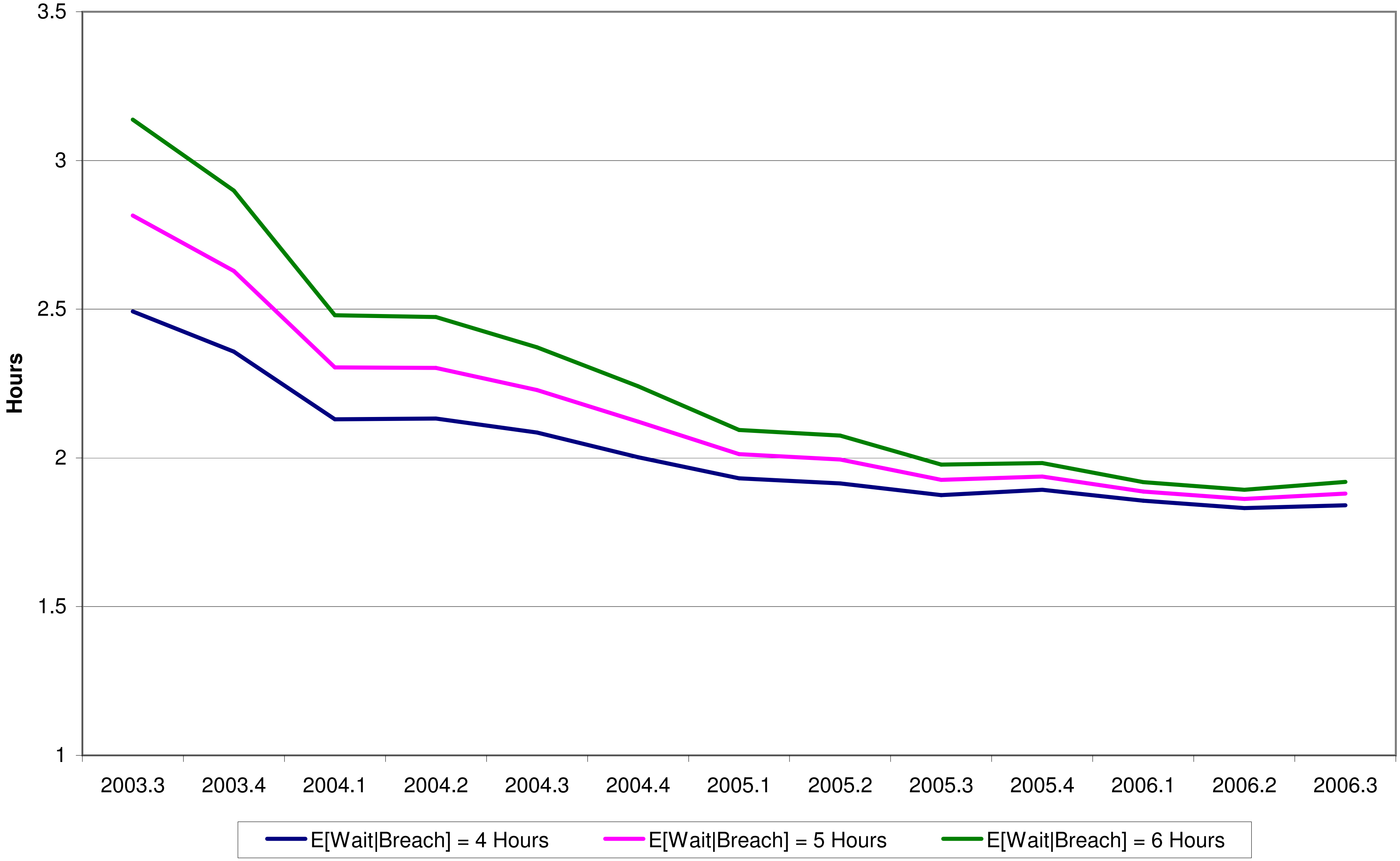
3.5
3
2.5
2
1.5
1
Hours
2003.3
2003.4
2004.1
2004.2
2004.3
2004.4
2005.1
2005.2
2005.3
2005.4
2006.1
2006.2
2006.3
E[Wait|Breach] = 4 Hours
E[Wait|Breach] = 5 Hours
E[Wait|Breach] = 6 Hours

Figure 7: Distribution of "Mean" Waiting Time Across Hospitals

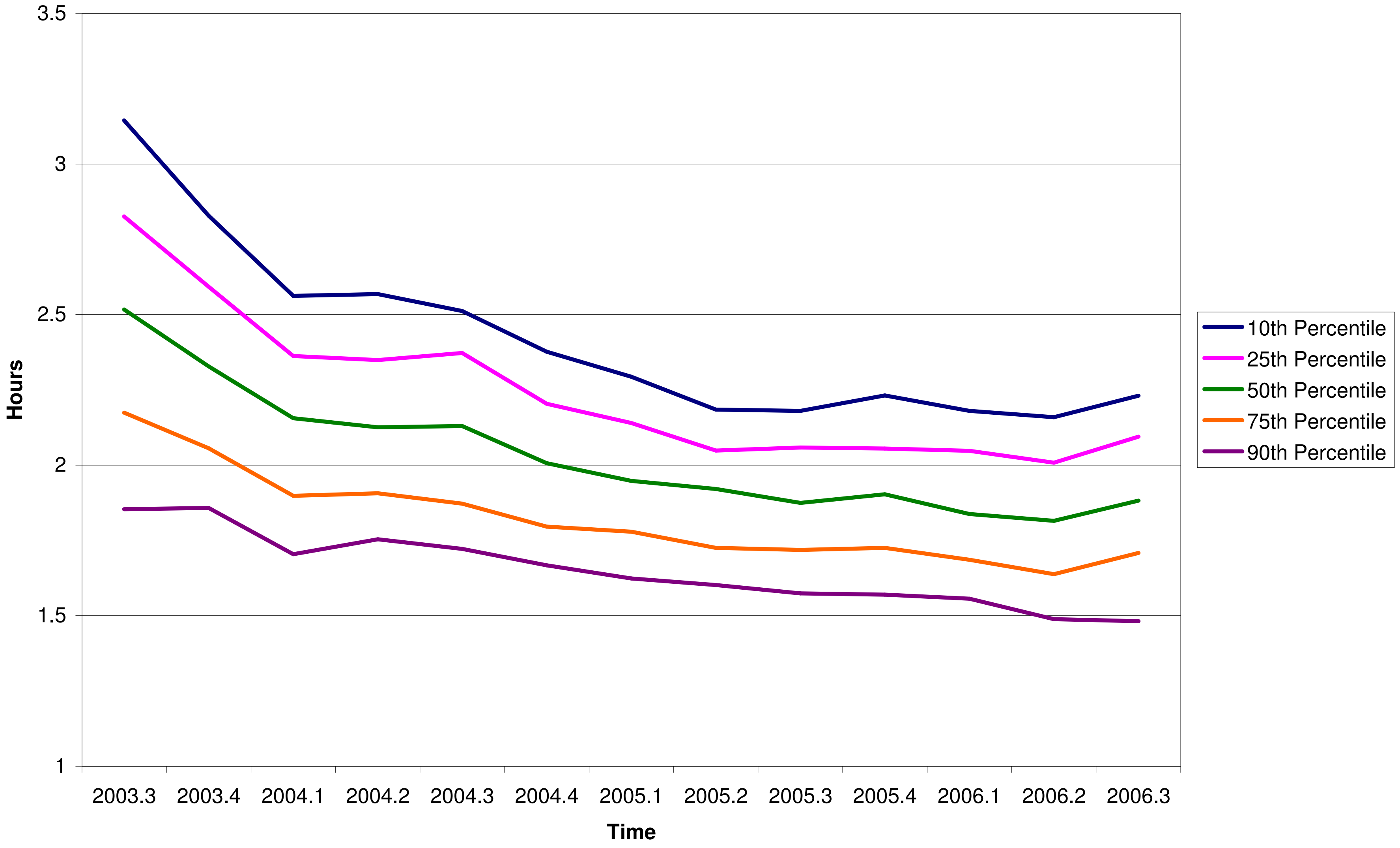

Table 1: Threshold Clearance Rates

Threshold	Rate
94%	52.6%
95%	43.5%
96%	42.2%
97%	42.9%
98%	27.9%
Met All Thresholds	9.1%
Met No Thresholds	20.8%

Table 2: Performance Summary Statistics

Panel A: Entire Sample

Variable	Mean	Std. Dev.	w/i Trust Std. Dev.	N
Performance: Type 1 Patients	93.73%	6.41%	2.26%	35865
Performance: All Patients	96.51%	3.46%	1.89%	23870
Type 1 Attendances	1625	696	123	35865
Type 1 Breaches	76.2	87.7	51.3	35865

Panel B: Type 1 Performance, Period by Period

Period	Variable	Mean	Std. Dev.	w/i Trust Std. Dev.	N
Pre-Sweeps	Performance	82.37%	9.05%	2.84%	7510
	Attendance	1441	629	106	7510
	Breaches	263.9	194.1	69.1	7510
Sweeps Week	Performance	93.01%	5.02%	-	155
	Attendance	1606	709	-	155
	Breaches	114.4	95.6	-	155
Middle	Performance	88.34%	6.20%	2.96%	6355
	Attendance	1564	668	110	6355
	Breaches	188.7	137.1	63.8	6355
Cash Incentive	Performance	94.54%	4.01%	2.37%	9765
	Attendance	1608	687	127	9765
	Breaches	92.1	87.5	57.4	9765
Post Incentive	Performance	97.79%	2.00%	1.51%	11780
	Attendance	1702	718	130	11780
	Breaches	40.0	42.9	31.8	11780

Table 3: Time Series Analysis of Performance

	Dependent Variable:					
Explanatory	*Type 1 Performance*			*-LOG(1-Performance)*		
Variables:	(1)	(2)	(3)	(4)	(5)	(6)
"Sweeps Week" Jump	7.20**	7.23**	7.22**	80.4**	82.8**	80.1**
	(0.49)	(0.49)	(0.49)	(5.54)	(5.16)	(5.62)
Post "Sweeps Week" Bump	1.91**	1.64**	1.67**	17.5**	17.3**	2.55
	(0.42)	(0.43)	(0.43)	(3.69)	(3.89)	(4.81)
"Cash Period" Jump	1.76**	-0.45	0.27	(0.98)	5.52	83.44
	(0.26)	(0.45)	(0.47)	(4.04)	(6.54)	(8.25)
Post-Cash Jump	0.21	1.24**	1.13**	28.1**	33.6**	-15.3**
	(0.13)	0.20	(0.20)	(3.81)	(5.12)	(5.45)
Pre-"Sweeps" Slope	0.52**	-	-	3.19**	-	-
	(0.06)			(0.49)		
Post-"Sweeps" ΔSlope	-0.50**	-0.45**	-0.48**	-3.06**	-1.41	-4.39**
	(0.07)	(0.10)	(0.10)	(0.53)	(0.90)	(1.25)
"Cash Period" ΔSlope	0.07**	0.21**	0.18**	1.95**	1.19*	-5.15**
	(0.02)	(0.04)	(0.04)	(0.16)	(0.52)	(0.71)
Post-"Cash" ΔSlope	-0.10**	0.04	-0.02	-2.21**	0.13	-2.73**
	(0.01)	(0.03)	(0.03)	(0.11)	(0.66)	(0.76)
Quintic in Time?	no	yes	yes	no	yes	yes
Trust Effects?	yes	yes	yes	yes	yes	yes
Month Effects?	no	no	yes	no	no	yes
R-Squared	0.6681	0.6706	0.6708	0.6880	0.6917	0.6965
N	29915	29915	29915	29915	29915	29915

Statistical significance is denoted with the system: * 5%, ** 1%. Standard errors are clustered at the trust level.

Table 4: Convergence Across Hospitals over Time

Explanatory Variables:	*Dependent Variable:* *Type 1 Performance* (1)	(2)	*-LOG(1-Performance)* (3)	(4)
Pre-Sweeps Slope	0.583**	-	3.65**	-
	(0.027)		(0.27)	
Pre-Sweeps Convergence	-0.709*	-0.709*	0.344	0.344
	(0.347)	(0.347)	(0.594)	(0.598)
Post-"Sweeps" ΔSlope	-0.553**	-0.495**	-3.52**	-1.86**
	(0.027)	(0.032)	(0.28)	(0.37)
"Cash Period" ΔSlope	0.069**	0.211**	1.95**	1.20**
	(0.004)	(0.013)	(0.06)	(0.21)
Post-"Cash" ΔSlope	-0.102**	0.038**	-2.21**	0.14
	(0.002)	(0.010)	(0.05)	(0.31)
Post-Sweeps, Δ Converg.	0.476	0.476	-0.557	-0.557
	(0.367)	(0.366)	(0.593)	(0.598)
Cash Period, Δ Converg.	0.038	0.038	0.134	0.134
	(0.113)	(0.108)	(0.113)	(0.114)
Post-Cash, Δ Converg.	0.314	0.314	0.132	0.132
	(0.108)	(0.101)	(0.090)	(0.090)
Quintic in Time?	no	yes	no	yes
Trust Effects?	yes	yes	yes	yes
R-Squared	0.8499	0.8525	0.7857	0.7894
N	29605	29605	29605	29605

Statistical significance is denoted with the system: * 5%, ** 1%. Standard errors are clustered at the trust level.

Table 5: Summary Statistics for Non-Incentivized Performance Measures

Panel A: Entire Sample

Variable	Mean	Std. Dev.	N
% Patients Treated < 2 Hours	56.73%	10.92%	1979
"Mean" Patient Wait Time	2.02	0.37	1979
% Patients Admitted to Hospital	18.53%	5.37%	2008
% Patients on Return Visit	5.12%	3.53%	2002

Panel B: Type 1 Performance, Period by Period

Period	Variable	Mean	Std. Dev.	N
Pre-Sweeps	% Patients Treated < 2 Hours	49.27%	12.19%	288
	"Mean" Patient Wait Time	2.41	0.45	288
	% Patients Admitted to Hospital	18.17%	4.96%	471
	% Patients on Return Visit	6.40%	4.34%	471
Middle	% Patients Treated < 2 Hours	55.04%	11.04%	458
	"Mean" Patient Wait Time	2.12	0.35	458
	% Patients Admitted to Hospital	17.79%	5.20%	458
	% Patients on Return Visit	5.56%	3.40%	457
Cash Incentive	% Patients Treated < 2 Hours	57.61%	9.73%	154
	"Mean" Patient Wait Time	2.01	0.29	154
	% Patients Admitted to Hospital	19.40%	5.69%	154
	% Patients on Return Visit	5.02%	2.76%	149
Post-Cash	% Patients Treated < 2 Hours	59.31%	9.56%	1079
	"Mean" Patient Wait Time	1.88	0.26	1079
	% Patients Admitted to Hospital	18.94%	5.55%	925
	% Patients on Return Visit	4.26%	2.97%	925

Table 6: Testing for Classical Effort Substitution

	Dependent Variable:							
Explanatory	*Follow Up Visits*		*Hospital Admits*		*% Waits > 2 Hours*		*"Mean" Wait Time*	
Variables:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Performance	0.013	-0.018	-0.030	-0.046	-0.662**	-0.659**	-3.207**	-3.186**
	(0.169)	(0.037)	(0.021)	(0.032)	(0.067)	(0.088)	(0.180)	(0.236)
*Performance*Incentives*	-	0.034	-	0.018	-	-0.004	-	-0.028
		(0.036)		(0.026)		(0.071)		(0.202)
ΔPerformance	-0.016	-0.046*	0.003	-0.018	-0.105	-0.345**	-0.527**	-1.369**
	(0.009)	(0.022)	(0.014)	(0.034)	(0.061)	(0.070)	(0.198)	(0.196)
*ΔPerformance*Incentives*	-	0.037	-	0.025	-	0.275**	-	1.053**
		(0.023)		(0.039)		(0.085)		(0.247)
Log Performance	0.003	0.000	-0.002	-0.001	-0.050**	-0.044**	-0.221**	-0.184**
	(0.002)	(0.003)	(0.003)	(0.002)	(0.006)	(0.007)	(0.018)	(0.020)
Log Performance	-	0.004	-	-0.001	-	-0.012*	-	-0.067
**Incentives*		(0.003)		(0.002)		(0.006)		(0.018)
ΔLog Performance	-0.001	-0.004**	-0.002	0.000	-0.028**	-0.029**	-0.107**	-0.102**
	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)	(0.004)	(0.008)	(0.012)
ΔLog Performance	-	0.005**	-	-0.002	-	0.002	-	-0.010
**Incentives*		(0.002)		(0.003)		(0.006)		(0.017)
Trust Effects?	yes	yes	yes	yes	yes	yes	yes	yes
Quarter Effects?	yes	yes	yes	yes	yes	yes	yes	yes
N	2002 / 1841		2008 / 1847		1979 /1979		1979 /1979	

Statistical significance is denoted with the system: * 5%, ** 1%. Standard errors are clustered at the trust level. The number of observations records two sample sizes: that for the level regressions, and that for the difference regressions, respectively. The dependent variables, when percentages, are scaled to range from 0 to 100. Performance as an independent variable is scaled from 0 to 1.