EVALUATING THE IMPACT OF RESULTS-BASED FINANCING ON HEALTH WORKER PERFORMANCE

Theory, Tools and Variables to Inform an Impact Evaluation

Christophe Lemière, Gaute Torsvik, Ottar Mæstad, Christopher H. Herbst, and Kenneth L. Leonard

January 2013
EVALUATING THE IMPACT OF RESULTS-BASED FINANCING ON HEALTH WORKER PERFORMANCE:

Theory, tools and variables to inform an impact evaluation

Christophe Lemière, Gaute Torsvik, Ottar Mæstad, Christopher H. Herbst, and Kenneth L. Leonard

January 2013
Health, Nutrition and Population (HNP) Discussion Paper

This series is produced by the Health, Nutrition, and Population (HNP) Family of the World Bank's Human Development Network (HDN). The papers in this series aim to provide a vehicle for publishing preliminary results on HNP topics to encourage discussion and debate. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors and should not be attributed in any manner to the World Bank, to its affiliated organizations or to members of its Board of Executive Directors or the countries they represent. Citation and the use of material presented in this series should take into account this provisional character.

For information regarding the HNP Discussion Paper Series, please contact the Editor, Martin Lutalo, at mlutalo@worldbank.org or Erika Yanick at eyanick@worldbank.org

© 2013 The International Bank for Reconstruction and Development / The World Bank
1818 H Street, NW Washington, DC 20433
Health, Nutrition and Population (HNP) Discussion Paper

Evaluating the Impact of Results-Based Financing on Health Worker Performance

Christophe Lemièrea, Gaute Torsvikb, Ottar Mæstadc, Christopher H. Herbstd and Kenneth L. Leondre

a Health, Nutrition & Population–Global Practice, World Bank, Washington DC, United States
b Department of Economics, University of Bergen, Norway
c CHR Michelsen Institute, Bergen, Norway
d Health, Nutrition & Population–Global Practice, World Bank, Washington DC, United States
e Department of Agricultural and Resource Economics, University of Maryland, USA

This working paper was prepared as an output of the Africa Region Human Resources for Health (HRH) Program, financed by the Government of Norway. The Program supports Governments in the Africa Region to address challenges related to their health workforce.

Abstract:
In order to advance our understanding of why Results Based Financing (RBF) works or not, it is crucial that evaluations not only measure the impact of such an arrangement on final outcomes (population health), but also assess the changes in variables in the causal chain between intervention and final outcomes. Health worker performance is a key variable in this chain; it is only by changing health workers’ behaviors—their performance—that RBF can influence health outcomes. Careful assessment of impacts on health worker performance is therefore a natural and important element of any RBF impact evaluations. This paper discusses various approaches to evaluating the impact of RBF on health worker performance. The first part is a discussion of possible ways in which RBF may affect health worker behavior, based on economic theory and empirical evidence. The second part is a more practical discussion of how health worker performance and other relevant variables can be measured and how impacts can be estimated. This is followed by some practical steps that can be taken to ensure that the evaluation leads to actions that can be implemented; a brief conclusion completes the paper.

Keywords: Results Based Financing, Human Resources for Health, Health, Nutrition and Population, Behavioral Economics,

Disclaimer: The findings, interpretations and conclusions expressed in the paper are entirely those of the authors, and do not represent the views of the World Bank, its Executive Directors, or the countries they represent.

Correspondence Details: Christophe Lemiere, Clemiere@worldbank.org; and Christopher H. Herbst, Cherbst@worldbank.org, The World Bank, Washington DC
TABLE OF CONTENTS

EVALUATING THE IMPACT OF RESULTS-BASED FINANCING ON HEALTH WORKER PERFORMANCE: ................................................................. I
ACKNOWLEDGMENTS........................................................................................................ 6
PART I: INTRODUCTION .................................................................................................. 7
PART II – HOW CAN RBF AFFECT HEALTH WORKER PERFORMANCE? .... 10
   II.I Incentives and Performance: The Standard Principle Agent Model................. 10
      The Principal-Agent Model Applied to Health Care ............................................. 11
         Performance Pay, Monitoring, and Absenteeism ........................................... 12
         Empirical Evidence of Impact of Performance Pay ....................................... 13
   II.II Extension of the Standard Principle Agent Model ........................................ 13
      The Folly of Rewarding A, While Hoping for B: The Multitask Problem ....... 14
      Financial Incentives May Crowd Out Nonfinancial Motivation .................... 15
      Financial Incentives and Non-selfish Work Motivations ............................... 16
      The Role of Monitoring and Supervision ...................................................... 18
         Managerial Monitoring and Supervision ....................................................... 18
         Peer Monitoring .......................................................................................... 19
         Community Monitoring ............................................................................. 20
      Unconditional Pay May Also Increase Effort ................................................. 20
      Incentives May Increase Motivation but Impede Performance .................... 20
   II.III Summing Up ..................................................................................................... 21
PART III – MEASURING AND EXPLAINING THE EFFECTS OF RBF ON HEALTH WORKER PERFORMANCE ...................................................... 23
   III.I Measuring Health Worker Performance: What should be measured .......... 23
      Attendance/absenteeism ............................................................................... 24
      Productivity ........................................................................................................ 25
      Quality of Care ............................................................................................... 26
         Clinical quality .............................................................................................. 26
         Responsiveness ............................................................................................. 29
         Financial Issues ............................................................................................ 29
      The Design and Implementation of the RBF Scheme .................................... 30
         Linear or Nonlinear ......................................................................................... 30
         Relative vs. Absolute Performance .............................................................. 31
         Implementation Arrangements and Processes ............................................. 32
   Intrinsic Work Motivation .................................................................................. 32
   Monitoring and Supervision ............................................................................ 33
      Changes in the Level of Formal Monitoring ................................................... 34
      The Quality of Monitoring ........................................................................... 34
      Managerial Incentives for Enhanced Monitoring and Supervision ............ 34
      The Nature of Supervision and Feedback ...................................................... 34
      Observability and Group Size ..................................................................... 34
      Community Awareness About Performance Levels .................................... 34
   Opportunity Costs ............................................................................................... 35
      Outside Income Opportunities ................................................................... 35
The Know-Do Gap ..................................................................................................................... 35
Workload .............................................................................................................................. 36
III.III Measuring the Costs of RBF ....................................................................................... 36
Administrative Costs ............................................................................................................. 36
Salary Top-Ups (Risk Costs) ............................................................................................... 37
PART IV – SOME PRACTICAL STEPS TO EVALUATE IMPACT OF RBF ON
PERFORMANCE ..................................................................................................................... 38
IV.I Four Key Steps to Consider .......................................................................................... 38
    Designing of the Scheme ................................................................................................. 38
Identifying the Dependant Variables .................................................................................... 39
Selecting the Relevant Comparison Interventions ................................................................. 39
Selecting the Tools for Measuring Performance ................................................................. 40
    Using Non-experimental Methods to Assess the Influence of Mediating Factors. 41
PART V – CONCLUSION ....................................................................................................... 42
REFERENCES ......................................................................................................................... 43
ACKNOWLEDGMENTS

This working paper was jointly prepared by Christophe Lemière, Gaute Torsvik, Ottar Maestad, Christopher H. Herbst, and Kenneth L. Leonard. The authors are grateful to Damien De Walque (DEC-HD) and Sophie de Witter (Edinburgh University) who provided valuable input and reviewed the paper. The authors are grateful to the World Bank for publishing this report as an HNP Discussion Paper.
PART I: INTRODUCTION

The quality and utilization of health services is low in many developing countries, leading to unnecessary morbidity and premature deaths (Black, Morris, and Bryce 2003; Jones et al. 2003). In most countries in Sub-Saharan Africa, less than 50 percent of births take place with a skilled birth attendant. Higher utilization of delivery and newborn health services, in combination with improved quality of service could substantially reduce maternal and child mortality and morbidity (Montagu et al. 2011).

Results-based financing (RBF) schemes are currently being introduced in several low-income countries in an attempt to improve the utilization and quality of health services. To let resources follow results is a key element of the “new public management” approach, and variants of RBF schemes have been implemented in the health sectors of many developed countries. Their implementation in low-income countries, however, is still in its infancy. A rigorous impact evaluation of the RBF scheme in Rwanda demonstrates that RBF can have strong positive impacts (Basinga et al. 2011), but our understanding of why RBF has such positive effects in Rwanda is still very limited, and thus it is unclear whether such schemes could have a similar potential in other settings.

In order to advance our understanding of why RBF works or not, it is crucial that evaluations not only measure the impact of such an arrangement on final outcomes (population health), but also assess the changes in variables in the causal chain between intervention and final outcomes. Health worker performance is a key variable in this chain; it is only by changing health workers’ behaviors—that RBF can influence health outcomes. Careful assessment of impacts on health worker performance is therefore a natural and important element of any RBF impact evaluations.

This paper discusses various approaches to evaluating the impact of RBF on health worker performance. The first part is a discussion of possible ways in which RBF may affect health worker behavior, based on economic theory and empirical evidence. The second part is a more practical discussion of how health worker performance and other relevant variables can be measured and how impacts can be estimated. This is followed by some practical steps that can be taken to ensure that the evaluation leads to actions that can be implemented; a brief conclusion completes the paper.

Explanations for the low volume and poor quality of health services in low-income settings usually belong to one of the following categories: (1) low demand—the result of limited access and high user costs, or of insufficient education; (2) insufficient provider capacity—the result of few health workers, limited knowledge, or an insufficient supply of equipment and drugs.

There is an increasing awareness, however, that insufficient motivation of health workers also plays a key role. Direct evidence of this lack of provider motivation is the significant gap between what health workers are capable of doing and what they actually do in practice (Das and Hammer 2007; Das, Hammer, and Leonard 2008; Leonard Masutu, and Vialou 2007; Maestad, Torsvik, and Aakvik 2010), also known as the know-do gap. Another indication of low motivation is found in high rates of absenteeism at health facilities.

1 Basinga et al. compared two groups of districts receiving similar amounts of money, but one group was receiving this funding through RBF while another group received funding through the traditional input-based mechanism. In the RBF group, utilization of maternal and children health services increased sharply, sometimes up to 60 percent.
In a recent study from India, Muralidharan et al. (2011) find that as much as 43 percent of the doctors were absent from work (see also Chaudhury and Hammer 2007).

A basic presumption for the expansion of RBF schemes in health care in low-income countries is that enhanced motivation is a key to improved provider performance. The theoretical basis for the motivation effect of performance pay is straightforward. Since most people welcome money, making their earnings contingent on the results delivered will induce them to work harder and smarter to produce more of the output that is rewarded. Performance pay can motivate health workers to be more present (that is, less absent from their post), to increase service quality, and to work more actively to enhance the utilization of the services they provide. They may also be persuaded to make greater efforts to obtain adequate equipment and medicines to enhance their capacity to deliver quality health care. Yet another positive effect on performance in the long run may arise from a change in the composition of the workforce: performance pay may attract and retain workers who produce at optimal levels when they are rewarded financially. In the literature, this is called the selection effect of financial incentives (Lazear 1986).

Despite its simple logic and its widespread use in the health sector, critics of RBF in health care remain. A key objection is that the perceived motivation and selection effects of performance pay are too simplistic in the health care setting. This objection is not without support in the literature on financial incentives, which point at three problems that may arise when financial incentives are used in sectors where workers have complex and potentially rewarding jobs.

First, if jobs are multifaceted, performance pay may reduce the attention workers pay to tasks that are not rewarded. This is known as the multitask problem. This problem may arise when rewards are provided for only a subset of the tasks needed to do a good job—for instance, when workers are rewarded for the number of patients treated and not for the quality of the service they provide.

Second, performance pay may negatively affect other, nonmonetary motivations that workers have for exerting effort in their jobs. This is known as the crowding-out problem. Performance pay may undermine the intrinsic motivations driving workers to act in the interest of their employers (or their patients), and enhanced focus on financial rewards may disrupt other extrinsic motivations workers have for doing a good job (for example, to attain status, pride, respect, and glory).

Third, high-powered incentives may create “too much” motivation or expose workers to too much risk. When the stakes are large and tied to events that are uncertain, workers may make more mistakes because of the stress and strain that come with the incentives or they may be unwilling to accept the fluctuation in pay that would result.

Another issue is that switching from input- to output-based financing may involve other profound changes in the governance of the health sector. RBF often entails changes in the level and structure of monitoring and accountability. When compensation is linked to output, those providing incentives are encouraged to define and monitor performance. When a group of workers (a health facility) receives resources based on their collective performance, this may induce more peer monitoring. These changes in supervision and monitoring will affect the impact of financial incentives, but they may also have a separate impact on the workers’ performance.

Since the impacts of RBF in health care are not yet well known and the ideology supporting its premises is contentious, a careful empirical assessment of incentive programs is essential. This paper proposes an analytical framework for impact evaluation of RBF in the health sector. This framework is grounded in the
literature on financial incentives and program evaluation as well as in practical approaches to measuring health worker performance and those variables that affect performance.
PART II – HOW CAN RBF AFFECT HEALTH WORKER PERFORMANCE?

The performance of health workers can be affected by several different features of the health system. In the context of developing countries, improving performance is a crucial element in the improvement of health and service delivery outcomes. In this section we focus on the impact of RBF on health worker performance, building on the principle agent model as it applies in general and then considering RBF as it applies to the health care field specifically.

II.I Incentives and Performance: The Standard Principle Agent Model

The principal-agent model within economic theory is helpful in analyzing the impact of financial incentives used to enhance performance of service providers who are not governed by market forces. In the simplest arrangement, one organization (the principal) hires one agent (another person or organization) to do a job. The standard model has two key elements: (1) the interests of the principal and the agent are not perfectly aligned (for example, health workers do not have interests or goals that are congruent with the employer’s goals), and (2) the principal cannot link payment directly to the actions taken by the agent (that is, the principal or one doing the hiring cannot link payment directly to the effort expended by the agent or worker but can condition payment on some output measure that correlates with that “effort.”

The principal-agent model is used to study how financial incentives can improve workers’ efforts and thereby further the goals of the organization. Note that, because of its focus on monetary incentives, the standard model disregards nonfinancial motivations.

To give a brief recount of the basic model, consider the case where one unit of effort ($e$) produces one unit of output ($y$) worth benefits $b > 0$ for the organization. For effort below a certain level, say $e^0$, the worker’s marginal costs of increased effort is zero (or negative), implying that the worker would be willing to provide effort up to the level $e^0$ without any financial compensation. (It could be the case that the worker earns a fixed salary and is willing to provide $e^0$ without any further expectation of compensation.) For workers with nonmonetary motivations for exerting effort, $e^0$ will be greater than zero. In the standard principal-agent model, the effort that is exerted without any financial incentives—($e^0$)—is assumed to be a constant, not affected by the design of incentive schemes. An examination of this assumption follows.

To increase effort beyond $e^0$ will be costly for the agent. It is reasonable to assume that the agent’s marginal costs of effort $c’(e)$ are increasing for $e > e^0$ (see Figure 1). But as long as the agent’s marginal costs are lower than the value to the organization of increased effort ($b$), increased effort creates a surplus ($S$ in Figure 1) beyond the increase to the worker’s compensation for his higher effort.

To induce agents to work harder (that is, to convince them to exert effort above $e^0$ and thus to produce more than $y^0$), the principal must offer compensation that varies with the result that is delivered. The principal must offer some type of performance pay. This can be implemented in many different ways. One option is to offer a piece rate, and another approach is to pay two different levels of salary, conditional on

---

2 The term effort is here used broadly to include all activities that impact on the outcome that the employer cares about.
output levels, making sure that it is in the agent’s interest to choose a high effort/high output alternative. The latter approach is inspired by the same reasoning as Shapiro and Stiglitz’s efficiency wage model; see Shapiro and Stiglitz 1984. Note that these schemes do not rely on changes in workers’ total pay—by lowering the fixed salary, the same level of $e$ (and associated benefits) can be achieved without increasing total pay. The key point is to make pay depend on effort.

**Figure 1: Incentives and effort**

![ marginal costs and benefits of effort ]

Source: authors

Figure 1 also illustrates how an organization that offers performance pay will exert a pull on productive workers. The dotted curve to the right of the solid one is the marginal cost curve of a person who is highly skilled at undertaking the task that is rewarded and therefore has lower marginal costs of effort. For simplicity, it is assumed that this person has the same nonfinancial motivation as the other one; he or she will exert effort $e^0$ in the absence of financial incentives. We observe that, for any level of effort above $e^0$, the generated surplus will be larger for the more productive worker. As long as part of the additional surplus is retained by the worker, he will be better off than less productive workers within the same organization. Hence, the incentive scheme will pull more productive workers toward the organization.

**The Principal-Agent Model Applied to Health Care**

In health care the output $(y)$ can be measured as the number of people that utilize the health care services $(n)$ times the average quality $(q)$ of the services they get: $y = nq$. Utilization depends on the (perceived) quality of the services that are provided at the clinic, on the costs of access $(c)$ (that is, cost of transport, fee for service, etc.), and on cultural practices and social norms $(s)$: $n(q,c,s)$. 

11
The quality of health care increases in the skills of the health worker \((a)\) and in the effort \((e)\) he or she exerts. Quality also depends on a number of enabling factors \((x)\)—for example, equipment and supplies—that are (partly) outside the control of the provider: \(q(e, a, x)\). Recall that effort is a broad concept that includes any activity workers undertake to improve the quality of the health services; a health worker exerts effort if he or she is present more frequently at the clinic (that is, reduced absence) and if, while at the clinic, he or she exerts effort to perform proper history taking, examination, treatment, and tutoring of patients. Health workers also exert effort if they take actions to make sure they have sufficient equipment and drugs in stock (although in many cases access to these inputs is beyond the control of frontline service providers).

The relationship between health workers’ effort and output \((y = q(e, a, x) \cdot n(q(e, a, x), c, s))\) is more complicated in health care than in the example shown in Figure 1 above, but the same basic logic applies: health workers can work harder to enhance the utilization and quality of health care and thereby produce more health care. It is also realistic that the cost of effort is as illustrated in Figure 1: health care workers have (to varying degrees) an intrinsic desire to work, but beyond a certain level it is costly for them to work harder. The personal costs of exerting extra effort are, however, lower than the value of improved health outcomes—at least up to a certain level. If the health workers’ financial rewards increase with an increase in the utilization and quality of health care (in output \(y)\) this will (1) motivate agents to enhance their work effort (beyond \(e^0)\) and (2) attract more able health workers, because output increases with a higher level of skills, and therefore better-abled health workers will automatically earn higher payment.

According to the argument above, linking pay to performance will improve health outcomes over and above the improvement generated by an unconditional transfer of resources of the same magnitude. An unconditional transfer of resources may improve the availability of drugs and equipment (increase \(x)\) more generally), which will enhance the quality and utilization of health care. Unconditional resources will, however, not—in the model above—directly motivate workers to make an extra effort to improve the quality and outreach of their health care services. To enhance motivation and effort above the intrinsic level, resources must be provided through a performance-pay mechanism. In the next section we consider health workers who are motivated by a desire to provide reciprocity; such workers may indeed exert costly effort even if the pay supplement is not related to their performance.

The discussion above illustrates the motivation and selection effect of performance pay. The most fundamental prediction is that, after a switch from paying a fixed salary to a scheme with financial performance incentives, we expect to observe higher output. Note again that the motivation and selection effects of financial incentives do not require that income be the only reason why people work. It requires only that money is one driver of effort, and—crucially—that appealing to this motivation (of greater pay) does not strongly undermine other work motivations (an assumption we will return to later).

**Performance Pay, Monitoring, and Absenteeism**

An RBF scheme may also reduce the problem of absenteeism. Consider a health worker who earns an income \(F\) if he spends all his working time at an unofficial job, moonlighting in the private sector for example. Suppose the health worker is offered a bonus \(B\) at his official job, and that the likelihood of receiving the pay supplement increases in his or her attendance rate \((e)\). To simplify assume that the worker’s base salary is not affected by absenteeism—health workers do not lose their jobs even if they are often absent—but the pay supplement may be withheld if the health worker is found to be absent from work on a regular basis. The probability that absence is detected \((p)\) increases in monitoring \((m)\)
and decreases in the worker's attendance rate ($e$). The expected income ($I$) for a health worker is given by:

$$I(e) = w + p(e;m)B + (1-e)F$$

Suppose the health worker chooses his attendance rate to maximize expected income, and that an attendance rate somewhere between 0 and 100 percent is chosen (an interior solution). The optimal choice of attendance ($e^*$) is then characterized by the first order condition

$$I'(e) = 0 \Rightarrow p'(e^*;m)B + (1-e^*)F = 0$$

It is easy to understand (and to show) that the chosen attendance rate ($e^*$) increases in the bonus $B$ and in monitoring $m$. When workers have selfish preferences (workers choose attendance in order to maximize their own income), monitoring works only because the worker knows that he may lose some of his income if inadequate behavior (too much absence) is detected. Hence, a pay supplement must be performance related in order to have an effect on the behavior of income maximizing workers. The argument here is parallel to the one used by Stigler and Becker (1974) to study corruption; see also Besley and McLaren (1993).

**Empirical Evidence of Impact of Performance Pay**

One of the most cited empirical studies of the impacts of performance pay is Lazear's (2000) case study of the Safelite Glass Company. Safelite switched from paying hourly wages to using an individual piece rate scheme. Lazear estimates that the implementation of variable pay increased productivity by 44 percent and that roughly half of the gross increase in productivity came from improved motivation; the other half of the gain was the result of the selection and retention of high-productivity agents (agents who can produce high quantities at low costs).

The Safelite study is unique. Impact evaluators rarely have access to information that allows a separation of the motivation and selection effect of performance pay. There are, however, several other case studies that assess the overall impact of financial incentives on productivity. Positive output effects of pay-for-performance have been documented in tree logging (Haley 2003), tree planting (Paarsch and Shearer 1999), and fruit picking (Bandiera, Barankay, and Rasul 2009, for example. Prendergast (1999) and Lazear and Oyer (2009) provide a more comprehensive list of studies of performance pay.

The simple theory of incentives and the empirical case studies supporting this model seem to make a case for using financial incentives to enhance performance in the health sector. This conclusion is, however, premature. By acknowledging the complex nature of health service provision, taking into account the fact that the level of intrinsic motivation is not necessarily independent of financial incentives, studying the potential impact of RBF on governance structures, and taking into account the significance of social or other-regarding preferences, the potential impacts of RBF are seen to be both more complex and more ambiguous than RBF in simpler contexts. We now turn to a discussion of these issues.

**II.II Extension of the Standard Principle Agent Model**

A characteristic feature of the case studies referred to above is that they examine pay-for-performance in organizations where workers undertake relatively simple work tasks, where it is easy to measure the output, and where there are few (if any) professional guidelines specifying how workers ought to do their
jobs. Health workers operate in a totally different work environment. Additional aspects of performance pay must be considered when workers have complex, challenging, professional jobs.

**The Folly of Rewarding A, While Hoping for B: The Multitask Problem**

Whereas the employer’s ultimate concern may be with the level of \( y \) (where \( y = nq \)), this variable may be difficult to reward because of difficulties in measuring quality \( q \). The only factor that can be rewarded may therefore be \( n \). However, paying for \( n \) when what you really want is more \( y \) can backfire. It is easy to understand the basis of this problem: If an organization starts to reward the number of units produced, employers may compromise on the quality of production. In the long run, perhaps even in the short run, lower-quality output may be detrimental to the organization and it is possible that \( y \) will remain constant, or even fall, despite increased \( n \). The employer may then prefer to pay a fixed wage rather than instituting an imperfect scheme for performance pay.

It is not only the quality of output that may suffer under piece-rate financing. The more general worry is that monetary incentives may pull workers’ attention away from work tasks that are not rewarded. The multitask problem arises where workers are required to perform multiple tasks as part of their job. Suppose that the agent performs two activities, \( e_1 \) and \( e_2 \), and that the output the organization cares about is given by \( y = e_1 e_2 \). The multi task problem arises when the principal cannot link pay to \( y \). At best he can condition pay on an incomplete output measure. To make an extreme case, assume that performance pay can be based only on the measure \( g = e_1 \). Now, if the principal starts to give a bonus that increases in response to observed \( g \), this will enhance the effort the agent assigns to task 1 (\( e_1 \) increases). The problem is that greater effort expended on task 1 may lead the agent to reduce effort on task 2. Such a negative spillover will occur, for example, if the agent’s effort costs increases in aggregate effort (\( e_1 + e_2 \)). The overall effect of incentive pay may then become negative.

In a classic article, Kerr (1975) provides several examples of incentive programs where the principal gets exactly what he pays for, which turns out to be something other than what he was hoping for. Holmstrom and Milgrom (1991) and Baker (1992) are the first of several articles that examine multitask problems in formal principal-agent models. There are now many careful empirical works documenting the relevance of the multi-task problem; see, for example, Asch (1990) and Prendergast (1999) for a review.

Those opposing pay-for-performance mechanisms in the public sector often appeal to the multitask problem. Financial incentives do not work well in schools and hospitals because it is hard to measure “education” and “health care,” and when teachers and doctors get rewards for reaching certain targets, they will direct their attention toward the targets; this may eventually not promote learning and care taking. If teachers receive a bonus based on their students’ test scores, they will be tempted to teach to the test, or worse, to cheat and inflate the score that their students get on tests (Jacob and Levitt 2003). If health personnel get paid based on the number of cancer screenings they perform or the number of births they assist, they may screen too many patients to be careful with each one and may discourage women with severe birth complications from coming to the clinic because those births would take too long, or they may not use sufficient time and energy to educate women on how to prevent unwanted pregnancies. Another less vivid effect of pay-for-performance in health care may be that health personnel do not spend enough time on history taking, examination, and education of patients.

It should be noted that rewarding quantity does not necessarily give workers an incentive to reduce quality. In some cases, high quality may contribute to higher output quantity. This is a valid point in health care where the number of patients seeking assistance is likely to increase in response to an increase in
the quality of the services provided \( n \) increases in response to \( q \), to use the notation introduced above). In this case, even if doctors are paid according to how many patients they treat for a certain condition or illness, they have an incentive to increase the quality of their services in order to bring more patients to the clinic; see Ma (1994) and Maestad and Torsvik (2011). A presumption here is that patients can perceive the true quality of health care, and that true quality matters for their decision to utilize the services.

The graveness of the multi-task problem depends both on the character of the jobs and on the character of the workers. If workers are sufficiently unscrupulous, they may not only adjust their behavior to the incentives offered, but they may also actively use time and energy to bend and misuse the scheme in order to make the most of it. Such gaming of incentive schemes can be a problem in health care: diagnoses can be miscoded and patients may be selected not based on the need for treatment but rather on the extent to which they add to the performance indicators that are rewarded (cherry picking) (Rosenthal and Frank 2006).

**Financial Incentives May Crowd Out Nonfinancial Motivation**

A long-standing issue in psychology concerns how financial incentives interact with other work motivations (Deci 1971). This question does not arise in the standard principal-agent model because the model centers around the impact of financial incentives on worker effort. The key assumption in this model is the notion that nonfinancial motivation is constant—that is, that motivation is not affected by the power of financial incentives and the structure of pay.\(^3\)

Over the last two decades, an influential body of laboratory experiments has shown that the behavior of many individuals is strongly and systematically influenced by other-regarding social preferences. Economists have used a set of simple games to test the self-interest model. The general picture that has emerged from these experiments is that individuals are much more generous (willing to share their money with strangers in anonymous one-shot encounters), trusting (willing to trust strangers who have an economic incentive to cheat), and fairness oriented (willing to sacrifice resources in order to adhere to a fairness principle) than the economic self-interest hypothesis predicts (see Camerer 2003 for an extensive review of the experiments). The experiments have also revealed a dark side of human motivation: some of those who participate in the experiments are status oriented, spiteful and envious, and willing to punish others even if doing so lowers their own income (see Fehr et al. 2008). We refer to this view of individual behavior as the **behavioral economic view** (as opposed to the standard principal-agent view).

These lab experiments also show that there is a lot of heterogeneity in behavior: individuals differ with respect to what extent they possess (or at least display) social preferences. In the dictator game (A is given 10 dollars and must anonymously decide how much to share with another unknown person B), the distribution of donations will typically be bimodal with one peak at 50 percent and one at 0 percent and with an average donation around 30 percent. Individuals motivated only by extrinsic monetary rewards will give 0 and individuals who follow fairness norms will typically give 50 percent; clearly both types exist in significant numbers within the same population. The same pattern is found in many other sharing games or experiments.

---

\(^3\) The fact that nonfinancial motivation is entirely dropped from the model may explain why it took economists so long to realize that there may be interesting interaction effects between nonfinancial motivations and financial incentives.
Some of the lab experiments are framed as employment relationships, and many of the subjects who are assigned the role of agents do not respond to monetary incentives in the way that the principal-agent model predicts. The reason for this discrepancy appears to be that the “workers” have concerns other than maximizing their own income; see Fehr and Schmidt (2006) and Charness and Kuhn (2011) for an overview of this literature. The results from these experiments have led economists to pay more attention to how financial incentives may affect other work motivations.

**Financial Incentives and Non-selfish Work Motivations**

Extending employers’ motivations beyond material self-interest has two important implications for the use of incentives in firms. First, it enlarges the set of instruments a principal can use to motivate agents. A worker who is inclined to reciprocate generosity (or spite), who wants to adhere to fairness norms, who wants to do good for others (altruism), who craves recognition and respect, and who feels guilt and shame if he or she performs below expectations can be induced to work harder and smarter in many more ways than can workers who have their eyes fixed only on their own net incomes (pay minus effort-costs). Put differently, workers who care for many outcomes of their own performance, not only the monetary outcome, can be rewarded (incentivized) in many different ways. It can, for example, be possible for an employer to increase nonfinancial motivation (to increase $e^0$) by increasing the respect, rather than the money, paid to workers (Ellingsen and Johannesson 2007).

These ideas are especially important in health care because professional health workers are supposed to adhere to higher standards of social behavior and, even when well paid, are not supposed to be driven by extrinsic motivations. A professional’s dedication to group goals and values can be extended to include a sense of service to the greater good (Akerlof and Kranton 2000, 2005; Cullen 1978; Freidson 2001). Indeed Leonard and Masatu (2010) show that, in Tanzania, doctors who appear to be professional provide higher-quality care, even for aspects of care that are difficult to monitor by their employers and for which patients are unlikely to pay extra. In general, we should expect health workers to be more intrinsically motivated than the average individual.

Second, the effectiveness of various pay-for-performance mechanisms becomes more involved and uncertain when agents have a variety of reasons for exerting work effort and do not rely only on self-regarding reason. Workers who feel uncomfortable when their colleagues are lagging behind will work harder under a piece rate than they will if they are competing against their own colleagues in a tournament (Bandiera, Barankay, and Rasul 2005). Workers who dislike income inequality may produce more under a team bonus than they would under an individual bonus (Engelmaier and Wambach 2010).

In discussing how nonfinancial motivations are affected by financial incentives, it is instructive to distinguish between intrinsic and extrinsic motivations. *Intrinsic motivation* can be defined as “the propensity to perform a given behavior in absence of external rewards or reinforcement” (Donovan 2001). According to this definition, workers who find some of their work activities fulfilling, meaningful, and joyful in their own right are intrinsically motivated to work. This definition includes a wide range of reasons why workers perform: “I work hard because it is interesting to solve the kind of puzzles I encounter at work”; “I work hard because my employer treats me well and I think it is appropriate to reciprocate his or her

---

4 Historically, in the United Kingdom, though it was acceptable to be paid, it was not acceptable for a physician to care too much about being paid. “Under an ancient legal fiction, English law regarded the services of physicians as wholly philanthropic. While surgeons and apothecaries could sue for their fees, physicians could not” (Starr 1982, 61–62).
kindness”; “I work hard because I believe it is important to relieve pain and prevent premature deaths.” If, however, the reason for working hard is to earn status, respect, and glory, the motivation is grounded in a craving for external rewards and is better described as nonfinancial extrinsic motivation. The distinction between intrinsic rewards and nonfinancial extrinsic rewards is not always clear in medicine—what is the difference between caring for your patients and enjoying the respect you earn from being seen to care for your patients? In this review of RBF, it is clear that externally driven financial rewards are neither intrinsic nor nonfinancial extrinsic, and it has been argued that financial incentives may supersede both of these alternative motivations for performance.

Results from lab experiments—and from field experiments on various nonmarket activities—indicate that this crowding-out problem is not merely a theoretical possibility. Titmuss (1970) is an early warning of the “crowding-out” likelihood, albeit in a different context. He argued that rewarding blood donations with financial incentives may lead to a decrease in donations, but he did not provide significant evidence for this hypothesis. Later Johannesson (2008) performed controlled experiments with “paying for blood.” He found that paying for blood had a motivational crowding-out effect for women but not for men. Another well-known field experiment that tests the crowding-out hypothesis is presented in Gneezy and Ruchtichini (2000a). In collaboration with a population of 10 day-care centers in Haifa, they randomly selected six centers and introduced a fine for parents who came late to pick up their children. Surprisingly, the fine caused an increase in late pick-ups. The day-care study provided a first rigorous evidence of a crowding-out effect. There are now several studies—mostly lab experiments—that examine the crowding-out hypothesis. Bowles (1998) provides a review of this literature.

Although the day-care experiment shows that introducing a financial incentive to encourage an activity (timely pick up) may in fact have a discouraging effect (more late pick up), it is not obvious why this could happen. One possibility—which also may apply to the blood donation case—is that financial incentives, when used to regulate behavior, may change subjects’ perception of what is appropriate behavior in this situation: By using a fine to deter late-coming or by paying money to induce blood donations, these acts are moved from the moral domain into the market domain. There are, however, several other mechanisms that may—in other situations—explain why financial incentives may lower performance. Concerns about social and self-esteem, for example, may make some individuals less willing to perform an activity when they are offered monetary incentives, as suggested by Bénabou and Tirole (2006) and Ellingsen and Johannesson (2008).

If financial incentives undermine other work motivations, it is possible that the impact of performance pay is negative for low-powered incentives but positive for high-powered incentives. Small financial incentives may drive out an intrinsic motivation to perform but will on their own not have any substantial positive effect on effort; hence the overall impact on performance may be negative. If, on the other hand, financial incentives are more powerful, they will eventually drive performance above the initial level. This is exactly what Gneezy and Rustichini (2000b) find in their experimental study of how the level of financial incentives affects behavior (hence the telling title of their paper: “Pay Enough or Don’t Pay At All”).

Although less often discussed, it is of course possible that performance pay can complement and crowd-in other work motivation. Introducing performance pay, for example, may be seen as a generous or fair gesture that generates a desire to perform (for example, as an act of reciprocity), which would add to the direct impact of the financial incentives.
The Role of Monitoring and Supervision

The use of pay-for performance may increase different levels of monitoring and supervision that are performed on several levels: by managers, by peers, and by the community.

Managerial Monitoring and Supervision

Monitoring and supervision may increase under RBF schemes because the provider of incentives needs to keep account of performance. Moreover, since RBF schemes tend to be team-based bonuses where line managers also receive a share, managers may have an additional incentive to increase the level of monitoring and supervision to enhance performance.

In a standard principal-agent model, nonfinancial work motivation is independent of the financing system workers are offered, and in this case more monitoring will always enhance the agents’ effort (for example, reduce absenteeism). The reason is of course that, with more monitoring, the link between effort (performance) and pay becomes tighter, so it becomes more likely that high effort is rewarded and low effort penalized.

When workers have other-regarding concerns, the link between monitoring and motivation is no longer that simple. Nagin et al. (2002) find that increased monitoring does not enhance the efforts of all types of workers. Their interpretation is that some workers are “trustworthy” at the outset and therefore immune to increased monitoring (see also Falk and Kosfeld 2004). The impact of monitoring on performance may also be negative. One of the basic motivations that seems to drive a lot of the nonselfish behavior in the lab is reciprocity and fairness—that is, individuals want to behave generously to someone who is generous and fair, and spiteful toward someone who is unfair and unpleasant. Workers may therefore respond negatively to an increase in monitoring if they interpret this as a signal of distrust that limits their autonomy. Falk (2009) finds evidence for a negative effect of monitoring in an experiment, calling it “the hidden cost of control.” Several studies cited in Oldham and Cummings (1996) found that supervision that was perceived as controlling was indeed associated with lower intrinsic motivation. There is some evidence that the hidden costs of control are higher when there is a higher degree of personal closeness between workers and supervisors (Barkema 1995; Dickinson and Villeval 2004).

However, when workers have social preferences, there is also a possible additional positive effect of monitoring, a hidden benefit, that may be relevant in the case we are discussing: monitoring may show that someone takes interest in one’s work, which may be taken as a sign of recognition that may on its own enhance the workers’ willingness to perform. Frey (1997) and Frey and Jegen (2000) have argued that monitoring will not reduce intrinsic motivation if it is perceived as supportive rather than controlling. The way in which feedback is communicated to workers may also affect performance, as demonstrated by Brandt and Cooper (2007). They found that motivation to perform was significantly higher in a group

---

On the basis of Deci and Ryan (1987), Oldham and Cummings (1996) provide a definition of a supportive style of supervision: “when supervisors are supportive, they show concern for employees’ feeling and needs, encourage them to voice their own concerns, provide positive, chiefly informational feedback, and facilitate employee skill development.” Conversely, “when supervisors are controlling, they closely monitor employee behavior, make decisions without employee involvement, provide feedback in a controlling manner, and generally pressure employees to think, feel, or behave in certain ways.” The notion of supportive style of supervision can be expanded so as to include “organizational trust” and procedural justice at work.
where workers had the opportunity to respond to feedback from managers (two-way communication) than in a group without such opportunities (one-way communication only).

**Peer Monitoring**

Peer pressure arises when colleagues are monitoring each other, pushing each other to become more productive. Kandel and Lazear (1992) point out that peer pressure will occur only if two conditions are met: (1) the member’s effort must affect the well-being of the rest of the team (in other words, profit-sharing is necessary, as well as some complementarity in production); and (2) team members must have the ability to affect the choices of other members.

The first condition is typically met in an RBF setting: not only is the bonus usually a team bonus that depends on the contributions of all members, but there are also complementarities in the production process because the ability of each worker to increase service provision depends on the effort of others. Whether the second condition is also met will typically depend on the nature and strength of social preferences. The ability to affect the choices of other workers through peer monitoring usually depends on the activation of feelings such as guilt or shame. Charness and Dufwenberg (2006) have argued that people tend to be “guilt averse” in the sense that they dislike letting other people down. When someone expects a lot from us, we tend to make an extra effort. Interestingly, therefore, peer pressure may have an impact on performance (through guilt) even if performance cannot be directly observed by the peers. If performance can be observed, peer monitoring may additionally affect performance through feelings of shame and/or pride.

A majority of field studies have found evidence of the effect of peer pressure. A compelling experiment was carried out by Falk and Ichino (2006). They set up two comparable groups, which were paid a fixed amount. The first group was allowed to observe the efforts of each other, while the second could not. Productivity of the first group was significantly higher, demonstrating that peer pressure was a strong motivational lever, even in the absence of any financial incentives. In another study, Mas and Moretti (2009) explored feedback among supermarket cashiers. These workers are paid a flat wage, but they can free ride in unloading the workload on to other cashiers. The authors find that when colleagues are more productive, fellow cashiers also increase their productivity, but this happens only when they can observe each other’s output. Falk and Ichino (2006) and Knez and Simester (2001) also provide evidence of positive impact from peer monitoring.

The strength of peer pressure and its impact on health worker motivation may depend on contextual variables such as group size. One hypothesis is that when the group gets larger, mutual monitoring becomes more difficult and free riding more frequent. However, many studies (Isaac and Walker 1988; Isaac, Walker, and Williams 1994; Fehr and Gächter 2000; Carpenter 2007) have not found any evidence that monitoring is more difficult when group size increases. The existence of social preferences may explain these findings: when the group size is larger, the shame effects may be stronger, which will prevent shirking. As discussed above, shaming requires that performance be observable. Therefore, large group size is expected to have a less negative effect on peer mentoring when the production process is transparent (Carpenter 2007).

---

6 Bellemare (2009) is one of the few studies that could not find a positive impact of peer pressure.
Community Monitoring

RBF mechanisms often include some community monitoring features that might have a separate impact on health worker performance. Community monitoring can take several forms: (1) community-based organizations may be subcontracted to verify accuracy of reported results (see Naimoli and Vergeer 2009 for a review); (2) community representatives may be members of committees overseeing RBF programs at a local level (as occurs in Rwanda, for example); (3) RBF results achieved by health facilities may be made public in villages and cities; (4) more generally, because RBF bonuses are proportional to the quantity and quality of services available to local communities, health workers may feel more accountable to these communities (that is, a sort of implicit “customer power” may be at play).

As with peer monitoring, community monitoring depends on the ability of communities to impact health worker performance. Despite the absence of formal mechanisms for sanctions or rewards, a number of informal mechanisms may be available. These include various kinds of social mechanism (for example, guilt/shame/recognition) as well as material incentives (for example, informal payments or practical assistance with various private matters, such as housing).

Evidence of the impact of community monitoring remains limited, but what there is nevertheless suggests that this type of intervention may have a strong positive impact on health worker behavior. Bjorkman and Svensson (2009) conducted a randomized experiment measuring the impact of community monitoring on the quantity and quality of health care in Uganda and found significant improvements in health care utilization and children’s health. An important part of that intervention was that communities received accurate information about the quality of health services. The introduction of RBF may have a similar effect by making public the results obtained by the health facility (although this is not always done systematically) However, Bjorkman and Svensson also ensured that the information was used as a basis for a dialogue between providers and community, suggesting that it may be insufficient merely to make information available to the community.

Unconditional Pay May Also Increase Effort

Above we discussed the importance of various nonfinancial motives for exerting work effort. If workers are motivated by fairness and reciprocity, they may increase their effort in response to unconditional pay raises. This is contrary to the predictions of the standard principal-agent model, where the prospect of a pay increase induces agents to improve their performance. Hence, the pay supplement must somehow be linked to performance, and performance must be monitored in order to motivate. When workers are reciprocity motivated, on the other hand, they will engage in a kind of “gift exchange”; if workers feel that their organization treats them well (increases pay) they want to repay by increasing their performance (without monitoring and without payment being related to performance). This is an important insight because unconditional pay increase is far less costly to administer than an RBF scheme and therefore needs to be considered as an alternative to RBF.

Incentives May Increase Motivation but Impede Performance

The basic premises of the simple principal-agent model builds on two auxiliary assumptions: (1) financial incentives imply greater motivation to produce the output that the organization cares for and (2) greater motivation leads to better performance. The multitask problem and the motivational crowding-out problem question the first assumption: financial incentives may enhance motivation to perform some (rewarded) activities but will reduce the motivation to do other (nonrewarded) activities; financial incentives may
undermine other work motivations and the total effect on motivation may be uncertain (the crowding-out problem).

A recent lab experiment shows that the second assumption may also be problematic (Ariely et al. 2009). The title of that study, “Large Stakes and Big Mistakes,” brings out the main message: if high performance is rewarded with high prizes, individuals may choke under the pressure of the incentives; they are highly motivated to perform, but stress, anxiety, and excessive pondering may lower their achievements. Although this finding is interesting, it is not at all clear that the mechanism is relevant for the kind of incentives we find in the health care sector of developing countries.

Perhaps more relevant is the fact that health workers will face increased risk in pay. Risk-averse health workers are worse off with an uncertain wage than they are with a fixed wage. If RBF prevents a comparable increase in fixed salary, health workers will eventually be worse off than they otherwise would be. This may discourage qualified health workers from working in the health sector.

II.III Summing Up

Introducing a pay-for-performance mechanism in health care implies that workers are given (1) a pay supplement that is (2) conditional on some performance indicators that (3) normally evoke more monitoring and supervision. All of these factors may affect the effort and ability of the health workers and hence how they perform their services. The theoretical and empirical literature discussed above demonstrates that, although the standard principal-agent model makes fairly sharp predictions with respect to the impact of pay-for-performance, the behavioral economics perspective does not.

The standard model predicts that sufficiently strong financial performance incentives will enhance the effort of health workers (and in the longer run will pull workers with higher performance skills into the health sector). This will improve the quality and quantity of health services, which will lead to better health outcomes. In this framework only the power of the monetary incentives that matters—the generosity of the payment scheme (the amount of resources the facility gets) and the fairness of the procedures that lie behind the design of the reform are irrelevant for how the workers respond to the scheme. Furthermore, monitoring will motivate if more monitoring leads to a potential loss of income.

The multitask extension of the standard model points out that health workers may bend their focus and effort toward performance indicators that are rewarded. This may improve the quantity and perhaps also the quality of the health services that are incentivized, but it may also reduce the quality and quantity of other types of health services. The overall effect of performance pay on population health is therefore uncertain unless the multitask problem is solved through balanced incentives.

If health workers possess different types of other-regarding (social) preferences and/or intrinsic work motivations, the route from financial incentives to improved health worker performance becomes much more convoluted. With these other motivations, it is possible that financial performance incentives will undermine other work motivations and hence that the overall level of motivation to deliver quality health services actually declines. But, as noted in the discussion above, the interaction between financial incentives and intrinsic motivation may also go the other way: performance pay may strengthen the

---

7 Unconditional recourses (drugs or equipment) may improve health outcomes by enabling the health workers to produce better health outcomes for the same “effort.”
intrinsic drives to work because, for example, the pay supplement is interpreted as a supporting and generous act. Another important point is that if workers have reciprocity preferences, they may exert greater effort even if the pay supplement does not vary with performance (the “gift exchange” argument). Finally, when workers are motivated by more than their personal income and effort costs, the impact of monitoring and supervision also becomes unclear. Workers may strive harder if their work is monitored (even if there are no financial incentives) because they seek social approval and it is embarrassing and unpleasant to be observed not doing an adequate job. Monitoring can, however, also destroy motivation if it is interpreted as controlling and unsupportive.

If financial performance incentives can incite so many different mechanisms, often leading to opposing outcomes, it becomes important to identify how they work in reality. Theory reveals that the impact of performance pay will depend on a number of parameters, such as the intrinsic motivation of the workers, how they perceive the intentions behind the scheme, the span of activities covered by financial incentives, and so on. It is therefore unsatisfactory to know simply whether or not a particular RBF program had a desirable impact. What we need is a theory-informed and comprehensive program assessment that can shed light also on why it is working or not. The next section explores in greater detail how to measure both the level of health worker performance and some of the parameters that are crucial for the impact of RBF on health worker behavior.
In order to consider fully the effects of RBF on health worker performance, this section first determines exactly what should be measured and the elements of that performance that need to be taken into account. The section then explores both how and why RBF has an impact on performance, and concludes with a look at how to measure its cost.

III.I Measuring Health Worker Performance: What should be measured

Figure 2 summarizes our theoretical discussion of the impact of RBF on the provision of health care services. The key variable in the causal chains between RBF and population health is health worker behavior—it is only through changes in worker behavior that RBF can influence health outcomes. The arrow pointing from health worker behavior toward population health captures the hope that improved health performance will reduce morbidity and mortality. There is also an indirect route here, as improved health worker performance may increase utilization of health care services, which may further improve population health. The association between quality of care and utilization is, however, complicated by the fact that a higher number of patients—and thus a higher workload—may force health workers to spend less time with each patient, possibly reducing the quality of care (Mæstad and Torsvik 2010).

Figure 2: Analytical framework for the possible influence of RBF on health workers performance

As displayed in Figure 2, RBF may affect health worker performance via three channels: one direct route and two indirect ones. The direct impact comes from the monetary incentives inherent in the RBF scheme. A RBF scheme makes the remuneration of health workers depend—to some extent—on the results they deliver. Being financially accountable for their performance may induce health workers to work harder and smarter to produce the results that are rewarded. In the longer run, it may also affect their capacities to
deliver rewarded results through a change in the workforce composition (selection of workers who perform better).

In addition to this direct motivational effect, monetary incentives may—perhaps in the longer run—indirectly affect health workers’ underlying work motivations (preferences). A second indirect causal mechanism is triggered through changing levels of monitoring and supervision. The use of RBF may require enhanced monitoring. In addition, it may spur various forms of additional monitoring and supervision mechanisms that may separately affect (positively or negatively) health worker behavior. In order to understand the impact of RBF schemes, it may be important to distinguish between these different causal pathways.

The following discussion focuses on how we can go about measuring health worker performance and the various intermediary variables that influence the impact of RBF on health worker performance. Finding relevant and measurable indicators of health worker behavior and performance is a challenge because work behavior is multifaceted and complex. Figure 2 illustrates the three different aspects of worker behavior that may be influenced by an RBF scheme: attendance (presence/absenteeism), productivity, and quality of care. Below we discuss ways to measure each of these performance indicators, but first we stress the importance of evaluating a sufficiently broad set of behavior indicators.

Because of the above discussed issue of multitasking, it is not enough to measure how financial incentives affect activities that are rewarded; we should also assess how the program affects work duties that are not incentivized. One problem with the evaluation of the Rwandan RBF scheme (Basinga et al. 2011), for example, is that it focuses only on those aspects of performance that were incentivized. The payment in that scheme is based on 14 service indicators, primarily associated with maternal and child health services. A possible consequence of such an incentive scheme is that health workers will pay less attention to other important patient groups (TB patients for example). Before drawing conclusions about the overall impact and desirability of the program, such potential negative side effects need to be assessed.

Note that the problems associated with multitasking may take some time to materialize. It may take some time for workers to learn how to game the system and adapt to the new incentive structure. Evaluations should therefore allow for a sufficient time lag between the implementation of the intervention and the measurement of such impacts.

**Attendance/absenteeism**

We define **absenteeism** as being away from the working station during duty hours. The main challenge with measuring absenteeism is to establish how many workers are supposed to be present at the working station at any given point in time. One way to determine this figure is to look at facility duty rosters. But rosters are not always available. The alternative is then to obtain a list of all workers at a facility and then enquire about each worker whether he or she is absent because he or she is on a different shift.

Measurement of absenteeism has to be unannounced. If the enumerators who are given the task to assess levels of absenteeism may be unable to establish the number of workers who should be present (if, for instance, no one is present when the enumerators show up or the facility is closed), this figure may need to be established through an earlier, announced visit.
Absenteeism is not necessarily illegitimate. Participation in training courses, for example, may require health workers to be absent. It is challenging to distinguish empirically between legitimate and illegitimate absenteeism because it requires enumerators to trace each individual health worker who is absent. To evaluate RBF schemes, it is interesting to know the total impact of RBF on absenteeism, whether illegitimate or not, because reduced absenteeism increases the availability of health services. It could also be interesting to know how RBF affects legitimate relative to illegitimate absenteeism, because legitimate absenteeism presumably may have some positive impacts on health service delivery that illegitimate absenteeism does not have. However, we regard this to be of secondary importance.

There may be systematic differences in the level of absenteeism over the working day, over the week, and over the seasons. An accurate measurement of absenteeism therefore ideally requires careful planning to ensure that all different timings are represented in the data (see Chaudhury et al. 2006).

**Productivity**

Another core aspect of health worker performance is the number of clients served per health worker. A central purpose of most RBF schemes is to increase the quantity of service delivery by increasing the number of clients. The number of clients served is therefore one of the main indicators used to determine the financial rewards in RBF schemes.

As discussed in Part II, the number of clients is not necessarily a neutral performance indicator, since the number of clients may be increased by reducing the provision of preventive services and possibly also by relinquishing some aspects of the quality of care to increase the number of re-attendances (although low quality also may make patients stop using the health service altogether). Therefore, ensuring that measures of outputs include the provision of preventive services and that measures of quality are also produced is essential (see below).

Output can in principle be increased either through increased productivity (a higher number of clients per health worker) or by attracting additional health workers (if there are vacant posts, increased remuneration through RBF may attract more health workers). Both causal mechanisms are worth exploring, but this discussion focuses on the more challenging aspect of higher output: how to measure changes in productivity.

Productivity is a measure of outputs relative to inputs. One challenge in productivity analysis is that health facilities are multiproduct entities that produce a number of different services (outpatient consultations, immunizations, family planning, surgeries, etc.), making it hard to construct an aggregate measure of outputs (Vujcic et al. 2009). One possible solution is to construct weights for each type of service (antenatal services, outpatient consultations, family planning consultations) based on the average time expended per consultation. However, as we discuss further below, more advanced techniques are also available.

A second challenge is to adjust for any changes in input levels. Meeting this challenge requires a measurement of aggregate inputs across many different inputs. If market prices are available, the aggregate level of inputs can be measured through an aggregate monetary measure. But market prices of inputs in health care are not necessarily readily available. In this case, we are left with an aggregation challenge similar to the one on the output side. In addition, some of the most important inputs—such as health worker time use—may be difficult to estimate precisely if we do not simultaneously measure absenteeism.
One of the techniques that may be used to deal with the aggregation challenge is Data Envelopment Analysis (DEA). DEA is a linear programming technique that constructs a productivity measure for each health facility based on data from the whole sample of facilities, without any need to define input and output weights a priori (see Cooper et al. 2006; Maestad and Mwisongo 2012). The DEA technique is very cheap in terms of data collection because it typically requires only routine administrative data (see World Bank 2008, 2009) for case studies from Senegal and Serbia.

Another approach to productivity measurement is the use of time and motion studies, where enumerators follow health workers over a long time (such as a week) to note how they spend their time. Such studies will produce measures of what health workers spend their time on, and in particular whether they spend more or less time on productive activities. However, the share of the time spent on productive activities is not necessarily an important indicator of the effect of RBF schemes; we would typically be more interested in how the productive time is spent (that is, in the quality of care). We therefore do not recommend this approach in this context.

Quality of Care

The quality of health services is usually defined as “the degree to which health services . . . increase the likelihood of desired health outcomes and are consistent with current professional knowledge” (Institute of Medicine, Washington). High quality is important not only because of its direct impact on health outcomes, but also because of its potential indirect effect through an increased utilization of health services.

In the context of RBF, we think it is particularly important to assess the following aspects of quality:

- **Clinical quality:** The extent to which health services are provided according to accepted medical standards
- **Responsiveness:** The social aspect of how people are treated and the physical environment in which they are being served
- **Financial issues:** The level of user fees and informal payments, if any

**Clinical quality**

The ideal way to measure clinical quality would be to measure the level of health improvements resulting from the health service provided. This is normally not possible to do, and standard ways of measuring clinical quality therefore typically focus on input quality and/or process quality.

The availability of inputs such as equipment and drugs should be measured as part of the RBF evaluation. Most facilities will be able to influence these variables to some extent—for example, through timely ordering of new supplies, careful maintenance, and so on—although it is also clear that these factors to some extent are beyond the control of the frontline service providers. These variables can be measured through a standard facility survey.

The measurement of process quality is normally far more involved. Process quality refers to the procedures a doctor or other health worker performs in order to arrive at a diagnosis, the treatment provided, and the education given to the client about his or her condition and what actions to take to improve it. The challenge with measuring process quality is to define a benchmark against which to measure actual performance. Each patient is unique, and a standard optimal way of going about the case is often difficult to get at.
Fortunately, some conditions require fairly standardized procedures, and these conditions are therefore natural candidates for performance assessments. For instance, antenatal visits should include the following elements: inform the client about signs of pregnancy complications; measure weight, height, and blood pressure; take a urine sample and a blood sample to identify a medical condition. Similarly, the guidelines for integrated management of childhood illness (IMCI) offer a streamlined procedure for how to assess and treat children with symptoms of fever, cough, and diarrhea. In these cases, process quality can be measured as the degree to which health workers adhere to the guidelines.

Table 1 summarizes a set of possible methods that can be used to assess process quality. The methods fall into two groups: those that use information from real patient consultations and those that are based on hypothetical or fake patients.

<table>
<thead>
<tr>
<th>Data collection technique</th>
<th>Description</th>
<th>Advantages and disadvantages</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct clinical observation (DCO)</td>
<td>A surveyor attends a consultation with a real patient and observes how the clinician performs relative to a predefined checklist of procedures (based on clinical guidelines).</td>
<td>Useful for obtaining information about the quality of the diagnostic process and level of communication between provider and patient. Less suitable for generating information about diagnosis and treatment. Suitable only for standardized services, where a checklist can be developed and agreed. Quite costly for rare diseases (may have to stay for a long time).</td>
<td>Leonard and Masatu (2005); Mæstad, Torsvik, and Aakvik (2010) in Tanzania</td>
</tr>
<tr>
<td>Exit interviews</td>
<td>Patients leaving a health facility are asked about care they received (usually according to a checklist, similar to DCOs).</td>
<td>Similar to DCOs, but only those procedures that can be easily observed and remembered by patients can be studied.</td>
<td>Leonard and Masatu (2005)</td>
</tr>
<tr>
<td>Household survey</td>
<td>Households are interviewed about care recently received.</td>
<td>Similar to above, but because of recall problems, fewer items can be reliably measured.</td>
<td>DHS surveys (for example, antenatal care services). Bjerkman and Svensson 2009)</td>
</tr>
<tr>
<td>Reexamination</td>
<td>A surveyor trained as a health worker reexamines the patient after the consultation.</td>
<td>Better than DCO in measuring whether diagnosis and treatment is correct. Does not measure the quality of the diagnostic process as such.</td>
<td>Rowe…. REF IMCI evaluations</td>
</tr>
<tr>
<td>Vignettes</td>
<td>Clinicians are asked to diagnose and treat hypothetical patient cases.</td>
<td>Avoids potential problems of comparison across clinicians that result from a different case mix. Easy and quick to implement, and can also study more rare conditions. Certain procedures may be easier to capture through vignettes than through DCOs (for example, those procedures that require the doctor to look for something). The artificial situation may bias the results.</td>
<td>Leonard and Masatu (2005), in Tanzania; Das and Hammer (2005) in India.</td>
</tr>
</tbody>
</table>

*List is selective, not exhaustive.
Information about real patient consultations can be obtained either by the surveyor being present in the consultation room (direct clinical observation), by interviewing patients immediately after the consultation (exit interview), or later when the patients have returned to their homes (household surveys). In all these cases, the quality of the service is assessed by comparing what was done in the consultation (history taking, physical examinations, and communications) with a predefined set of procedures that should be performed in order to ensure high quality. Obviously, the number of procedures that can be monitored is larger and more reliable when obtained by direct observation than it is when relying on information from patients. Also when the time from consultation to data collection is longer, recall bias is also greater—that is, patients are not always able to recall actual circumstances accurately. On the other hand, direct clinical observation, and possibly also exit interviews, may be biased because the doctors or other health care workers may increase their effort because they are being assessed (Hawthorne effect). However, Leonard and Masatu (2005) demonstrated that these bias circumstances not necessarily be very large and may be quite short lived (see Box 1).

**Box 1: Hawthorne Effects in DCO: Does It Really Matter?**

Leonard and Masatu (2005) measured the magnitude of the Hawthorne effect by comparing the results of exit interviews from patients who had been consulted before and after the research team arrived, and also by comparing results between doctors who had an observer (termed *direct clinical observation*, or DCO) in the room with those who did not. As shown in the Figure 1a, there is a Hawthorne effect, but the effect quickly fades away. The performance of doctors who are observed returns to the normal level after 10 to 15 consultations.

**Figure 1a: Measuring the Hawthorne Effect**

![Graph showing adherence over time](source:Leonard and Masatu 2005)

- **y axis:** Adherence (% items correct)
- **x axis:** Number of previous consultations under observation
- **key:** solid line: Doctor observed from $t = 1$
  dotted line: Doctor never observed
An alternative way of establishing the quality benchmark is by letting a medically trained person reexamine the patients after the consultations and compare diagnosis and treatment from the real consultation with the results from the reexamination. This method is well suited for assessing the correctness of the diagnosis and whether treatment is provided in accordance with diagnosis, but it has to be combined with other methods (direct clinical observation or exit interviews) to assess the quality of the diagnostic process.

A second group of methods use hypothetical or fake patients to assess process quality. The method that has been most widely applied in low-income settings is vignettes (patient case simulations). In this case, one of the surveyors acts as a case study patient with some specific symptoms. The clinician, who is informed of the simulation, is asked to proceed as if the surveyor is a real patient, while another surveyor acts as an observer. The methodology presents several advantages: (1) all clinicians are presented with the same case study patients, thus making it easier to compare performance across clinicians; (2) the method is quick to implement and does not require waiting for patients with particular diagnoses; (3) intrusion and ethical issues that would arise if we were studying real patient cases are avoided. The method also has its drawbacks, however. The most important one is that the situation is a not a real one, which may bias the results.

Comparisons of vignettes with direct clinical observation in low-income contexts have revealed that performance scores are typically higher with vignettes, but that the correlation between the two measures is substantial (for example, Das, Hammer, and Leonard 2008). The performance score in vignettes has thus sometimes been interpreted as a measure of competence rather than actual performance (Das and Hammer 2005; Leonard et al. 2007). There is reason to believe that vignettes measure a blend of competence and actual performance, and that the blend depends on the actual design and framing of the tool. When vignettes are used as a proxy of process quality, the tool should be framed in a way that makes clinicians behave as similar as possible to real patient consultations.

**Responsiveness**

Responsiveness to people’s expectations in regard to nonhealth matters is another important aspect of the quality of health services (World Health Report 2000). The World Health Organization (WHO) has defined responsiveness along dimensions such as dignity, autonomy, confidentiality, prompt attention, access to social support networks during care, and quality of basic amenities.

Responsiveness, like informal payments, has to be measured through household surveys. Only in this way will it be possible to include the views of nonusers. Examples of survey tools can be found at [http://www.who.int/responsiveness/surveys/en/](http://www.who.int/responsiveness/surveys/en/).

Client’s judgments about responsiveness are bound to reveal a considerable degree of subjectivity. Comparability across people and communities can be enhanced by letting clients assess the degree of responsiveness in standardized test cases (vignettes) and using these responses to calibrate their assessment of responsiveness in real-world situations.

**Financial Issues**

It may be more common to discuss financial costs in relation to access to care than as a component of quality of care, but conceptually it makes perfect sense to include this aspect as one dimension of quality since it is a core factor determining the net benefit to the client of the services received.
It is of great to interest to measure the impact of RBF schemes on official user fees and informal payments, since health workers may decide to reduce these costs to attract a larger number of patients (although official user fees sometimes will be beyond the control of the individual health facility).

Official user fees can easily be measured through health facility surveys. Informal payments, on the other hand, are most reliably measured through household surveys, as exit interviews are likely to be more biased. Note that informal payments does not necessarily involve cash; food, animals, and other valuables can also be used.

**III.II How and Why RBF Affects Performance**

We have stressed that it is not sufficient to know that an RBF reform affected health worker performance—it is also important to ask why there was a change in performance. According to Figure 2, RBF may affect behavior directly, by offering monetary incentives for performance, or indirectly, by changing monitoring and supervision or by changing the underlying motivation of health workers. In addition, the impact of RBF schemes will depend on the scope for behavioral change, which is influenced by factors such as health worker competence and skills, workload, and so on. This section discusses how to go about measuring these influences.

**The Design and Implementation of the RBF Scheme**

RBF—or pay-for-performance—comes in many versions. All RBF schemes include some kind of payment contingent on performance or results, but the specification of the relationship between performance and payments takes a number of different forms. Below we first discuss the pros and cons of various designs and then turn to a discussion of implementation issues.

**Linear or Nonlinear**

A bonus may vary linearly or nonlinearly with measured performance. The Rwandan RBF scheme for example, offers a piece rate for a number of services (for example, US$ 0.18 per curative visit, US$ 0.92 dollars per complete vaccination course, US $ 4.59 dollars per delivery at the facility). Piece rates comprise a linear payment scheme.

In other cases, the relationship between pay and performance is less straightforward. One example of a nonlinear scheme is when there is a cap on the aggregate bonus in a piece rate scheme. Another would be a requirement to reach a specific performance target in order to receive a bonus (Not recommended in the PBF toolkit). Some countries (for example, Senegal) explore RBF schemes where health facilities will be rewarded only when they reach a predefined target. RBF schemes may also use a linear scheme for some indicators and a linear scheme for others (for example, the Tanzanian RBF pilot).

---

8 As pointed out in Part II, nonconditional payouts may also enhance workers’ effort (for example, through “gift exchange” driven by reciprocity motivation). In Basinga et al. (2011), units in the control group got the same amount of resources unconditionally as units with RBF, thus separating the “resource effect” from the “incentive effect” in the provision of RBF.

9 The Rwandan scheme is complicated by the fact that the overall bonus is multiplied by the facility’s score on a quality index.
From principal-agent theory, it is well known that a linear scheme is optimal only under restrictive conditions. However, a linear scheme is simple to understand and implement, and it is also robust against various forms of dysfunctional behavior. As far as nonlinear incentives are concerned, Oyer (1998) and Larkin (2008) provide strong empirical evidence that workers may exploit nonlinearities in the scheme in a way that harms the principal. On the other hand, Nalbantian and Schotter (1997) find that target-based (that is, nonlinear) schemes have a stronger impact on workers’ efforts than linear schemes.

**Team Bonus vs. Individual Bonus: Allocation Mechanisms**

With a team bonus, the payment to each individual depends on the aggregate performance of a team of workers (for example, all workers at a health facility). A team bonus is the only option when outputs cannot be assigned to a single individual. This will often be the case in health services production, where a number of individuals can contribute to a particular output. However, for certain services (for example, outpatient services, vaccination services) it may in principle be possible to use individual bonus schemes.

Standard economic agency theory predicts that unless there are strong complementarities among workers’ performances, an individual bonus is more powerful than a team bonus. This is because of the externality that is associated with a team bonus: a worker has to share the fruits from extra work with his colleagues but must absorb the full cost of extra effort alone. This collective action problem is a strong argument against team incentives. The problem can, however, be mitigated if workers can monitor each other closely and exert peer pressure. Workers may then enforce an implicit contract with high effort (Kandel and Lazear 1992). In health service provision, normally strong complementarities also exist between the efforts of various groups of professionals (for instance, between doctors and nurses).

A key assumption in the standard model is that workers have strictly self-regarding preferences—that is, they care only about their own incomes and costs. If workers possess other-regarding preferences, sharing the bonus with others will be seen as less of a sacrifice. Furthermore, a team bonus may ignite additional self-regarding (warm glow) motivations because high performance now will be recognized as generous and kind behavior (and lax performance as the opposite).

To our knowledge, all current RBF schemes in the health sector in developing countries are team based, because rewards are based on performance at the facility level. Note, however, that the impact of a team-based scheme may depend very much on how the bonus is allocated among workers at the health facility. Is everyone included? Does everyone receive the same? From an efficiency point of view, we expect the impact on performance to be larger when stronger incentives are provided to individuals who have greater influence on the performance indicators (for example, a doctor has greater influence on the number of consultations than a cleaner, which may justify a larger share of the bonus going to the doctor). In addition, fairness issues are likely to be important: allocation rules considered unfair may undermine worker motivation. In the evaluation of RBF schemes, it is therefore important to assess how the team bonus is shared within the team and to measure the perceived fairness of the system across various groups of employees.

**Relative vs. Absolute Performance**

An RBF scheme may condition pay on absolute or relative performance. In the latter case, performance is assessed relative to the performance of other comparable units (tournament pay) or against own past behavior. One advantage of a tournament-based RBF scheme is that it filters out any noise in the performance measure that is common to the units participating in the tournament (Holmstrom 1979). A
disadvantage of tournaments is that they do not give workers any incentives to cooperate—they may in fact encourage sabotage. Other-regarding and social preferences will again complicate matters. If the agents who compete for extra financing have altruistic concerns for each other, a tournament may work poorly. Bandiera, Barankay, and Rasul (2005) find, for example, that fruit pickers paid under relative incentives partially internalize the negative externality their effort imposes on others, especially when they work alongside their friends. On the other hand, Nalbantian and Schotter (1997) found that tournament pay elicited much higher levels of efforts than pre-defined performance targets.

Some RBF schemes encourage service providers to “compete” against their own past performance. If health care facilities have different capacities to deliver results, it may be efficient to accommodate the incentives scheme accordingly (for example, if the bonus is achieved only if a predefined target is achieved and the target is considered beyond reach, the incentive is without effect). However, if the benchmark is changed over time, agents may strategically hold back on their performance to get a laxer performance benchmark in the next period (this is known as the “ratchet problem” in the agency literature). Where relevant, these mechanisms should be acknowledged in an RBF impact evaluation.

**Implementation Arrangements and Processes**

Detailed knowledge of the implementation process may be of great value in understanding why RBF schemes work or not. First, there is a need to document whether the RBF is implemented consistently according to the specifications of the design. For example, even if the RBF scheme is designed to share the bonus equally between all health workers at a facility, there is no guarantee that the money will be actually shared equally. Similarly, even if rewards are scheduled to be paid every three months, this may be far from what actually takes place. Documenting such deviations from the design—in particular the variation in such deviations across facilities—is important.

Second, impact may be influenced by a number of implementation factors that are not specified in the design of the scheme. Examples of such factors might include how the scheme is communicated to health workers, who communicates it, how comprehensively the health workers understand how the system works, and so on.

A health worker survey would be the recommended tool for measuring these implementation aspects.

**Intrinsic Work Motivation**

Health workers are often portrayed as professional workers who have a mission to help their patients. The possibility that financial incentives may drive out other work motivations is therefore a concern that needs to be taken seriously when we assess the impact of pay-for-performance schemes in health care. Furthermore, in a population where intrinsic motivation varies across workers, monetary incentives may encourage some to work harder (those with little intrinsic work motivation, for example), while others become demotivated by such a program. It is useful to be aware of and, if possible, to measure this heterogeneity in work motivation when we assess how workers respond to financial incentives. Finally, RBF schemes may affect the motivation of the average health worker by affecting entry and exit from the workforce. If workers differ with respect to their social concerns and intrinsic motivation, one possible effect of RBF schemes is that performance pay attracts workers who are driven more by money than by a mission.
It is not easy to measure work motivation; it is even more difficult to distinguish intrinsic from extrinsic work motivation. One option is to use questionnaires to gauge health workers' motivation. Apart from being a relatively inexpensive form of data collection, this approach makes it possible to separate extrinsic from intrinsic motivation and therefore to account for the possibility that performance pay may dilute intrinsic motivation. The disadvantage with using questionnaire data is that motivation is a sensitive issue with strong normative connotations, which may bias what workers say about their own motivation.

For self-reported motivation, two instruments have been thoroughly tested and validated: the Work Preference Inventory (WPI) (Amabile et al. 1994) and the Minnesota Satisfaction Questionnaire (MSQ) (Weiss et al. 1967). Note, however, that many of the questions in these tools are related more to job satisfaction than to the intrinsic motivation to perform. It is important to understand the difference between these concepts. Although high satisfaction may lead to higher intrinsic motivation, it is perfectly possible to be satisfied without having intrinsic motivation to perform. An RBF scheme can make health workers satisfied as a consequence of receiving a higher income without necessarily inducing them to exert higher effort. This important distinction needs to be acknowledged in the selection of measurement tools.

Alternatively, motivation can be measured with behavioral data from lab experiments. The degree of altruism can, for instance, be measured by letting people play the dictator game described earlier. It is possible to assess whether RBF affects health workers' behavior and thus their motivation through such games. One important challenge, however, is that these games are highly stylized and usually very different from the situation of a health service provider. The relevance of the results for actual service provision will therefore be up for debate.

In a recent study, Brock, Lange, and Leonard (2011) use different sharing experiments to measure the social preferences of health workers. They make no attempt to measure how financial performance incentives affect the workers' social behavior, but they do find evidence that among the health workers (just as in any other population of individuals) a lot of heterogeneity seems to be present in social preferences. As we might expect, health workers who are more generous in a sharing experiment provide higher quality care, but there is no evidence that they would be more or less responsive to additional financial incentives.

A third alternative would be to design a lab experiment where RBF mechanisms, similar to those that are implemented in the real world, are built in, and then assess how exposure to financial incentives changes the workers' motivation and performance.\(^\text{10}\)

If financial incentives crowd out intrinsic work motivation, the process will probably be gradual, implying that such effects are more likely to be observed in the long rather than in the short run. This provides yet another reason for measuring the long-run impacts of pay-for-performance programs.

**Monitoring and Supervision**

The effect of RBF depends on how it affects monitoring and supervision. Some of the variables that will be important to measure to capture the influence of monitoring and supervision are the following:

\(^{10}\) A study by Burks, Carpenter, and Goette (2009) find evidence (in a totally different work setting that workers who are exposed to performance pay become less cooperative than workers who receive fixed pay.
**Changes in the Level of Formal Monitoring**

The level of change in formal monitoring will depend on which data were routinely monitored before the implementation of RBF. The preexisting routine data collection system may have been implemented differently at different health facilities, however, implying that RBF may result in different levels of change across facilities. One way of measuring this aspect would be to assess the number of additional indicators that are measured at each facility as a result of the RBF scheme (including indicators that were not actually measured previously even if they were supposed to be).

**The Quality of Monitoring**

RBF schemes are implemented in countries with high levels of corruption. It is not inconceivable that the quality of monitoring in these countries may be low. If it is possible to cheat on performance data, the RBF system will be undermined and may have no impact despite reports of improved performance. It is crucial in the assessment of the impact of RBF to collect independent performance data. For instance, household surveys may be very useful for measuring levels of health service utilization as well as certain quality indicators.

**Managerial Incentives for Enhanced Monitoring and Supervision**

The incentives of managers to strengthen their monitoring and supervision depend on the share of the bonuses that are allocated to managers. This share could vary across health facilities.

**The Nature of Supervision and Feedback**

The literature emphasizes that workers’ perceptions of whether supervision is controlling or supportive, as well as the way in which feedback is communicated, may influence performance. The methods of communicating performance results can be collected either through facility or health worker surveys or through participatory observation. Perceptions of whether supervision is controlling or supportive are more difficult to measure. One would have to make these concepts concrete and relevant to the particular context in order to capture any differences across facilities in a reliable way.

**Observability and Group Size**

The potential impact of group size on the degree of peer pressure implies that one should collect data on differences in the size of the workforce across health facilities. Also, since “shaming” effects depend on the degree of observability, it could be interesting to classify different types of health service according to their degree of observability. For instance, the quality of those services that are provided behind closed doors is less likely to generate any shaming effects than indicators such as the number of patients and attendance at work.

**Community Awareness About Performance Levels**

A prerequisite for effective community monitoring is that the community has knowledge about the level of performance. Such data can be collected through household surveys or through interviews with key informants.
Opportunity Costs

The impact of an RBF scheme will be smaller when it is difficult for health workers to improve their performance or, more generally, when the opportunity costs of improved performance are high. Here we discuss three factors that can have a significant impact on the opportunity costs of behavioral change for health workers in low-income settings: outside income opportunities, the magnitude of the know-do gap, and workload.

Outside Income Opportunities

Because of low salaries, health workers in many low-income countries supplement their income by engaging in various other income-generating activities, such as private health service provision, agriculture, trade, and so on. In order for RBF schemes to reduce absenteeism, the incentives must be strong enough to outweigh the benefits from these alternative activities. To understand why RBF schemes work or not, it will therefore often be important to collect information about the extent and value of such extra income opportunities.

The Know-Do Gap

The ability of health workers to improve their performance will depend on the difference between their actual performance and their level of knowledge and skills (the know-do gap). If the gap is small, we should not expect RBF to have a large impact on the level of performance.

One way to measure the know-do gap is to compare the results from direct clinical observation with the results from vignettes (note that if we want vignettes to be a measure of knowledge, they may require suitable framing and possibly also some incentives to make health workers do their best). Leonard and Masatu (2005, 2010) used this method to measure the know-do gap in outpatient consultations in Tanzania; their results are displayed in Figure 3.

The horizontal axis of Figure 3 measures competence drawn from vignettes and the vertical axis measures performance drawn from the DCO. The solid line at 45 degrees from the axes represents the points where knowledge is the same as practice. Though some points are close to the line, most of the points lie below the line where competence and practice are equal, meaning health workers do less than they know how to do. These points, therefore, represent health workers whose capacity is greater than their performance. The vertical distance between a point and the line is the know-do gap. The dashed line is the fitted curve through the data, and can be seen as the predicted relationship between capacity and performance. An RBF program that addressed motivation should close the know-do gap by increasing the effort of health workers who lie below the 45 degree line.
Workload

Because of the low number of health workers in many low-income countries, the workload may in some places be so high that it is essentially impossible to serve more patients, at least not without reducing time used per patient and therefore probably also the quality of service. In such places, RBF schemes are likely to do more harm than good. Mæstad, Torsvik, and Aakvik (2010) showed, however, that even in a country such as Tanzania, with very few health workers per capita, there are large (rural) areas where the number of patients per health worker is rather low. This suggests that there is considerable variation in workload between facilities and thus in their capacity to respond to financial incentives, which may give rise to highly diverse impacts of RBF schemes.

Workload can be measured by combining data on the number of patients from facility records with data on the number of health workers from facility surveys. If health workers provide several types of services, the techniques presented above for measuring productivity may also be used here to generate a measure of relative workload across facilities. Collection of data on the level of absenteeism may be required in order to measure the true level of human resource inputs.

III.III Measuring the Costs of RBF

RBF impact evaluations should include cost measurements to be able to assess the cost-effectiveness of the intervention.

Administrative Costs

RBF requires the financer to use resources to negotiate (and renegotiate) targets and to monitor performance. The administrative costs associated with performance pay may therefore be substantial and should be compared with the administrative costs of alternative methods of pay. For RBF to be
worthwhile, it is not sufficient that the incentive scheme produce benefits (more output) for the organization. These benefits must also cover the extra costs associated with performance pay. Freeman and Kleiner (2005) study a firm that switched from performance-pay (piece rates) to input-based (time rates) compensation. They find that although productivity fell, the firm’s profit increased because of the lower costs under fixed pay.

**Salary Top-Ups (Risk Costs)**

If RBF schemes could be implemented by redefining existing fixed salaries to performance-based pay, the costs of the scheme would be confined to administrative costs. However, the standard principal-agent model emphasizes that performance-based pay will generate risk costs: high-powered financial incentives often imply that the total compensation an agent receives will fluctuate a lot, and risk-averse workers must be compensated for being exposed to such uncertain income. This will increase the expected cost of the transfer to the agent. Therefore, RBF schemes cannot normally be implemented simply by redefining a fixed salary to a (partly) performance-based salary. This is in line with the way RBF is implemented in most low-income countries: as a top-up on existing salaries.
PART IV – SOME PRACTICAL STEPS TO EVALUATE IMPACT OF RBF ON PERFORMANCE

The following section outlines some of the practical steps required to evaluate the impact of RBF on health worker performance. It discusses the four key steps to consider.

IV.I Four Key Steps to Consider

In practice, the evaluation of the impact of RBF on health worker performance involves the following steps: (1) designing the RBF scheme to be evaluated, (2) identifying the dependent variables to be measured, (3) selecting one (or several) relevant comparison intervention(s), and (4) selecting tools for measuring health worker performance and its explanatory factors.

Designing of the Scheme

Assuming that our goal is to address a real-world problem—to improve health worker performance—and not to study RBF from a theoretical point of view, the key priority in the design of the RBF scheme should be to maximize cost-effectiveness. Note that high cost-effectiveness does not necessarily imply that impact is high: high impact and medium costs may result in lower cost-effectiveness than medium impact and very low costs.

The first stage in the design process should be to ask which behaviors are most important to change, followed by in-depth qualitative work to get a grasp on the amount of financial payments that would constitute a meaningful incentive, which institutional opportunities and constraints are in place for effective implementation, and so on. This analysis should include a comparison of various incentive scheme designs, such as linear schemes vs. schemes with threshold levels (or other nonlinear schemes) vs. relative performance incentives, and so on. A key point throughout the design phase is that anything that is implemented be grounded in solid knowledge of key contextual factors that may influence the impact of the RBF scheme.

The qualitative work may result in some clear recommendations about how to implement the RBF scheme and in others that are more unclear. The more ambiguous aspects will often be primary candidates for further testing. In some cases, such testing can be implemented through smaller, rather than larger, experiments. If resources allow, alternative designs can be tested by implementing several variants of the full RBF scheme. If the RBF scheme is implemented as a randomized controlled experiment, the causal impact of alternative designs can be identified with high reliability.
Identifying the Dependant Variables

In accordance with Figure 2, exploring the relations between RBF and health worker performance requires several different analyses. Indeed, as shown in the previous section, health workers’ performance encompasses several dimensions:

- Quality of care
- Productivity
- Presence at work
- Responsiveness

Ideally, each of these four dependent variables should be measured. Instruments for that purpose have been described in Part III.

Selecting the Relevant Comparison Interventions

When selecting the control group, a decision has to be made about what constitutes the relevant comparison to the implementation of an RBF scheme.

The point here is to decide which features of RBF should be explored experimentally. An obvious consideration here is to ask which RBF feature has the biggest expected impact on health worker performance:

- Is the monitoring and supervision feature expected to have the most impact? If so, a control group could have no incentives and only monitoring and supervision.
- Is the additional money given to health workers (whether conditional or not) expected to have the most impact? In that case, a control group could give unconditional bonuses to health workers.
- Is the level of performance bonus expected to have the most impact? If so, a control group could have “low” bonuses and a treatment group could have “high” bonuses.
- Is the rule used for calculating the bonus expected to have the most impact? If so, one a control group could have nonlinear bonuses and possibly a second control group could have relative bonuses, while the treatment group would be a linear scheme.

Other designs are obviously possible. The one mentioned above is simply given as examples. As long as the evaluation budget allows it, these various questions should be combined in the same design. Box 2 provides several examples of these combinations.

11 In addition, the evaluation must look at other important questions: (1) Is this a policy relevant dimension? (2) Is this a dimension where we are uncertain about the effect?

12 We have deliberately not mentioned the issue of individual bonuses versus team bonuses. We believe that in a low-income country such an experiment would be extremely expensive to carry out and therefore would not be practically relevant. Conversely, another issue not treated here is how the team bonuses are allocated among health workers. This is an important variable that deserves further attention.
### Box 2: Some Examples of RBF Impact Evaluation Designs

<table>
<thead>
<tr>
<th>Country</th>
<th>Financial bonuses</th>
<th>Monitoring</th>
<th>Increased management autonomy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kyrgyzstan</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Financial bonuses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1 (RBF)</td>
<td>Yes, conditional</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>T2</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Cameroon</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Financial bonuses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1 (RBF)</td>
<td>Yes, conditional</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>Yes, unconditional</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>T3</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Benin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Financial bonuses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1 (RBF-1)</td>
<td>Yes, conditional</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>T2 (RBF-2)</td>
<td>Yes, conditional</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>T3</td>
<td>Yes, unconditional</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>T4</td>
<td>Yes, unconditional</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>C</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Central African Republic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Financial bonuses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1 (RBF-1)</td>
<td>Yes, at 100% payment level</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>T2 (RBF-2)</td>
<td>Yes, at 50% payment level</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

*T=Treatment. Source: authors

### Selecting the Tools for Measuring Performance

A theory of change should be the basis for choosing which performance variables to monitor. For instance, if we think that the main causal pathway is that RBF will make health workers more pleasant toward their patients because increased pleasantness will attract more patients to the health facilities, then we should focus on measuring health workers’ responsiveness. On the other hand, if the RBF scheme is designed primarily to increase the quality of consultations, those are the aspects we should focus on measuring.

In addition to measuring the primary mechanisms through which RBF is supposed to work, impacts in dimensions that may suffer because they are not incentivized should also be measured (see, by way of comparison the multitasking issue). Again, a theory of change will be an important tool in selecting which activities to monitor. Typically, they should be activities performed by the same health workers who have some of their activities incentivized.
The choice of tools for measuring health worker performance ultimately depends on the RBF intervention itself (which activities are incentivized), the advantages and disadvantages of various tools, and cost considerations.

For example, if the aim of the RBF scheme is to improve the quality of antenatal care, exit interviews may be as useful as direct observation because many of the procedures that are supposed to be done during an antenatal care visit are relatively easy for the client to observe and remember. Exit interviews are less useful when we are interested in procedures that a lay person would have difficulty in observing (see Table 1 for an overview of advantages and disadvantages of the various methods).

**Using Non-experimental Methods to Assess the Influence of Mediating Factors**

Table 2 summarizes some of the mediating factors discussed in Part III and the tools that can be used to measure them. Some of the factors may be incorporated into the experimental design, which of course is preferable from an analytical point of view. The effect of other factors must be assessed through more imperfect methods (mediation analysis). Qualitative methods will also often be useful in combination with quantitative techniques.

**Table 2: Mediating Factors and Assessment Tools**

<table>
<thead>
<tr>
<th>Mediation factors</th>
<th>Data collection tools</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RBF design and implementation</strong></td>
<td></td>
</tr>
<tr>
<td>How the bonus is shared among workers, when the bonus is paid, level of knowledge about the RBF scheme among health care workers, etc.</td>
<td>Health worker survey</td>
</tr>
<tr>
<td><strong>Intrinsic motivation</strong></td>
<td></td>
</tr>
<tr>
<td>Level of intrinsic motivation</td>
<td>Health worker survey</td>
</tr>
<tr>
<td></td>
<td>Lab experiments</td>
</tr>
<tr>
<td><strong>Monitoring and supervision</strong></td>
<td></td>
</tr>
<tr>
<td>Change in level of formal monitoring</td>
<td>Health facility survey</td>
</tr>
<tr>
<td>Quality of RBF data collection and monitoring</td>
<td>Household survey</td>
</tr>
<tr>
<td>Managerial incentives for enhanced supervision</td>
<td>Health facility survey</td>
</tr>
<tr>
<td>The nature of supervision and feedback</td>
<td>Health worker survey</td>
</tr>
<tr>
<td>Group size</td>
<td>Health facility survey</td>
</tr>
<tr>
<td>Community awareness about performance</td>
<td>Household survey</td>
</tr>
<tr>
<td><strong>Opportunity costs</strong></td>
<td></td>
</tr>
<tr>
<td>Outside income opportunities</td>
<td>Health worker survey</td>
</tr>
<tr>
<td>Know-do gap</td>
<td>Direct clinical observation and vignettes</td>
</tr>
<tr>
<td>Workload</td>
<td>Health facility survey (including absenteeism survey)</td>
</tr>
</tbody>
</table>
PART V – CONCLUSION

With this paper, our objective was not to offer evidence on what kind of RBF features work (an entire book would not be enough), but rather to (1) identify what variables have been analyzed in the literature to better understand the impact of incentives on workers performance, and (2) to determine how this can be done in future evaluations, especially in African health sectors.

By examining RBF from the perspective of two important theoretical views of motivation (the principal-agent model and the behavioral-economics model), we have shown that it is not at all clear that RBF can always work, or even how often it may be expected to be helpful. In part the answer of the impact of RBF on HRH is unknown (because we have not yet done enough research); in part it is unknowable (because it will always depend on local institutions and cultures). The principal-agent model argues strongly that when an administrator can specify what he wants and can pay for the outcomes he seeks, RBF should work. The behavioral economics model suggests that health workers may have multiple sources of motivation and pushing strongly on one of the sources (extrinsic monetary motives) can have ambiguous impacts on the other sources of motivation. The interplay between these forces is complicated.

In part, careful design and consideration of the local circumstances are necessary to avoid the most obvious mistakes of an RBF program. If we cannot specify what we want health workers to do, tying performance pay to an arbitrary, narrow measure is almost certain to result in a failed program. However, even when the program is carefully designed there remain too many unknowns for anyone to say with certainty that it will be successful. That is why we propose that programs test the structure of motivation within themselves.

We have outlined a program for testing the sources of motivation and for managing those sources to the best advantage. We have shown that most of the various RBF effects can be unbundled or disentangled. Each of these effects can be assessed by combining experimental control and statistical control. Such a program is not without costs, even when compared with an impact evaluation of one simple RBF program. Testing multiple theories within a program means that somewhere you are making a mistake. However, it is also much more likely that somewhere you are doing the right thing, and proper design means that everyone will be able to recognize where the right thing is being done, and we will all be able to learn from that carefully calibrated understanding of success.

By following an agenda that seeks to understand the details of RBF—its advantages and pitfalls, where it works and where it does not—we can take advantage of the current excitement for RBF while at the same time preparing the necessary groundwork for its maturing as a tool. It is not possible that a simple RBF is exactly the right tool for all health measures; in some cases, RBF may do more damage than good. However, we believe that the early evidence combined with the poor performance of all other schemes so far clearly merits experimentation and further exploration.
REFERENCES


43


44


The Contribution of Traditional Herbal Medicine Practitioners to Kenyan Health Care Delivery

Results from Community Health-Seeking Behavior Vignettes and a Traditional Herbal Medicine Practitioner Survey

John Lambert, Kenneth Leonard with Geoffrey Mungai, Elizabeth Omindi-Ogaja, Gladys Gatheru, Tabitha Mirangi, Jennifer Owara, Christopher H. Herbst, GNV Ramana, Christophe Lemiere

September 2011