Can Pay-for-Performance Improve Child Health in Low- and Middle-Income Countries? A Review of the Evidence

JUSTIN ABRAHAM[1]

JAN COOPER[2]

OCTOBER, 2019

[1] Department of Economics. University of California, San Diego. jabraham@ucsd.edu.

[2] J-PAL South Asia, New Delhi, India | *Present address:* Department of Global Health and Population, Harvard T. H. Chan School of Public Health, Boston, MA. jancooper@hsph.harvard.edu

This Review has been commissioned under the Cash Transfers for Child Health (CaTCH) Initiative, a J-PAL South Asia regional research initiative funded by the Bill & Melinda Gates Foundation. The Initiative is led by J-PAL Affiliates Pascaline Dupas (Stanford University) and Seema Jayachandran (Northwestern University).

Table of Contents

Executive Summary

Governments in low- and middle-income countries have implemented pay-for-performance schemes (P4P) for health care providers with the objective of improving health outcomes. P4P are incentive schemes that reward providers on the basis of measurable performance indicators. Such programs attempt to incentivize healthcare providers to deliver high-quality, timely health services. Despite enthusiasm for these policies, the evidence about their effectiveness is inconclusive, especially for the health of children in resource-limited settings. Furthermore, the evidence is less clear about how to design optimal pay-for-performance programs.

This article reviews 36 recent evaluations of P4P for providers of child and maternal healthcare in 15 low- and middle-income countries. To draw policy-relevant conclusions about the impact of P4P programs, our paper updates previous literature reviews and focuses solely on experimental and quasi-experimental evidence that can speak to the causal effect of P4P on healthcare delivery and the health outcomes of children in low- and middle-income countries. First, we summarize the impact of P4P on malnutrition, anemia, childhood morbidity, and infant mortality. We link these results to evidence on the provision of services including antenatal care, institutional deliveries, immunization, and nutrition supplementation. Second, we draw on a conceptual framework and use our evidence base to motivate a discussion on the effects of various design features. Finally, we identify opportunities for future research on the application of P4P for healthcare providers.

We find strong evidence that P4P can work to reduce malnutrition among children under five. Evidence from India, Indonesia, Rwanda, and the Philippines suggests that incentivizing nutrition-targeted tasks has short-term positive impacts on weight-for-age as proxies for nutrition status. These results can be traced, in part, to successfully motivating health workers to provide supplementation services, which are inexpensive and relatively simple to implement.

The evidence is mixed about how P4P can affect anemia prevalence, childhood morbidity, and neonatal mortality despite improved institutional deliveries, immunizations, and antenatal care. Seemingly contradictory findings can be attributed to program and contextual differences. Furthermore, these health outcomes have complex determinants, many of which are outside of the provider's discretion.

There is limited evidence on the cost effectiveness of P4P programs. Cost-benefit analyses suggest that P4P in certain contexts are more cost-effective than demand-side interventions like insurance provision. Evidence further suggests that there may be important complementarities between P4P and other health interventions. Cost-effectiveness and program interaction remain large gaps in the empirical literature.

We draw some conclusions from our evidence base regarding the design of P4P programs and how different features of these programs can work. First, although health outcomes are not the result of provider services alone, it may be beneficial to tie performance rewards to these rather than inputs if it allows providers flexibility in meeting policy objectives. Second, our evidence base suggests that manipulation of performance indicators remains a risk for providers under P4P. Third, incentives tied to relative performance can be used to ensure equitable allocation of resources among facilities with differing capabilities but that policymakers should its effects on workplace relationships. Fourth, there is a potentially important role for non-financial incentives and intrinsic motivation in improving performance among health providers in low- and middle-income countries. Comparisons of P4P policies on the basis of cost-effectiveness and the direct testing of design features are promising opportunities for further research.

1 Introduction

The past two decades have brought significant gains in child health worldwide. Better diagnostic and treatment technologies, changes in health behaviors, and more consistent access to health care have been a powerful combination for improving outcomes for infants and children. However, not all gains have manifested equally across populations, and progress for the health of children in low- and middle-income countries has stagnated, particularly for preventable diseases.

While the causes of poor child health are complex, one important driver is the supply of health care. Poor child health outcomes can often be traced, in part, to the lack of high-quality, timely, and accessible health care. Resource constraints, weak organizational capacity, and gaps in care delivery contribute to poor quality of services (Berendes et al. 2011). Also, providers are often overburdened with patient care and paperwork, which can limit the quantity of care they give. Yet at the same time, high absentee rates are common. In India, for example, high absentee rates among health workers is compounded by substandard clinical practice (Chaudhury et al. 2006; Das et al. 2016). To address this, policymakers have increasingly turned to pay-for-performance (P4P)—incentive schemes that directly reward providers for achieving certain standards of performance. Governments in low- and middle-income countries have increasingly been adopting performance incentives in healthcare systems at multiple levels. However, do these programs really work? Only recently has there been rigorous evidence testing the impact of P4P on provider behaviors and, in turn, child health status.

The goal of improving stagnating health outcomes has led to the growing interest in incentivizing health care workers. Yet several questions emerge around how best to do so. Policymakers require information about how to design P4P schemes, in particular, who should receive what kind of incentive, for which activities, and how often. More broadly, designing and implementing successful P4P schemes raise further questions about workforce dynamics and structures, such as how incentives might change who chooses to become a healthcare worker, or how incentives might affect competition between workers. New evidence is emerging as individual P4P schemes are evaluated. Analyzed together, these new studies can help answer some of these questions and identify, more specifically, areas for future research.

There are a handful of recent systematic reviews that address the impact of P4P on the behavior of health providers and health outcomes. Yet, there is no consensus on the effectiveness of performance pay since these existing reviews summarize very different programs over a wide variety of settings. Petersen et al. (2006) examine 17 empirical studies testing effects of financial incentives on process-of-care measures and reports moderate impacts on adverse selection and manipulation of indicators. Rosenthal and

Frank (2006) are unable to attribute any effect on health outcomes to financial incentives apart from the effect of feedback to providers or better monitoring introduced alongside P4P schemes. The most recent reviews specifically concentrate on P4P in low- and middle-income countries. Witter et al. (2012) attempt to conduct a meta-analysis but found that the evidence base at the time was too weak to make any general conclusions. Rowe et al. (2018) also conduct quantitative meta-analysis on general strategies to improve provider practices. However, the authors review only three studies that test financial incentives alone and document moderate effects. These reviews document very little work on the cost-effectiveness of any of these programs. Other reviews stand on a thin evidence base that relies on evaluations in areas outside of health. Furthermore, the evidence base for these reviews included everything from qualitative work, randomized trials, pre-post studies, and interrupted time series studies. A narrowing of the empirical literature to rigorously identified evaluations can help form consensus around the most robust set of results. Furthermore, newer reviews of P4P will need to pay attention to how design features change potential impacts. For instance, Miller and Babiarz (2014) provide a recent handbook chapter on P4P in developing countries that organizes its survey on design-centered questions but does not conduct any systematic review of the literature.

Previous reviews highlight the need for an updated, rigorous survey focusing on policy design and that address unanswered questions including: should health-care workers be paid based on their "inputs", that is, the effort and time spent on their work; or on "outputs" such as the quantity or frequencies that health services are delivered? In turn, a discussion about the design of P4P programs in low- and middle-income countries requires a conceptual framework attuned to the specificities of providing health care in those contexts.

The objective of this literature review is to summarize the most up-to-date, credible evidence on P4P among frontline healthcare providers in low- and middle-income countries. We conducted a review of published manuscripts from the last 10 years and unpublished working papers from the past 5 years. We limited our search to studies that evaluate supply-side incentives to improve child health in low- and middle-income countries. Further, we included only studies that use experimental or quasi-experimental approaches to evaluate the impact of the P4P program. By limiting our analysis to these empirical approaches, we hope to build on existing reviews by assessing the newest and most rigorous evidence of P4P.

Our conceptual framework is based on a principal-agent model and its extensions (Section 3). This framework provides predictions of how P4P could work to improve health outcomes and provides the lens through which we organize and interpret results. We first summarize the results on child health outcomes in Section 4.1. We then examine in Section 4.2 how different P4P interventions affect provider inputs in an attempt to explain impacts

on health outcomes and to test the theoretical implications presented in Section 3. Section 5 uses the empirical findings to discuss the potential effects of different P4P design features and concludes with opportunities for future research.

Overall, we find that P4P, broadly defined, is an effective supply-side intervention for improving child and maternal health in low- and middle-income countries. Evidence suggests at short-term effects on indicators of malnutrition, anemia, and neonatal mortality. P4P improves access to health services as well as the quality of provision. P4P contracts can be complex; we discuss evidence surrounding several design considerations motivated by our theoretical framework. Finally, we identify promising areas of research for performance pay for health providers.

2 Methods

Our review focuses on programs implemented in low- and middle-income countries that provide explicit incentives to health care providers conditional on service delivery. Moreover, we limit our attention to those evaluations which meet certain methodological criteria. Finally, we provide a general characterization of the evidence base at the end of this section.

2.1 Systematic search

We performed a systematic search of publications in peer-reviewed academic journals and unpublished academic working papers. First, we identified key existing reviews on the effect of P4P on health outcomes and provider behavior. We included studies covered in those reviews that satisfy criteria summarized below. Second, we conducted a search for journal articles using the EconLit and PubMed databases using the keywords in Table 1.[3]

[3] Search terms for developing countries include Afghanistan, Benin, Burkina Faso, Burundi, Central African Republic, Chad, Comoros, Congo, Eritrea, Ethiopia, Gambia, Guinea, Guinea-Bissau, Haiti, Korea, Liberia, Madagascar, Malawi, Mali, Mozambique, Nepal, Niger, Rwanda, Senegal, Sierra Leone, Somalia, South Sudan, Syria, Tajikistan, Tanzania, Togo, Uganda, Yemen, Zimbabwe, Angola, Bangladesh, Bhutan, Bolivia, Cabo Verde, Cambodia, Cameroon, Congo, Côte d'Ivoire, Djibouti, Egypt, El Salvador, Georgia, Ghana, Honduras, India, Indonesia, Kenya, Kiribati, Kosovo, Kyrgyz Republic, Laos, Lesotho, Mauritania, Micronesia, Moldova, Mongolia, Morocco, Myanmar, Nicaragua, Nigeria, Pakistan, Papua New Guinea, Philippines, São Tomé and Principe, Solomon Islands, Sri Lanka, Sudan, Swaziland, Timor-Leste, Tunisia, Ukraine, Uzbekistan, Vanuatu, Vietnam, West Bank and Gaza, Zambia, Albania, Algeria, American Samoa, Armenia, Azerbaijan, Belarus, Belize, Bosnia and Herzegovina, Botswana, Brazil, Bulgaria, China, Colombia, Costa Rica, Cuba, Dominica, Dominican Republic, Ecuador, Equatorial Guinea, Fiji, Gabon, Grenada, Guatemala, Guyana,

Table 1: Keywords used in database search

	Keywords
At least one of the following	*performance pay, P4P, pay-for-performance, PFP, performance-based incentives, PBI, performance-based pay, PBP, results-based financing, RBF, performance-based financing, PBF, performance incentives, financial incentives, extrinsic rewards, bonus, target outcome*
and	the name of at least one LMIC country, *poverty, poor, development, developing, low-income*
and	*community health worker, frontline workers, physician, nurse, health, medicine*

We included working papers using a more stringent exclusion criteria described below. We reviewed abstracts to eliminate duplicate papers. In cases where multiple papers discuss the same results from the same program, we included only the most recent and most comprehensive. We then reviewed papers in full to screen according to the following criteria.

1. *Quality criterion*: Papers published in peer-reviewed academic journals since 1990 or unpublished working papers written by 2005. Unpublished working papers written before 2005 are deemed to have methodological weaknesses that warrant not having been published.

2. *Program criterion*: Evaluations of payment or bonus schemes that target healthcare providers and conditional on behavior or measures of performance. This definition is meant to encompass everything from performance-based funding of entire facilities to performance bonuses for individual workers. Throughout the review, we use 'P4P' to refer to individual incentives and 'PBF' for higher-level incentives.[4]

3. *Contextual criterion:* Evaluations taking place or implemented in low- and middle-income countries (as defined by the IMF). The study of P4P is extensive so a focus on

Iran, Iraq, Jamaica, Jordan, Kazakhstan, Lebanon, Libya, Macedonia, Malaysia, Maldives, Marshall Islands, Mauritius, Mexico, Montenegro, Namibia, Nauru, Paraguay, Peru, Romania, Russian Federation, Samoa, Serbia, South Africa, St. Lucia, St. Vincent and the Grenadines, Suriname, Thailand, Tonga, Turkey, Turkmenistan, Tuvalu, Venezuela

[4] In general, P4P, PBF, results-based pay, etc. are all overlapping terms that describe the wide category of programs which this review discusses.

low income settings helps refine the scope of the review and makes salient the unique economic conditions of those environments.

4. *Health criterion*: Payment schemes implemented in the health sector or studies measuring health outcomes related to maternal or child health.

5. *Methodological criterion:* Randomized controlled trials or evaluations using quasi-experimental methods (difference-in-differences, regression discontinuity, or instrumental variables). These empirical strategies control for selection bias or confounding factors in the evaluation of P4P. Thus, effects derived from these methodologies can be understood as causal.

2.2 Compilation of results

This review is neither exhaustive nor a quantitative meta-analysis since program details, study design, and outcomes of interest differ across papers. Rather, we organize our discussion around the most commonly studied health outcomes and around themes precipitated by our conceptual framework. We report effect sizes for health indicators for which we have the most evidence (wasting, hemoglobin concentration, illness, and neonatal mortality) and discuss program differences in the text.[5] We also report null results but do not attempt to distinguish between the absence of an effect and inadequate statistical power.

2.3 Summary of the evidence base

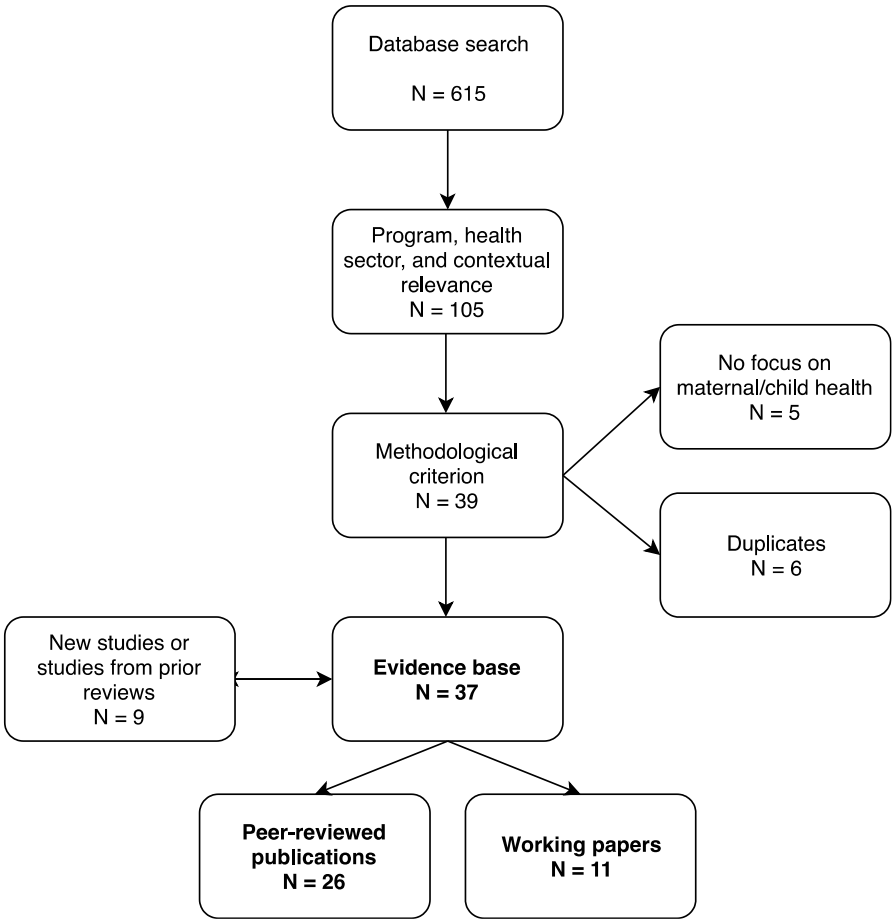
The initial database search yielded 615 papers, which we narrowed to 105 after reviewing abstracts for program relevance, focus on health care providers, and having taken place in a low- or middle-income country. This number was reduced to 39 after screening for methodology. We removed 6 duplicates and 5 studies that did not address child health. Finally, we included 19 recently published or reviewed studies meeting the screening criteria. The final evidence base consists of 37 studies: 11 working papers and 26 peer-reviewed publications. Figure 1 summarizes the compilation of the evidence base. Table 2 lists the included studies.

The majority of studies consist of evaluations of national PBF schemes delivered at the institutional level from 11 countries. Of the included studies, 9 employ difference-in-

[5] Table 2 lists each study and the primary outcomes they report. This review also discusses secondary analyses where relevant.

differences and 27[6] are randomized controlled trials. Most of the evidence comes from after 2010 with none from before 2000. All studies focus on some aspect of child and maternal healthcare.

Figure 1: Compilation of the evidence base by search criteria



3 Conceptual Framework

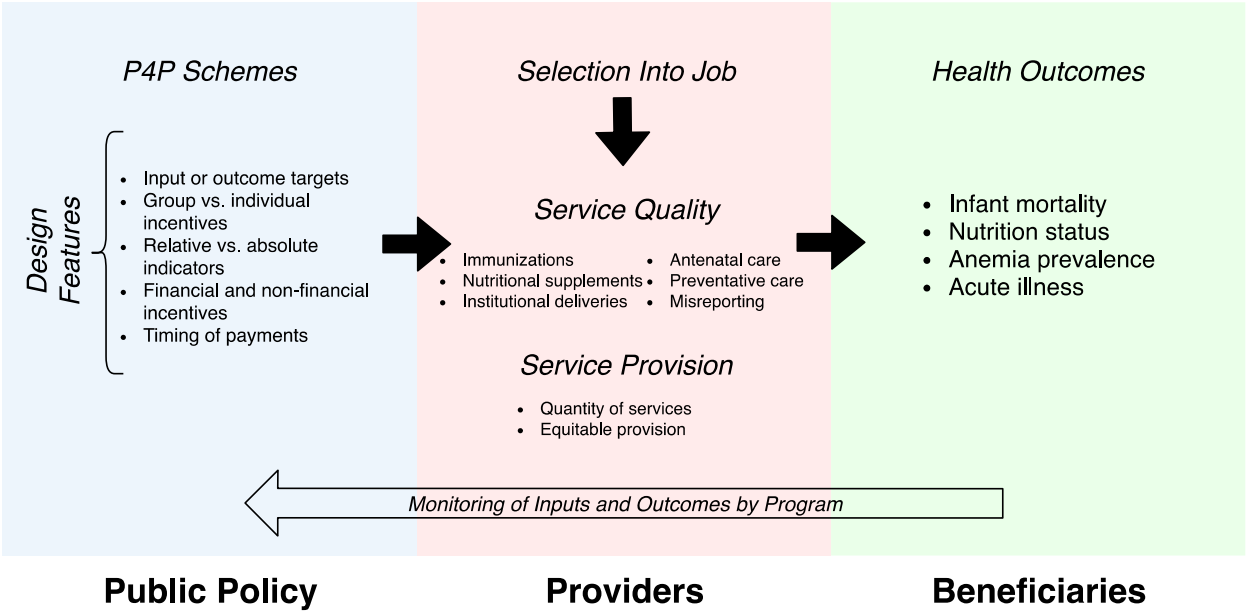
This section outlines a conceptual framework that we use to ground empirical questions regarding the design of P4P incentives and to help interpret empirical findings. The problem of selecting an optimal payment scheme to incentivize frontline health care providers to produce a certain quality of service is well-described by principal-agent

[6] J-PAL affiliated researchers conducted 12 of the 27 RCTs included in this review.

models. These refer to a set of game theoretic models in which one individual (the principal) contracts with another (the agent) to undertake some action for the principal. The result of principal-agent models is the characterization of contracts that principals (e.g., the health ministry) might offer to agents (e.g., frontline workers) to induce specific behavior.

We begin with a textbook principal-agent model as a basis for further extensions. Suppose that the ministry of health, whose ultimate objective is to minimize child illness, hires a community health worker to provide immunizations. Healthcare services partly determine child outcomes but are costly for the worker to provide in that she must expend effort. In this simplistic model, the providers seek to maximize her wage and minimize the work she has to do. A contract specifies wages paid to the provider that can depend on the service inputs, the final health outcomes, both, or neither. The decision problem for the principal is to design a contract that induces providers to work towards policy objectives. Figure 2 maps the potential pathways between performance pay, healthcare provision, and health status.

Figure 2: Overview of conceptual framework



3.1 Rewarding inputs versus rewarding outcomes

Policymakers need to decide which indicators to incentivize. Our simple setup suggests two possibilities: contracting on provider inputs and contracting on health outcomes. Intuitively, P4P should target directly the policy-relevant outcome. This aligns the incentives of providers with that of the policymaker while at the same time giving the

provider scope to use local knowledge or to innovate to achieve policy goals. Such designs can also benefit from the relative ease of observing health outcomes compared to intensive monitoring of provider behavior. This is an important consideration in low- and middle-income countries, where providing effective public services is hampered by weak institutions.

However, importantly, child health outcomes are only partially determined by the performance of healthcare providers. A P4P scheme that contracts on outcomes thus ties provider pay to factors that are outside of the providers' control. Incentivizing in this manner comes at the cost of greater financial risk faced by the provider so incentives will have to be higher than if inputs were incentivized. To put it another way, performance pay design trades off "insurance" from uncontrollable outcomes, which P4P for inputs does better on, and directly incentivizing the policy objective, which P4P for inputs does better on. Primary care in low-income settings is largely provided by health workers who are themselves vulnerable to further financial risk.

3.2 Multitasking

The previous discussion suggests that contracting on specific inputs or tasks is desirable when inputs have strong and well-understood effects on outcomes. But community health workers, as primary care providers, are responsible for multiple tasks that make competing demands on their time. Accordingly, there are multiple inputs that determine child health outcomes (immunization, improving sanitation, guarding against malnutrition, providing first aid, etc.). Assuming a single dimension of effort for the agent in the basic framework thus seems ill-suited in modelling P4P in this environment.

Holmstrom and Milgrom (1991) incorporate scope for effort across many different tasks to study how agents allocate their attention to different tasks under an incentive scheme. A concern is that input incentives on a subset of tasks will concentrate effort towards those tasks to the detriment of the final outcome (multitasking problem). If tasks are substitutes in the production of the outcome, then incentives will tend to shift attention to one task. If tasks are complements, then providers might be induced to perform even unincentivized tasks; she engages more in the incentivized task, which raises the productivity of performing the second task (by the definition of complementarity), so she chooses to also engage more in the second task. One lesson from the multitasking literature is that when multiple inputs contribute to a set of outcomes, outcome-based P4P allows the provider to use local knowledge creatively to allocate effort over the different inputs to maximize contracted outcomes. For example, outreach activities and consultations may be best left under the purview of providers.

The issue of multitasking also calls attention to potential unintended consequences of P4P. If health care providers can influence outcomes by increasing quality *and* by manipulating records, then a P4P scheme may incentivize the latter if it is easier to misreport. Another concern is that providers may cherry-pick subpopulations to service and undermine equitable provision. Incentives can shift service away from the most vulnerable since they are also likely the most difficult to reach. In this case P4P could have unintended regressive impacts. Such considerations underscore the importance of the entire portfolio of provider actions related to the desired outcome.

3.3 Group and individual incentives

The framework has thus far focused on the responses of a single individual to a tailored payment scheme. In sectors like public health, however, service provision is accomplished by teams of health care providers. Schemes that reward entire health care units on performance only indirectly incentivize individual providers. Under these programs, operational budgets are sometimes allocated based on performance indicators. Group-level incentives can give rise to free riding problems which dampen the individual motivating effects of performance pay.

Groups of providers allow for the use of relative performance as the basis for reward. Competition can play a role in improving outcomes, but a focus on individually rewarding tasks may be counterproductive to the degree that health care provision requires cooperation. These examples highlight the importance of group dynamics and the nature of provider tasks in designing P4P schemes.

3.4 Intrinsic motivation

Although community health providers are often remunerated, they may also be acting on pro-social motivations. Bénabou and Tirole (2006) analyze a model in which individuals experience intrinsic motivations, extrinsic motivations, and a concern for one's self-image. Their model describes situations where financial reward can *reduce* effort because it changes how others infer the individual's motivations. A consideration of non-financial motivations among providers makes the effects of P4P on effort theoretically ambiguous. Simultaneously, these considerations also suggest that there are other, non-financial incentives that could be used by policymakers to motivate better quality of service.

3.5 Selection

We have so far examined the problem of creating optimal incentives after a contractual relationship between the principal and agent has already been established. Yet another aspect of different P4P schemes is how it might induce different types of agents to select

into jobs and in turn how it affects the quality of the services they provide. For example, P4P could induce more qualified individuals who can better take advantage of incentives to select into the position. Financial reward can also turn away intrinsically motivated agents.

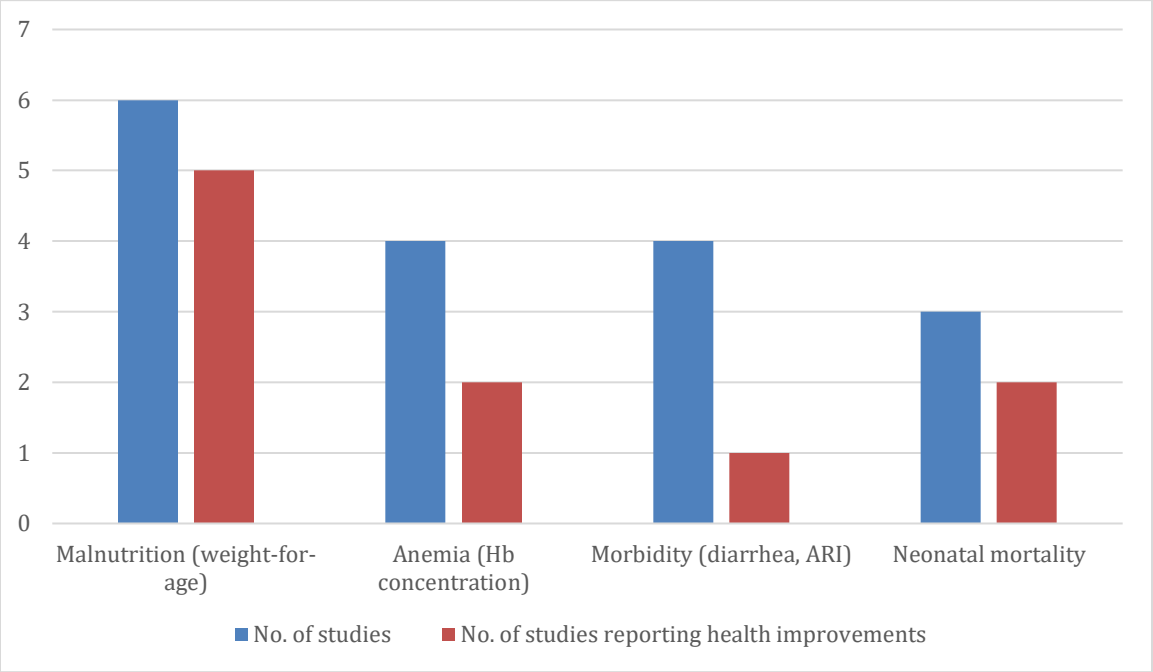
3.6 Long-term contracts

Our framework has not been able to address the impact of different timing of payments with respect to the work provided, the realization of the final outcome, or job tenure. How is provider behavior affected when contracts can specify the distribution of payments over time? Intuitively, a principal that can observe a history of performance can make more accurate inferences about agent behavior. It might also be optimal for principals to defer compensation because tenured workers face a larger incentive to perform and the principal can leverage the prospect of large future rewards to discipline new workers (Lazear 1981; Harris and Holmstrom 1982). Dynamic considerations suggest that it is important for public health providers to pay attention to the career concerns of its workforce.

4 Empirical evidence on impacts of P4P

In this section, we first summarize the effects of P4P on the outcomes of interest: child health status. Figure 3 summarizes the number of available studies for each outcome and the number of studies reporting improvements in child health status. There is convincing evidence that performance-based financing (PBF) for care facilities reduces malnutrition in the short-term and that it could have a role in preventing neonatal mortality. We found little positive effects of incentives for anemia or prevalence of acute illnesses. We next review the evidence on the channels through which P4P affects downstream health outcomes—changes in provider behavior. Finally, we summarize the available evidence on the relative cost-effectiveness of P4P.

Figure 3: The impact of P4P on child health outcomes



4.1 Child health outcomes

4.1.1 Malnutrition

Health providers often play an important role in determining child nutrition by providing nutritional supplements, good nutrition advice, or meals. As such, there is scope for P4P to alleviate malnutrition by improving the quality of these services. Evidence from India, Indonesia, Rwanda, and the Philippines suggests that incentivizing nutrition-targeted tasks has at least short-term positive impacts on child anthropometric indicators as proxies for nutrition status.

Singh and Masters (2017) conduct a randomized controlled trial of a P4P scheme among Anganwandi workers, who provide daycare services as part of India's Integrated Child Development Services (ICDS) program. Workers affect child nutrition by directly providing midday meals to children at ICDS sites and by providing nutrition counselling on-site or through home visits. Motivated by documented deficiencies in service quality, Singh and Masters (2017) test a P4P scheme which provides a bonus of about USD 3 to workers for each child whose malnourishment status improves in terms of weight-for-age. An advantage of their empirical strategy is the comparison of P4P against workers who were randomly assigned a fixed bonus of USD 3 to isolate the effect of incentives from any income effects arising from the bonuses. They found that P4P increases the average child's weight by 220 g starting at three months and persisting over a period of nine months compared to workers who receive no bonuses of any kind. They also document an increase

of 0.1 SD in weight-for-age and a decline of about 5.2 percentage points in weight-for-age malnutrition over the same time period. The study found no discernible effects of P4P against the fixed bonus group, who also have similar sized impacts on weight and weight-for-age compared to workers without bonuses. A similar experiment among 209 Integrated Child Development Services (ICDS) sites in Kolkata found that providing nutritional information to mothers as well as P4P induces nearly identical improvements in weight-for-age (0.1 SD) over three months compared to sites that receive neither bonuses nor information (Singh and Mitra. 2017).

Peabody et al. (2014) provide evidence that improvements in child nutrition can occur with physicians in a hospital setting who are not directly incentivized to address malnutrition. The Quality Improvement Demonstration Study (QIDS) included thirty hospitals in the Philippines that were randomly assigned a P4P scheme paying bonuses to physicians based on their performance in clinical vignettes (on average 5% of their salaries). They find, among other results, a 9.25 percentage point decrease in wasting among children under five who were admitted to the hospital for diarrhea or pneumonia. Improving clinical activities for diarrhea is likely to have an effect on anthropometrics since diarrheal illness can impair height and weight growth. The difference is with respect to the control group, who received no bonuses of any kind. As such, the estimates cannot distinguish between the effect of bonuses and that of the incentive alone. As in Singh and Masters (2017), there is only evidence for a *total* effect on malnutrition.

At the institutional level, Olken, Onishi, and Wong (2014) evaluate Indonesia's *Generasi* program, which provides block grants of about USD 8,500 each for villages to improve education and child and maternal health. The experiment randomly assigned subdistricts containing villages to receive either a fixed grant, for a portion of the grant to be based on villages' relative performance on twelve health and education indicators, or a control group. They report improvements in weight-for-age malnutrition by 2.6 percentage points in the incentivized villages compared to the fixed grant villages and attribute this effect to incentives apart from bonuses per se. This effect largely vanishes after 30 months as a result of fixed grant villages catching up to incentivized villages in prevalence of malnutrition.

Gertler and Vermeersch (2012) leverage the rollout of Rwanda's national PBF program which allocates funds to government-run primary care facilities based on a set of indicators for the number of child and maternal care services provided (curative care, ANC, contraceptive supply, and institutional deliveries) as well as indicators for the content of care provision. This study utilizes a difference-in-differences approach to compare facilities in program districts to facilities that received unconditional bonuses amounting to the average of PBF payments. As in the previous evaluations, this allows for the identification of incentive effects apart from differences due to financial resources. After 23 months,

infants younger than 12 months in treated districts gained a substantial 0.53 SD in weight-for-age while children between 24-47 months gained 0.25 SD in height-for-age. The results are consistent with research that suggests that in utero conditions affect height at an early age while nutrient intake determines weight.

4.1.2 Anemia

Anemia is estimated to encumber half of all school-aged children in low income countries (Hall et al. 2001). Anemia in turn is associated with poorer education outcomes likely as a result of impairment of cognition and brain function. One of the earliest evaluations of P4P to look at health outcomes is by Sylvia et al. (2013), who tested whether providing schools with bonuses tied to student prevalence of anemia could improve health and exam performance. The experiment assigned schools to a control group or to receive a fixed bonus, an information campaign, and a contract rewarding schools for reductions in the number of anemic students. The study documents an increase of 3.55 and 3.78 g/L in hemoglobin concentrations for students who are borderline anemic (< 120 g/L) and anemic (< 115 g/L) respectively at baseline. These are not statistically distinguishable from similarly sized effects of schools that received just the information and fixed bonus alone, however. In the same setting, Luo et al. (2019) conducts an experiment that assigned 300 schools varying amounts of a personal financial incentive for administrators alongside varying sizes of school budget supplements. The large incentive—amounting to 2 months of their annual salary—reduced anemia prevalence by 13.8 percentage points compared to control. Large block grants of USD 0.11 per student per day had similar effects (14.5 percentage points) and were also not distinguishable from the effect of the financial incentives.

Evidence from Rwanda and the Philippines on hospital-based performance pay is less promising with regards to anemia reduction. Neither Peabody et al. (2014) nor Sherry, Bauhoff, and Mohanan (2017) found statistically significant effects on hemoglobin concentration. The lack of an effect in Rwanda is surprising since PBF increased iron supplementation in incentivized district hospitals by 9 percentage points. That anemia has implications for a child's long-term development warrants further investigation on why and when P4P could improve child health in this dimension.

4.1.3 Morbidity

There are four well-identified studies that provide mixed evidence that incentives for hospitals and physicians can reduce child morbidity due to acute illness. The QIDS experiment in the Philippines directly incentivizes physician performance in care for diarrhea and pneumonia, which represent the greatest burden of child morbidity in low- and middle-income countries (Peabody et al. 2014). Children under five who receive

inpatient care from program hospitals are less likely to exhibit morbidity (wasting and self-reported health issues) due to diarrhea or pneumonia up to 4-10 weeks after being discharged.

Three randomized evaluations found no significant impact of PBF on child morbidity. An early study of Cambodia's Basic Health Services Project (BHSP) sought to examine the effect of two contracting arrangements for the provision of government health services (Bhushan et al. 2007). Twelve districts were randomized to have either its rural health centers managed by an NGO and staffed by government workers (contracting-in), full outsourcing of health services to NGOs (contracting-out) or retain management by the government (comparison). In both schemes contract renewal was tied to healthcare indicators and financing included individual P4P for their personnel so the study can only estimate total program impacts. They found no evidence of either program on incidence of diarrhea or reported illness, potentially due to a lack of statistical power or imbalance across treatment districts. Skiles et al. (2015) use data from the Demographic Health Survey to examine the effect of Rwanda's national PBF program on the prevalence of diarrhea, fever, and acute respiratory infection for children under five. They found no evidence that program districts experienced decreased morbidity, which the study attributes to a relatively short follow-up of six months, a broad decline in the prevalence of malaria due to the simultaneous implementation of a bed net program, and the fact that acute childhood illness are determined primarily by environmental factors. Finally, Olken, Onishi, and Wong (2014) found no effects of any type of block grant in Indonesia's *Generasi* program on prevalence of diarrhea or acute respiratory infection.

4.1.4 Mortality

Few studies have examined the effect of performance pay on infant and child mortality in part because mortality is a low probability event requiring a large enough sample to be able to statistically detect potential effects. Consequently, the three evaluations in our evidence base that do speak of child mortality are of large-scale national P4P programs. Olken, Onishi, and Wong (2014) report a 0.01 reduction of neonatal mortality after 18 months due to the fixed block grants compared to no bonuses but not the incentivized bonuses. After 30 months, this effect disappears and neonatal mortality in incentivized villages is actually 1.6 percentage points higher than in the non-incentivized villages. The authors interpret these results as possibly indicative of a multitasking problem among providers. With Cambodia's PBF scheme, Van De Poel et al. (2016) estimates a precise zero effect of performance bonuses on neonatal mortality in participating districts. They point to skepticism regarding the role of institutional deliveries—which were incentivized—in reducing infant mortality.

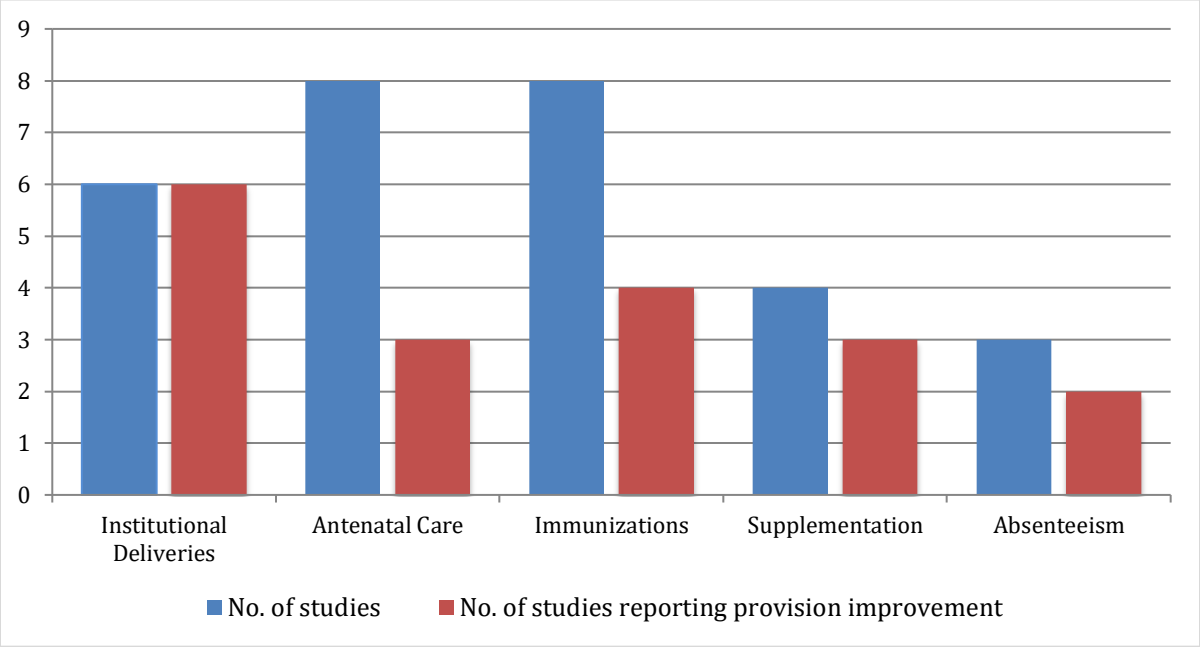
In Argentina, Gertler, Giovagnoli, and Martinez (2014) study the impact of *Plan Nacer*—a PBF scheme that allocates funding to provinces according to enrollment of beneficiaries and then to associated clinics as a budget supplement based on quality-controlled delivery of child and maternal health services. The study exploits the scale-up of the program within provinces to employ a difference-in-differences estimation of the intent-to-treat and on the treated effect. The impact of a clinic adopting performance pay results in a substantial 22 percent reduction in neonatal mortality. They attribute half of the effect to a 19% reduction in probability of low birth weight and half to improved care for low birth weight babies due to the program. The dearth of evidence for child mortality and the pattern of existing findings motivates further study.

4.2 Healthcare provision

The previous section directly examined the impact of P4P on the ultimate outcomes of interest for policy. It is also important to understand how P4P changes provider delivery of healthcare, which is the primary channel through which performance pay influences child health. Effects on provider behavior are interesting in itself as tests of the theoretical predictions outlined in Section 3. Namely, the results speak to provider multitasking, targeted provision, intrinsic motivation, and provider selection. These "first-stage" results also help interpret the pattern of effects on health outcomes.

4.2.1 Provision by service type

Figure 4: The impact of P4P on provision by service type



First, we review results on the quantity of healthcare services delivered, as this is frequently the dimension by which P4P schemes incentivize providers. The evidence base suggests PBF at the facility level yields across the board improvements for a variety of healthcare services, with the strongest evidence for institutional deliveries, immunizations, and supplementation. This is mostly consistent with our findings on malnutrition, morbidity, and neonatal mortality. Where P4P can reduce malnutrition and mortality through improved supplementation and maternal care, respectively, providers have less discretion over the determinants of child morbidity and services which depend on patient choice.

Institutional deliveries

Delivery by trained providers in health care facilities can help minimize neonatal mortality and morbidity. Evidence from PBF programs around the world suggest that institutional deliveries expand when tied to performance pay. Contracted management of public health centers in Cambodia increased the likelihood of delivery at a public facility from home by between 9 and 16 percentage points (van et al. 2016). In Rwanda, where baseline institutional delivery rates are at 30%, the national PBF program increased the likelihood of facility deliveries by on average 10 percentage points in treatment districts (Priedeman Skiles et al. 2013; Sherry, Bauhoff, and Mohanan 2017; Basinga et al. 2010). Bonfrer, Van de Poel, and Van Doorslaer (2014) examines the scale-up of PBF for government health care facilities in Burundi and reports 4 percentage point improvements in facility deliveries but only among non-poor respondents. A comparable PBF program in Zambia produces increased likelihoods for institutional deliveries with a birth attendant by 12.8 percentage points for PBF facilities and 17.5 percentage points for facilities receiving fixed bonuses (Zeng, Shepard, et al. 2018a). A comparison of these effect sizes suggest that the effect of the incentives alone is no greater than fixed increases to facility budgets.

Antenatal care

P4P has limited impact on the provision of broader preventative services in the form of antenatal care. Contracting-in during Cambodia's BHSP intervention increased the likelihood of expecting mothers receiving more than two antenatal care visits by 28 percentage points (Bhushan et al. 2007). Empirical evidence from Rwanda, however, reports no effect on antenatal care when incentivized by either PBF (Priedeman Skiles et al. 2013; Basinga et al. 2010) or for provider P4P among CHW cooperatives (Shapira et al. 2017). Evaluations of national PBF programs in Cameroon, Burundi, and Zambia also show no impact on the provision of antenatal care (De Walque et al. 2017; Zeng, Shepard, et al. 2018a; Bonfrer, Van de Poel, and Van Doorslaer 2014). The authors attribute this to a disproportionately higher cost of effort for ANC visits compared to other incentivized inputs. In Argentina, where PBF depends on birth outcomes rather than quantity of visits,

treatment clinics increase the average number of antenatal care visits by 34% (Gertler, Giovagnoli, and Martinez 2014) and that the effect persists even two years after the end of performance pay (Celhay et al. 2015). They attribute this long-term effect from a reduction in initial adjustment costs in the adoption of improved clinical practices.

Immunizations

Most studies report results on maternal immunization even when it is included as a service during antenatal care visits. As its own outcome, child and maternal immunizations can be improved by performance pay. Evaluations using routine and DHS data in Burundi document a 20-percentage point increase in the probability of receiving maternal anti-tetanus vaccination and 6.6 percentage point increase in children from poor households being fully vaccinated at one year old from PBF districts (Bonfrer, Van de Poel, and Van Doorslaer 2014; Falisse et al. 2015). These studies only estimate the total effect of funding changes but De Walque et al. (2017) and Zeng, Shepard, et al. (2018a) are able to identify positive effects of the incentive itself by comparison to fixed bonus treatment arms. Performance pay in Cameroon increases the number of completed anti-tetanus vaccinations by 69% while the program in Zambia produces no statistically significant effects on any type of immunization. Zeng, Shepard, et al. (2018b) report a surprising reduction of coverage of the third diphtheria, pertussis, and tetanus vaccinations even among improvements in other child care inputs. In Rwanda, where baseline immunization rates were already at 65%, PBF has no effects on child vaccination (Gertler and Vermeersch 2013; Basinga et al. 2011; Sherry, Bauhoff, and Mohanan 2017).

Supplementation

Nutritional supplements are complementary to adequate diets and counseling in determining child nutritional status. We expect to observe strong program effects for supplementation because it is relatively less effort-intensive than other targeted inputs. Indeed, contracted-out health centers in Cambodia document a 20-percentage point increase in receiving two doses of vitamin A in the past 12 months (Bhushan et al. 2007). PBF for clinics in the Republic of the Congo increased the number of children receiving vitamin A supplements by 155% (Zeng, Shepard, et al. 2018b). A similar program in Rwanda boosts iron supplementation by 9 percentage points (Sherry, Bauhoff, and Mohanan 2017). Individual pay for school administrators, however, was found to have no effect on supplementation despite improvements in prevalence of anemia (Luo et al. 2019).

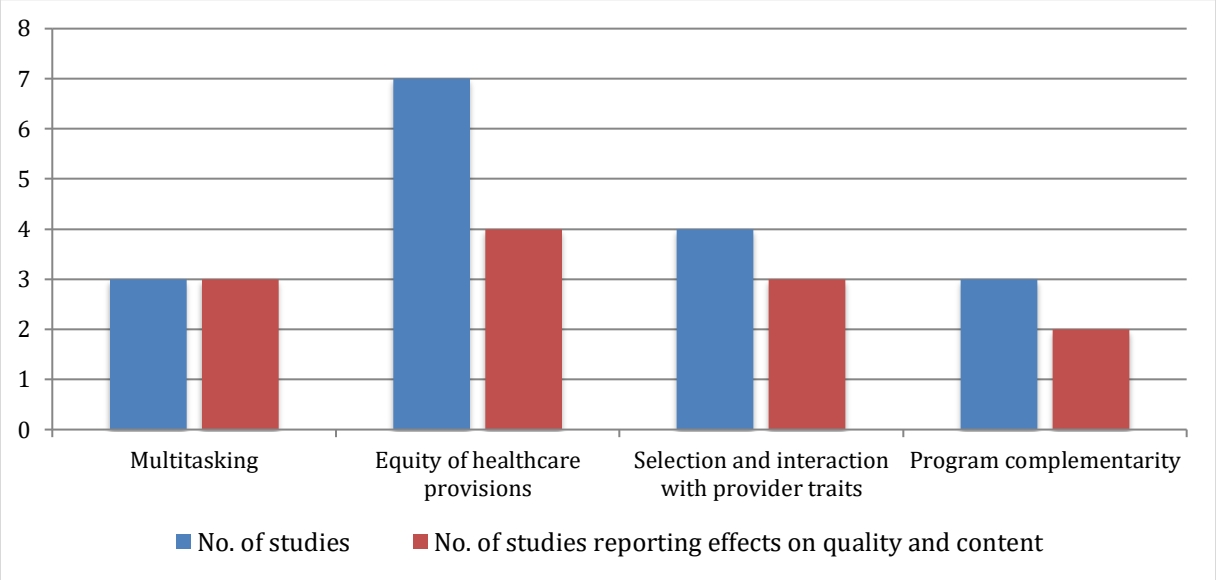
Absenteeism

Absenteeism among frontline healthcare providers is one immediate cause of deficiencies in healthcare provision (Chaudhury et al. 2006). Empirical evidence shows that absenteeism can be abated when incentive pay schemes are in place. The BHSP in

Cambodia improves the probability by an impressive 50-79 percentage points that all scheduled staff are present during unannounced visits to health centers (Bhushan et al. 2007). PBF in Rwanda increased maternity-related staff presence by 29% in treated districts (Ngo, Sherry, and Bauhoff 2017). An experiment by Banerjee, Duflo, and Glennerster (2008) highlights the importance of manipulation by staff of monitoring technologies. They found a short-term effect of monitoring with sanctions on the presence rate of ANM workers by 15% in India. In the following months, however, monitoring is undermined by nurses and their supervisors, rendering the program ineffective.

4.2.2 Quality and content of care

Figure 5: The impact of P4P on quality and content of care



Does incentivizing providers on quantity of services come at the cost of lower quality of care? A review of the evidence suggests that when health outcomes are targeted and when incentives include quality controls then P4P can improve the content of care in tandem with greater provision. Studies also demonstrate opportunities for alternative quality indicators.

Many P4P schemes explicitly account for the quality of care provision primarily by directly incentivizing quality measures or by calculating a quality multiplier for other targets. Unsurprisingly, service provision under these payment schemes are of higher quality. The PBF in Cameroon studied by De Walque et al. (2017) uses a composite quality score multiplier based on facility management, hygiene, and sanitation. They found no evidence of adverse effects on patient satisfaction and positive effects on hygiene above in addition to reported impacts on service provision. Patient satisfaction is an imperfect metric for

care quality but may be an important determinant of utilization of services. The Rwandan PBF also incorporates a quality multiplier tied to equipment and procedural criteria (Gertler and Vermeersch 2013). The program improves treatment facilities' adherence to clinical protocol (Gertler and Vermeersch 2013) as well as patient satisfaction with clinical services (Lannes 2015). Clinical vignettes offer a direct measure of provider capabilities by confronting providers with hypothetical patient cases and evaluating clinical procedures. The QIDS experiment in the Philippines ties bonuses to vignette performance and found a 10% improvement in quality scores measured this way that persists five years after the end of performance pay (Quimbo et al. 2016; Peabody et al. 2011). Such a measure proxies for provider knowledge but may be different from actual care delivered.

Even when quality controls are not incorporated into bonuses, there is little evidence of a quantity-quality tradeoff across the spectrum of different quality metrics. Data from patient exit interviews show that PBF in Cameroon increases patient satisfaction with providers during antenatal and child health consultations (De Walque et al. 2017). Experimental findings from PBF in the DRC shows that better patient satisfaction regarding facility visits is reflected in improved quality as reported by facility managers themselves (Soeters et al. 2011). The evaluation of Indonesia's *Generasi* indicates that in both incentivized and fixed grant villages, more prenatal care visits involve completion of the recommended procedures. A popular quality metric originating from the health provider literature is the "know-do gap"—the difference between the capabilities of a provider and her implementation of those skills. Gertler and Vermeersch (2013) collect data on provider knowledge using clinical vignettes and compares them with actual clinical services delivered. Providers in Rwanda knew on average 63% of clinical protocols but only implemented 45% of them. Gertler and Vermeersch (2013) estimate that performance incentives reduce this gap by 4 percentage points.

An RCT in the Democratic Republic of the Congo offers one piece of evidence that *perceived* quality of care may suffer from P4P despite improvements in provider effort. Huillery and Seban (2019) evaluate a government program that paid healthcare facilities based on the volume of patients seen for a set of predetermined services. They found that expecting women in PBF made 0.41 less prenatal visits in the past 12 months since the survey. This seems at odds with the results for providers, who on average increased attendance, held more preventative health sessions, and conducted more outreach. Huillery and Seban (2019) hypothesize that more intense promotion of health services could have sent a signal of low quality to users. This is supported by the fact that non-users in intervention areas were more likely to report that prenatal services were too far away, that they lack information, and that they perceive it to be low quality.

4.2.3 Multitasking

One way in which P4P can alter provider behavior is by changing the distribution of her effort over the multiple dimensions of provider responsibilities. In our conceptual framework, we explained how allocation of effort depends on the relative strength of incentives, the relative costs of tasks, and how multiple tasks contribute to final outcomes. This section reviews the evidence on multitasking, which draws from 1) the general pattern of effects on different healthcare services and 2) studies that observe both incentivized and non-incentivized tasks.

We found the most consistent evidence of positive impacts on institutional deliveries, maternal and child vaccination, and supplementation. This is consistent with theoretical predictions since institutional deliveries typically have the highest unit reward and that providing vaccinations and nutritional supplements require relatively less effort. Early initiation of ANC, however, requires that providers reach out to expecting mothers. Recommended completion of ANC involves patients receiving four prenatal care visits. As a result of we observe weaker impacts on this dimension.

Evidence from PBF programs in Rwanda and Cambodia observe impacts on non-incentivized tasks. They found little multitasking away from these towards paid tasks and report greater effort allocated to certain non-targeted inputs. Contracting-in health centers in Cambodia yielded an 8-percentage point increase in monthly outreach visits (Bhushan et al. 2007). Bonuses for clinics in Rwanda depended directly on incentivized services and indirectly on indicators that act as a bonus multiplier. Sherry, Bauhoff, and Mohanan (2017) report a 9.4 and 5.1 percentage point increases in iron supplementation and completed urinalysis, respectively, when those tasks enter into the quality multiplier. Ngo, Sherry, and Bauhoff (2017) for the same program attribute an 11% decrease in the monitoring of delivery statistics due to providers shifting towards better paying tasks (e.g. institutional deliveries).

Multitasking considerations suggest that there may be unintended adverse consequences to P4P. Chief among these is a concern that providers may simply manipulate indicators and monitoring systems as a response to receiving the incentive. Evidence of this behavior is scarce since P4P evaluations guard against illicit activity rather than study it. The study by Olken, Onishi, and Wong (2014) is unique in that they can compare provider immunization records against an independent measure. They found falsely recorded vaccinations increase in the short run in incentive villages by 3% compared to control villages but not compared to fixed grant villages. Banerjee, Duflo, and Glennerster (2008) examine an intervention that installed monitoring systems and sanctions to absentee ANM workers. Workers under the scheme circumnavigated sanctions by destroying record-keeping systems and claiming a greater number of exempt days. Thus, the little evidence

that is available on manipulation suggests that it remains a risk for workers under performance pay.

4.2.4 Equity of healthcare provision

Are the impacts of P4P distributed equitably or narrowly for a few beneficiaries? This section discusses the effects of performance pay for the most vulnerable, which are theoretically ambiguous since there is wider scope of improvement for those already underserved but providers may selectively target households that are relatively easier to reach. Overall, we found mixed evidence that P4P serves the poorest households.

Singh and Mitra (2017) test different varieties of individual P4P among ANM workers in ICDS centers in Kolkata. They found heterogenous effects by the degree of baseline child malnutrition: paying workers Rs. 100 per improvement in child malnutrition only occurs for children classified as mildly malnourished at baseline. The PBF scheme in Burundi increases the likelihood of completed vaccinations by age one of 6.6 percentage points but little differences in antenatal care by poverty status (Bonfrer, Van de Poel, and Van Doorslaer 2014). Olken, Onishi, and Wong (2014) report suggestive evidence from Indonesia that *Generasi* bonus grants were not associated with village poverty levels and that remote villages received more bonuses. In Rwanda, poor households are 44 percentage points more likely to receive treatment for childhood diarrhea and fever (Skiles et al. 2015). Separate studies in Rwanda, however, found that the poor are no more likely to receive ANC than the non-poor (Priedeman Skiles et al. 2013) and that those facing lower financial barriers to healthcare experience greater effects on institutional delivery and use of modern contraceptives (Lannes et al. 2016). Perhaps the clearest finding on targeted provision is due to Sylvia et al. (2013) who study a PBF program tied to anemia reduction among students in rural China. They found that PBF *reduced* the exam scores—their final outcome of interest—of children who were not anemic at baseline. The study posits that incentivizing anemia reduction could have allocated resources away from educational inputs that benefit all students towards health inputs that benefit only anemic students. Finally, a laboratory experiment with health providers in Burkina Faso gives support for no targeted provision. Banuri et al. (2018) allowed providers to undertake a real effort medical task in the form of clinical vignettes under different payment schemes. It simulates the feature that the poor may be harder to serve by allowing vignettes to present more complex information that took longer to convey. They found no targeted provision when piece rates do not depend on poverty status compared to fixed salaries and that only explicit incentives to target the poor create a more equitable delivery of services.

4.2.5 Selection and interaction with provider traits

A complete analysis of the impacts must consider how P4P can influence behavior among existing providers but can also affect the types of individuals that select into becoming providers. The only study in our evidence base that speaks to selection effects of incentives is an experiment in Zambia on government recruitment of rural health providers (Ashraf, Bandiera, and Lee 2014). They randomly assign districts into a recruitment campaign that highlights career advancement opportunities and a campaign that highlights the social benefits of the position. By keeping actual job rewards constant between these two groups, the study isolates the selection effect due to different types of recruited providers. They found that providers recruited with career prospects provide more inputs, utilization of health center services, offer higher quality of care, and, ultimately, decreases the proportion of surveyed children under five who are underweight by 5 percentage points. These observations provide strong evidence against the intuition that perhaps extrinsic rewards attract less intrinsically-motivated agents to the detriment of health outcomes.

Analysis of heterogeneous effects in P4P evaluations provides the remaining evidence on how provider characteristics can moderate program effectiveness. Overall, we found that impacts of P4P on inputs and health outcomes depend on provider skill, pro-social attitudes, and personality traits. Ashraf, Bandiera, and Jack (2014) conduct a field experiment in Zambia that tests the effect of financial versus non-financial incentives in the sales of condoms by hairstylists (acting as public health agents). They found that on average agents with non-financial rewards sell more than twice as many condoms as those with financial incentives. Additionally, both types of incentives improve sales especially for pro-socially motivated agents, which the authors interpret as evidence of crowding-in of intrinsic motivation. On the other hand, Huillery and Seban (2019) found that providers in PBF facilities in the DRC were more likely to see financial benefits as the main advantage of their position in self-reports. Gertler and Vermeersch (2013) examine another salient dimension of provider heterogeneity: skill. The study found that PBF induces a greater effect on measures of quality among providers above the median level of skill as measured by clinical vignettes. Lastly, an experiment among obstetric providers in India found that a P4P based on maternal health outcomes was less effective at reducing postpartum hemorrhages for more conscientious and more neurotic individuals at baseline (Donato et al. 2017).

4.2.6 Program complementarity

In this section we turn to a discussion about potential complementarities between P4P, other supply-side interventions, and demand-side (household) interventions. Not only does this convey which combinations of programs might produce the greatest returns, but it also identifies conditions under which we expect P4P to be more or less effective. In settings

where there is evidence, performance pay is a substitute for discretionary resources and is complementary to better service use induced by demand-side interventions.

The ideal study of program complementarities would observe units treated with just one of each program, units with both programs, and units with neither. The systematic search yielded one study that employs this design for supply-side interventions. Luo et al. (2019) reports on an experiment in China that provided schools with large or small block grants combined with incentives of varying magnitudes to school administrators to reduce student anemia prevalence. They found that a block grant of USD 2,717 fully crowds out the effect of large administrator incentives on hemoglobin concentration. That study found a similar pattern of substitution on inputs (iron supplementation and iron-rich foods) that suggests reallocation of administrator effort with larger discretionary resources.

The lesson from interactions between P4P and PBF with demand-side interventions is that healthcare use by beneficiaries is potentially important for the success of performance pay. The pilot of Cambodia's contracting scheme was differentially implemented in districts which simultaneously had in place donor-sponsored and government fee waiver schemes. Van De Poel et al. (2016) exploit this variation to examine heterogeneity of PBF impacts by accessibility of healthcare. They found that the effect of PBF on institutional deliveries is almost three times greater for districts with fee waivers than without. In addition to facility-based PBF, Rwanda also expanded performance pay to CHW cooperatives. Shapira et al. (2017) conducted an experiment that randomly assigned cooperatives to combinations of P4P and the ability to reward patients for utilization with an in-kind transfer. They found some effect of in-kind transfers on timeliness of ANC but no direct effect of P4P nor an interaction between the transfers and P4P. Shapira et al. (2017) write that this was due less to a lack of complementarity between P4P and conditional transfers per se and more to ineffectively small incentive amounts (USD 7.30 quarterly average).

5 Discussion

The objective of this review is to synthesize the empirical evidence on P4P in low- and middle-income countries in order to guide the design of performance pay. This section discusses the potential effects of P4P design choices. We only find limited evidence on the choice of targeted indicators, the level of providers, and the types of incentives. The array of empirical findings also reveal opportunities for future research that we discuss in this section.

5.1 Implications for policy design

5.1.1 Targeting inputs versus outcomes

The use of performance pay entails many design choices about who to pay, what to incentivize, and the type of incentives. We begin by considering evidence on the effect of tying incentives to provider inputs versus health outcomes, which comes exclusively from a comparison across studies. On balance, most of our evidence base consists of facility-level PBF and hospital-based P4P that incentivize the quantity and quality of healthcare services. Only four studies from two contexts on individual P4P target outcomes directly. There is opportunity for direct tests of these two design features that focus on provider innovation, targeted provision, and cost-effectiveness.

In the conceptual framework we noted how incentivizing outcomes allows providers to creatively combine inputs using local knowledge. Incentivizing outcomes is desirable when targeted agents do not have clearly defined tasks or when it is unclear a priori what combination of tasks maximize the outcome of interest. Singh and Masters (2017) and Singh and Mitra (2017) look at P4P for ANM workers, who are responsible for the provision of food, preschool education, primary healthcare, and consultations as part of ICDS. Here bonuses were tied to the reduction of the number of malnourished children. ANM workers achieved improvements in nutrition status by more frequent and more effective consultations with mothers. It would not have been feasible to contract on the content of consultations. Similarly, Sylvia et al. (2013) and Luo et al. (2019) work with school administrators who can influence health outcomes through a variety of means. When rewarded on the basis of reductions in anemia prevalence, schools would use a combination of iron supplementation, iron-rich foods, and information to students.

Another consideration for the rewarding of outcome is whether providers will target the worst-performing patients to the detriment of others. In both settings, providers were rewarded for improvements in child health status but were not sanctioned for deteriorating health. Singh and Masters (2017) shows that P4P did not worsen nutrition for normal weight children. Sylvia et al. (2013) found that exam scores worsened for students not classified as anemic but that anemia prevalence (the targeted outcome) diminished equitably.

5.1.2 Group incentives

Policymakers have a choice about what level of performance pay to provide. It can operate through hospitals and care centers as in PBF, in teams of providers, or at the individual level. It is difficult to generalize differences between PBF and P4P schemes in our evidence base to only this dimension since there are differences in scale, context, and objectives, to

say the least. Instead, we consider issues raised in our conceptual framework and addressed by three sources of evidence from Indonesia, India, Cameroon, and Rwanda.

An understudied facet of group performance pay is how it might impact the organizational dynamics of health care providers. We noted in the conceptual framework that it could affect cooperation between providers and, in turn, health outcomes and provider welfare. The PBF evaluation in Cameroon is one of the only studies to measure these variables (De Walque et al. 2017). They found no effects of PBF for health care facilities on self-reported measures of provider welfare but that providers in program facilities are 60% more satisfied with work benefits (including performance pay) compared to baseline. They report a 22% decline in good working relationships with managers significant at the 10% level. Shapira et al. (2017) studies PBF for CHW cooperatives in Rwanda and found no effects on work satisfaction or motivation. That study highlights another concern with group-level incentives that funds may be earmarked towards uses other than performance pay for individual providers. Only 30% of group bonuses went to P4P, resulting in no detectable impacts on behavior or health outcomes.

When performance pay is implemented for teams, incentives can be based on relative instead of absolute performance. Singh and Mitra (2017) directly tests absolute bonuses to bonuses based on an ANM worker's performance relative to other workers in her ICDS center. They find no effect of relative incentives compared to no P4P and absolute bonuses. Village-level performance bonuses in the *Generasi* study were based on service provision relative to other villages in the subdistrict as a way of minimizing the role of unobserved differences in regional capabilities. Olken, Onishi, and Wong (2014) show that even within subdistricts more affluent villages did not receive a disproportionately greater amount of grants.

5.1.3 Non-financial incentives

While all P4P programs in our evidence base use financial incentives, providers motivated by career advancement or social benefit might respond to non-financial incentives. Ashraf, Bandiera, and Jack (2014) tests the effect of financial incentives against displaying an agent's public sales record. Surprisingly, agents with non-financial rewards sell more than twice as many condoms as those with financial incentives and both types of incentives improve sales especially for pro-socially motivated agents. Through selection of providers at the recruitment stage, Ashraf, Bandiera, and Lee (2014) shows that the promise of career advancement selects for community health providers who provide better quality of care and more substantial reduction in child wasting down the line.

5.2 Opportunities for future research

Much progress has been made in recent years in collecting evidence for P4P in low- and middle-income countries. Indeed, all but three of the studies in our evidence base were conducted in the past decade. Beyond impacts, evidence points to promising directions for future performance pay research.

Direct evaluation of P4P design features

Direct comparisons of different P4P features within the same context are few and far between. Most of what we know about effective P4P design is a result of interpolation of effects across studies. The field is open for empirical work that can identify causal impacts of different design features. First, perhaps the most evident gap is on the effect of incentivization apart from income effects associated with bonus payments. A majority of the effects we present are the combination of incentives and better financing. It is important for forthcoming studies to control for this income effect using appropriate counterfactuals. Second, there is little research on individual P4P in low- and middle-income countries, let alone an experimental comparison to group incentives. Third, research in low income countries should catch up to evidence from developed countries on the motivating effects of non-financial incentives like career advancement. Virtually no research has been done on the role of timing and dynamic contracts for health providers despite its optimality on theoretical grounds.

Cost-effectiveness within and across programs

The value to policymaking of a review such as this is the ability to compare programs and program variants on the basis of cost-effectiveness. We report causal impacts of different interventions but can only speak about cost-effectiveness for a handful of them, due in part to a lack of reporting on program costs. For performance pay in particular, the cost of monitoring targeted indicators can vary across contexts. Future research with an eye towards policy-relevance should be transparent about program costs as a step towards standardized cost-effectiveness analysis.

Program complementarity

P4P exists alongside an assemblage of interventions aimed at lifting child and maternal health status in low- and middle-income countries. We found evidence that demand-side interventions that improve healthcare use can be a complement to P4P but there is space for further investigation in different settings and with different programs. A limitation of P4P is that it cannot address structural problems in weak health systems. As such, there is a demand for evidence regarding what systemic improvements in the supply of healthcare need to take place alongside P4P.

Focus on understudied outcomes

Existing evaluations cover impacts on the most important child health indicators and on clinically-proven healthcare practices to improve them. Nevertheless, there is scope for work on understudied outcomes. Gertler, Giovagnoli, and Martinez (2014) and Celhay et al. (2015) make significant strides in the investigation of impacts on neonatal mortality and identification of long-term effects. Future work requires empirical designs well-powered enough to detect changes in mortality and with a long enough time horizon to capture persistent effects. Finally, the literature has paid little attention to welfare effects on providers, which are important in their own right and for potential downstream consequences on healthcare provision.

References

Ashraf, Nava, Oriana Bandiera, and B. Kelsey Jack. 2014. "No Margin, No Mission? A Field Experiment on Incentives for Public Service Delivery." *Journal of Public Economics* 120 (December): 1–17. https://doi.org/10.1016/j.jpubeco.2014.06.014.

Ashraf, Nava, Oriana Bandiera, and Scott S. Lee. 2014. "Do-Gooders and Go-Getters: Career Incentives, Selection, and Performance in Public Service Delivery." *STICERD-Economic Organisation and Public Policy Discussion Papers Series* 54.

Banerjee, Abhijit V., Esther Duflo, and Rachel Glennerster. 2008. "Putting a Band-Aid on a Corpse: Incentives for Nurses in the Indian Public Health Care System." *Journal of the European Economic Association* 6 (2-3): 487–500. https://search.proquest.com/docview/56785075?accountid=14524.

Banuri, Sheheryar, B. C.M. De Walque Damien, Philip E. Keefer, and Paul J. Robyn. 2018. *Encouraging Service Delivery to the Poor: Does Money Talk When Health Workers Are Pro-Poor ?* https://search.proquest.com/docview/2160362705?accountid=14524.

Basinga, Paulin, Paul J. Gertler, Agnes Binagwaho, Agnes L. B. Soucat, Jennifer R. Sturdy, and Christel M. J. Vermeersch. 2010. *Paying Primary Health Care Centers for Performance in Rwanda*. https://search.proquest.com/docview/57013725?accountid=14524.

Basinga, Paulin, Paul J. Gertler, Agnes Binagwaho, Agnes LB Soucat, Jennifer Sturdy, and Christel MJ Vermeersch. 2011. "Effect on Maternal and Child Health Services in Rwanda of Payment to Primary Health-Care Providers for Performance: An Impact Evaluation." *The Lancet* 377 (9775): 1421–8.

Berendes, Sima, Peter Heywood, Sandy Oliver, and Paul Garner. 2011. "Quality of Private and Public Ambulatory Health Care in Low and Middle Income Countries: Systematic Review of Comparative Studies." Edited by Rachel Jenkins. *PLoS Medicine* 8 (4): e1000433. https://doi.org/10.1371/journal.pmed.1000433.

Bénabou, Roland, and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review* 96 (5): 1652–78. https://doi.org/10.1257/aer.96.5.1652.

Bhushan, Indu, Erik Bloom, David Clingingsmith, Rathavuth Hong, Elizabeth King, Michael Kremer, Benjamin Loevinsohn, and J. Brad Schwartz. 2007. "Contracting for Health: Evidence from Cambodia." *Brookings Inst*.

Bonfrer, Igna, Ellen Van de Poel, and Eddy Van Doorslaer. 2014. "The Effects of Performance Incentives on the Utilization and Quality of Maternal and Child Care in

Burundi." *Social Science & Medicine* 123 (December): 96–104. https://doi.org/10.1016/j.socscimed.2014.11.004.

Celhay, Pablo, Paul Gertler, Paula Giovagnoli, and Christel Vermeersch. 2015. *Long Run Effects of Temporary Incentives on Medical Care Productivity*. Vol. w21361. https://search.proquest.com/docview/1700661876?accountid=14524.

Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, and F. Halsey Rogers. 2006. "Missing in Action: Teacher and Health Worker Absence in Developing Countries." *Journal of Economic Perspectives* 20 (1): 91–116.

Das, Jishnu, Alaka Holla, Aakash Mohpal, and Karthik Muralidharan. 2016. "Quality and Accountability in Health Care Delivery: Audit-Study Evidence from Primary Care in India." *American Economic Review* 106 (12): 3765–99.

De Walque, Damien, Paul J. Robyn, Hamadou Saidou, Gaston Sorgho, and Maria W. Steenland. 2017. *Looking into the Performance-Based Financing Black Box: Evidence from an Impact Evaluation in the Health Sector in Cameroon*. https://search.proquest.com/docview/1934229159?accountid=14524.

Donato, Katherine, Grant Miller, Manoj Mohanan, Yulya Truskinovsky, and Marcos Vera-Hernandez. 2017. "Personality Traits and Performance Contracts: Evidence from a Field Experiment Among Maternity Care Providers in India." *American Economic Review* 107 (5): 506–10. https://search.proquest.com/docview/1904173742?accountid=14524.

Falisse, Jean-Benoît, Juvenal Ndayishimiye, Vincent Kamenyero, and Michel Bossuyt. 2015. "Performance-Based Financing in the Context of Selective Free Health-Care: An Evaluation of Its Effects on the Use of Primary Health-Care Services in Burundi Using Routine Data." *Health Policy and Planning* 30 (10): 1251–60. https://doi.org/10.1093/heapol/czu132.

Gertler, Paul, Paula Giovagnoli, and Sebastian Martinez. 2014. *Rewarding Provider Performance to Enable a Healthy Start to Life: Evidence from Argentina's Plan Nacer*. Policy Research Working Papers. The World Bank. https://doi.org/10.1596/1813-9450-6884.

Gertler, Paul, and Christel Vermeersch. 2012. *Using Performance Incentives to Improve Health Outcomes*. https://search.proquest.com/docview/1030495204?accountid=14524.

———. 2013. *Using Performance Incentives to Improve Medical Care Productivity and Health Outcomes*. Vol. w19046. https://search.proquest.com/docview/1361802906?accountid=14524.

Harris, Milton, and Bengt Holmstrom. 1982. "A Theory of Wage Dynamics." *The Review of Economic Studies* 49 (3): 315. https://doi.org/10.2307/2297359.

Holmstrom, Bengt, and Paul Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, & Organization* 7: 24–52. https://www.jstor.org/stable/764957.

Huillery, Elise, and Juliette Seban. 2019. "Financial Incentives, Efforts, and Performances in the Health Sector: Experimental Evidence from the Democratic Republic of Congo." *Economic Development and Cultural Change: 0013-0079, 1539-2988.* https://www.journals.uchicago.edu/doi/10.1086/703235.

Lannes, Laurence. 2015. "Improving Health Worker Performance: The Patient-Perspective from a PBF Program in Rwanda." *Social Science & Medicine* 138 (August): 1–11. https://doi.org/10.1016/j.socscimed.2015.05.033.

Lannes, Laurence, Bruno Meessen, Agnes Soucat, and Paulin Basinga. 2016. "Can Performance-Based Financing Help Reaching the Poor with Maternal and Child Health Services? The Experience of Rural Rwanda: Can PBF Help Reaching the Poor?" *The International Journal of Health Planning and Management* 31 (3): 309–48. https://doi.org/10.1002/hpm.2297.

Lazear, Edward P. 1981. "Agency, Earnings Profiles, Productivity, and Hours Restrictions." *The American Economic Review* 71 (4): 606–20. https://www.jstor.org/stable/1806184.

Luo, Renfu, Grant Miller, Scott Rozelle, Sean Sylvia, and Marcos Vera-Hernández. 2019. "Can Bureaucrats Really Be Paid Like CEOs? Substitution Between Incentives and Resources Among School Administrators in China." *Journal of the European Economic Association*, January. https://doi.org/10.1093/jeea/jvy047.

Miller, G., and K. S. Babiarz. 2014. "Pay-for-Performance Incentives in Low- and Middle-Income Country Health Programs." In *Encyclopedia of Health Economics*, edited by Anthony J. Culyer, 457–66. San Diego: Elsevier. https://doi.org/10.1016/B978-0-12-375678-7.00126-7.

Ngo, Diana K. L., Tisamarie B. Sherry, and Sebastian Bauhoff. 2017. "Health System Changes Under Pay-for-Performance: The Effects of Rwanda's National Programme on Facility Inputs." *Health Policy and Planning* 32 (1): 11–20. https://search.proquest.com/docview/1958569499?accountid=14524.

Olken, Benjamin A., Junko Onishi, and Susan Wong. 2014. "Should Aid Reward Performance?" *American Economic Journal. Applied Economics* 6 (4): 1–34. https://doi.org/10.1257/app.6.4.1.

Hall, A., E. Bobrow, S. Brooker, M. Jukes, K. Nokes, J. Lambo, H. Guyatt, et al. "Anaemia in Schoolchildren in Eight Countries in Africa and Asia." *Public Health Nutrition* 4, no. 3 (June 2001): 749–56. https://www.ncbi.nlm.nih.gov/pubmed/11415481

Peabody, John, Riti Shimkhada, Stella Quimbo, Jhiedon Florentino, Marife Bacate, Charles E. McCulloch, and Orville Solon. 2011. "Financial Incentives and Measurement Improved Physicians' Quality of Care in the Philippines." *Health Affairs* 30 (4): 773–81. https://doi.org/10.1377/hlthaff.2009.0782.

Peabody, John W., Stella Quimbo, Jhiedon Florentino, Riti Shimkhada, Xylee Javier, and David Paculdo. 2017. "Comparative Effectiveness of Two Disparate Policies on Child Health: Experimental Evidence from the Philippines." *Health Policy and Planning* 32 (4): 563–71. https://search.proquest.com/docview/1958565640?accountid=14524.

Peabody, John W., Riti Shimkhada, Stella Quimbo, Orville Solon, Xylee Javier, and Charles McCulloch. 2014. "The Impact of Performance Incentives on Child Health Outcomes: Results from a Cluster Randomized Controlled Trial in the Philippines." *Health Policy and Planning* 29 (5): 615–21. https://doi.org/10.1093/heapol/czt047.

Petersen, Laura A., LeChauncy D. Woodard, Tracy Urech, Christina Daw, and Supicha Sookanan. 2006. "Does Pay-for-Performance Improve the Quality of Health Care?" *Annals of Internal Medicine* 145 (4): 265. https://doi.org/10.7326/0003-4819-145-4-200608150-00006.

Priedeman Skiles, Martha, Sian L. Curtis, Paulin Basinga, and Gustavo Angeles. 2013. "An Equity Analysis of Performance-Based Financing in Rwanda: Are Services Reaching the Poorest Women?" *Health Policy and Planning* 28 (8): 825–37. https://search.proquest.com/docview/1511798790?accountid=14524.

Quimbo, Stella, Natascha Wagner, Jhiedon Florentino, Orville Solon, and John Peabody. 2016. "Do Health Reforms to Improve Quality Have Long-Term Effects? Results of a Follow-up on a Randomized Policy Experiment in the Philippines." *Health Economics* 25 (2): 165–77. https://search.proquest.com/docview/1792757642?accountid=14524.

Rosenthal, Meredith B., and Richard G. Frank. 2006. "What Is the Empirical Basis for Paying for Quality in Health Care?" *Medical Care Research and Review* 63 (2): 135–57. https://doi.org/10.1177/1077558705285291.

Rowe, Alexander K., Samantha Y. Rowe, David H. Peters, Kathleen A. Holloway, John Chalker, and Dennis Ross-Degnan. 2018. "Effectiveness of Strategies to Improve Health-Care Provider Practices in Low-Income and Middle-Income Countries: A Systematic Review." *The Lancet Global Health* 6 (11): e1163–e1175. https://doi.org/10.1016/S2214-109X(18)30398-X.

Shapira, Gil, Ina Kalisa, Jeanine Condo, James Humuza, Cathy Mugeni, Denis Nkunda, and Jeanette Walldorf. 2017. *Effects of Performance Incentives for Community Health Worker Cooperatives in Rwanda*. https://search.proquest.com/docview/1934216823?accountid=14524.

Sherry, Tisamarie B., Sebastian Bauhoff, and Manoj Mohanan. 2017. "Multitasking and Heterogeneous Treatment Effects in Pay-for-Performance in Health Care: Evidence from Rwanda." *American Journal of Health Economics* 3 (2): 192–226. https://search.proquest.com/docview/1924861058?accountid=14524.

Singh, Prakarsh, and William A. Masters. 2017. "Impact of Caregiver Incentives on Child Health: Evidence from an Experiment with Anganwadi Workers in India." *Journal of Health Economics* 55: 219–31. https://search.proquest.com/docview/1983803788?accountid=14524.

Singh, Prakarsh, and Sandip Mitra. 2017. "Incentives, Information and Malnutrition: Evidence from an Experiment in India." *European Economic Review* 93: 24–46. https://search.proquest.com/docview/1915309062?accountid=14524.

Skiles, Martha Priedeman, Siân L. Curtis, Paulin Basinga, Gustavo Angeles, and Harsha Thirumurthy. 2015. "The Effect of Performance-Based Financing on Illness, Care-Seeking and Treatment Among Children: An Impact Evaluation in Rwanda." *BMC Health Services Research* 15 (1). https://doi.org/10.1186/s12913-015-1033-7.

Soeters, Robert, Peter Bob Peerenboom, Pacifique Mushagalusa, and Célestin Kimanuka. 2011. "Performance-Based Financing Experiment Improved Health Care in the Democratic Republic of Congo." *Health Affairs* 30 (8): 1518–27. https://doi.org/10.1377/hlthaff.2009.0019.

Sylvia, Sean, Renfu Luo, Linxiu Zhang, Yaojiang Shi, Alexis Medina, and Scott Rozelle. 2013. "Do You Get What You Pay for with School-Based Health Programs? Evidence from a Child Nutrition Experiment in Rural China." *Economics of Education Review* 37: 1–12. https://search.proquest.com/docview/1511795699?accountid=14524.

Van De Poel, Ellen, Gabriela Flores, Por Ir, and Owen O'Donnell. 2016. "Impact of Performance-Based Financing in a Low-Resource Setting: A Decade of Experience in Cambodia." *Health Economics* 25 (6): 688–705. https://search.proquest.com/docview/1808035761?accountid=14524.

Witter, Sophie, Atle Fretheim, Flora L. Kessy, and Anne Karin Lindahl. 2012. "Paying for Performance to Improve the Delivery of Health Interventions in Low- and Middle-Income Countries." *Cochrane Database of Systematic Reviews*, no. 2. https://doi.org/10.1002/14651858.CD007899.pub2.

Zeng, Wu, Donald S Shepard, Ha Nguyen, Collins Chansa, Ashis Kumar Das, Jumana Qamruddin, and Jed Friedman. 2018a. "Costeffectiveness of Results-Based Financing, Zambia: A Cluster Randomized Trial." *Bulletin of the World Health Organization* 96 (11): 760–71. https://doi.org/10.2471/BLT.17.207100.

Zeng, Wu, Donald S. Shepard, Jean d. D. Rusatira, Aaron P. Blaakman, and Bernice M. Nsitou. 2018b. "Evaluation of Results-Based Financing in the Republic of the Congo: A Comparison Group Pre-Post Study." *Health Policy and Planning* 33 (3): 392–400. https://search.proquest.com/docview/2071167139?accountid=14524.

Table 2: Summary of evidence base by study

Citation	Year	Country	Intervention	Outcomes	Methodology
Ashraf, Nava, Oriana Bandiera, and B. Kelsey Jack. 2014. "No Margin, No Mission? A Field Experiment on Incentives for Public Service Delivery." Journal of Public Economics 120 (December): 1–17. https://doi.org/10.1016/j.jpubeco.2014.06.014.	2012	Zambia	Financial and non-financial incentives for condom sales	Condom sales	RCT
Ashraf, Nava, Oriana Bandiera, and Scott S. Lee. 2014. "Do-Gooders and Go-Getters: Career Incentives, Selection, and Performance in Public Service Delivery." STICERD-Economic Organisation and Public Policy Discussion Papers Series 54.	2016	Zambia	Hiring for career advancement versus social good	Household visits, community meetings, immunizations, child check-ups, institutional deliveries	RCT
Banerjee, Abhijit V., Esther Duflo, and Rachel Glennerster. 2008. "Putting a Band-Aid on a Corpse: Incentives for Nurses in the Indian Public Health Care System." Journal of the European Economic Association 6 (2–3): 487–500.	2008	India	Monitoring with penalties for absenteeism for ANMs	ANM absenteeism	RCT
Banuri, Sheheryar, B.C.M. De Walque Damien, Philip E. Keefer, and Paul J. Robyn. 2018. Encouraging Service Delivery to the Poor: Does Money Talk When Health Workers Are Pro-Poor ? https://search.proquest.com/docview/2160362705?accountid=14524.	2018	Burkina Faso	P4P for simulated diagnoses of poor/non-poor patients	Diagnostic behavior to poor/non-poor	RCT
Basinga, Paulin, Paul J. Gertler, Agnes Binagwaho, Agnes L. B. Soucat, Jennifer R. Sturdy, and Christel M. J. Vermeersch. 2010. Paying Primary Health Care Centers for Performance in Rwanda. https://search.proquest.com/docview/57013725?accountid=14524.	2010	Rwanda	P4P to facilities based on inputs	Quality of prenatal, institutional delivery, and child preventive care	RCT
Bhushan, Indu, Erik Bloom, David Clingingsmith, Rathavuth Hong, Elizabeth King, Michael Kremer, Benjamin Loevinsohn, and J. Brad Schwartz. 2007. "Contracting for Health: Evidence from Cambodia." Brookings Inst.	2007	Cambodia	Contracting services to NGO with subsidies based on maternal/child services	Service measures, child inputs, prenatal care, visits, spending	RCT
Binagwaho, Agnes, Jeanine Condo, Claire Wagner, Fidele Ngabo, Corine Karema, Steve Kanters, Jamie I Forrest, and Jean de Dieu Bizimana. 2014. "Impact of Implementing Performance-Based Financing on Childhood Malnutrition in Rwanda." BMC Public Health 14 (1). https://doi.org/10.1186/1471-2458-14-1132.	2014	Rwanda	PBF to facilities in different districts	Child height, weight	RCT
Bonfrer, Igna, Ellen Van de Poel, and Eddy Van Doorslaer. 2014. "The Effects of Performance Incentives on the Utilization and Quality of Maternal and Child Care in Burundi." Social Science & Medicine 123 (December): 96–104. https://doi.org/10.1016/j.socscimed.2014.11.004.	2014	Burundi	PBF tied to quantity of ANC, vaccinations, HIV care, family planning visits	Antenatal care, probability of institutional delivery	DD
Celhay, Pablo, Paul Gertler, Paula Giovagnoli, and Christel Vermeersch. 2015. Long Run Effects of Temporary Incentives on Medical Care Productivity. Vol. w21361. https://search.proquest.com/docview/1700661876?accountid=14524.	2015	Argentina	Provincial subsidies to improvements in maternal/child health	Prenatal care and birth outcomes	RCT
De Walque, Damien, Paul J. Robyn, Hamadou Saidou, Gaston Sorgho, and Maria W. Steenland. 2017. Looking into the Performance-Based Financing Black Box: Evidence from an Impact Evaluation in the Health Sector in Cameroon. https://search.proquest.com/docview/1934229159?accountid=14524.	2017	Cameroon	P4P and monitoring	Service utilization (child and maternal immunization, family planning, HIV testing) and quality of care	RCT

Dizon-Ross, Rebecca, Pascaline Dupas, and Jonathan Robinson. 2017. "Governance and the Effectiveness of Public Health Subsidies: Evidence from Ghana, Kenya and Uganda." Journal of Public Economics 156: 150–69.	2017	Ghana	Health worker incentive for LLIN distribution	Awareness of program, LLIN delivery	RCT
Donato, Katherine, Grant Miller, Manoj Mohanan, Yulya Truskinovsky, and Marcos Vera-Hernandez. 2017. "Personality Traits and Performance Contracts: Evidence from a Field Experiment among Maternity Care Providers in India." American Economic Review 107 (5): 506–10.	2017	India	P4P to obstetrics providers	Postpartum hemorrhages	RCT
Falisse, Jean-Benoît, Juvenal Ndayishimiye, Vincent Kamenyero, and Michel Bossuyt. 2015. "Performance-Based Financing in the Context of Selective Free Health-Care: An Evaluation of Its Effects on the Use of Primary Health-Care Services in Burundi Using Routine Data." *Health Policy and Planning* 30 (10): 1251–60. https://doi.org/10.1093/heapol/czu132Donato, Katherine, Grant Miller, Manoj Mohanan, Yulya Truskinovsky, and Marcos Vera-Hernandez. 2017. "Personality Traits and Performance Contracts: Evidence from a Field Experiment among Maternity Care Providers in India." American Economic Review 107 (5): 506–10.	2015	Burundi	PBF tied to quantity of ANC, vaccinations, HIV care, family planning visits	Vaccination of pregnant women, child vaccinations, outpatient visits, postnatal visits	DD
Gertler, Paul, Paula Giovagnoli, and Sebastian Martinez. 2014. Rewarding Provider Performance to Enable a Healthy Start to Life: Evidence from Argentina's Plan Nacer. Policy Research Working Papers. The World Bank. https://doi.org/10.1596/1813-9450-6884.	2014	Argentina	Provincial subsidies to improvements in maternal/child health	Birth outcomes and neonatal mortality	DD
Gertler, Paul, and Christel Vermeersch. 2012. Using Performance Incentives to Improve Health Outcomes. https://search.proquest.com/docview/1030495204?accountid=14524.	2013	Rwanda	P4P to facilities based on inputs	Children height-for-age, weight-for-age	RCT
Huillery, Elise, and Juliette Seban. 2019. "Financial Incentives, Efforts, and Performances in the Health Sector: Experimental Evidence from the Democratic Republic of Congo." *Economic Development and Cultural Change: 0013-0079, 1539-2988*. https://www.journals.uchicago.edu/doi/10.1086/703235.	2019	DRC	Budget allocation to facilities based on service utilization	Institutional deliveries, antenatal care, contraceptives, immunization, morbidity, child mortality, service delivery, patient/provider satisfaction	RCT
Lannes, Laurence. 2015. "Improving Health Worker Performance: The Patient-Perspective from a PBF Program in Rwanda." Social Science & Medicine 138 (August): 1–11. https://doi.org/10.1016/j.socscimed.2015.05.033.	2015	Rwanda	PBF to facilities in different districts	Patient self-reported satisfaction	RCT
Lannes, Laurence, Bruno Meessen, Agnes Soucat, and Paulin Basinga. 2016. "Can Performance-Based Financing Help Reaching the Poor with Maternal and Child Health Services? The Experience of Rural Rwanda: Can PBF Help Reaching the Poor?" The International Journal of Health Planning and Management 31 (3): 309–348. https://doi.org/10.1002/hpm.2297.	2016	Rwanda	PBF to facilities in different districts	Institutional deliveries, antenatal care, contraceptives, preventative child care	DD
Luo, Renfu, Grant Miller, Scott Rozelle, Sean Sylvia, and Marcos Vera-Hernández. 2019. "Can Bureaucrats Really Be Paid Like Ceos? Substitution Between Incentives and Resources Among School Administrators in China." Journal of the European Economic Association, January. https://doi.org/10.1093/jeea/jvy047.	2015	China	P4P to school admins to reduce anemia	Child anemia prevalence	RCT
Ngo, Diana K. L., Tisamarie B. Sherry, and Sebastian Bauhoff. 2017. "Health System Changes under Pay-for-Performance: The Effects of Rwanda's National Programme on Facility Inputs." Health Policy and Planning 32 (1): 11–20.	2017	Rwanda	P4P to facilities based on inputs	Presence of maternity related staff	RCT
Olken, Benjamin A., Junko Onishi, and Susan Wong. 2014a. "Should Aid Reward Performance? Evidence from a Field Experiment on Health and Education in Indonesia." American Economic Journal: Applied Economics 6 (4): 1–34.	2014	Indonesia	Village block grants conditional on child health outcomes	Prenatal visits, delivery by trained midwives,	RCT

				childhood immunizations, growth monitoring	
Peabody, John W., Stella Quimbo, Jhiedon Florentino, Riti Shimkhada, Xylee Javier, and David Paculdo. 2017. "Comparative Effectiveness of Two Disparate Policies on Child Health: Experimental Evidence from the Philippines." Health Policy and Planning 32 (4): 563–71.	2017	Philippines	P4P to doctors and provided insurance to patients	Prevalence of wasting in children	RCT
Peabody, John W., Riti Shimkhada, Stella Quimbo, Orville Solon, Xylee Javier, and Charles McCulloch. 2014. "The Impact of Performance Incentives on Child Health Outcomes: Results from a Cluster Randomized Controlled Trial in the Philippines." Health Policy and Planning 29 (5): 615–21. https://doi.org/10.1093/heapol/czt047.	2014	Philippines	Physician incentives tied to performance vignettes	Multiple child health outcomes	RCT
Peabody, John W., Riti Shimkhada, Stella Quimbo, Orville Solon, Xylee Javier, and Charles McCulloch. 2014. "The Impact of Performance Incentives on Child Health Outcomes: Results from a Cluster Randomized Controlled Trial in the Philippines." Health Policy and Planning 29 (5): 615–21. https://doi.org/10.1093/heapol/czt047.	2013	Rwanda	Performance-based subsidies with training	Service takeup, antenatal care, contraceptive use	DD
Quimbo, Stella, Natascha Wagner, Jhiedon Florentino, Orville Solon, and John Peabody. 2016. "Do Health Reforms to Improve Quality Have Long-Term Effects? Results of a Follow-Up on a Randomized Policy Experiment in the Philippines." Health Economics 25 (2): 165–77.	2016	Philippines	P4P to doctors and provided insurance to patients	Doctor performance vignettes	RCT
Shapira, Gil, Ina Kalisa, Jeanine Condo, James Humuza, Cathy Mugeni, Denis Nkunda, and Jeanette Walldorf. 2017. Effects of Performance Incentives for Community Health Worker Cooperatives in Rwanda. https://search.proquest.com/docview/1934216823?accountid=14524.	2017	Rwanda	P4P for cooperatives based on utilization	Antenatal care, in-facility delivery and growth monitoring of children	RCT
Sherry, Tisamarie B., Sebastian Bauhoff, and Manoj Mohanan. 2017. "Multitasking and Heterogeneous Treatment Effects in Pay-for-Performance in Health Care: Evidence from Rwanda." American Journal of Health Economics 3 (2): 192–226.	2017	Rwanda	P4P to facilities based on inputs	Institutional deliveries, contraceptive use	RCT
Singh, Prakarsh, and William A. Masters. 2017. "Impact of Caregiver Incentives on Child Health: Evidence from an Experiment with Anganwadi Workers in India." Journal of Health Economics 55: 219–31.	2017	India	P4P to caregivers based on child weight versus. fixed bonus	Child weight, home visits, center visits	RCT
Singh, Prakarsh, and Sandip Mitra. 2017. "Incentives, Information and Malnutrition: Evidence from an Experiment in India." *European Economic Review* 93: 24–46.	2017	India	P4P to caregivers based on child weight against varying amounts of fixed bonus	Caregiver behavior and child nutrition	RCT
Skiles, Martha Priedeman, Siân L. Curtis, Paulin Basinga, Gustavo Angeles, and Harsha Thirumurthy. 2015. "The Effect of Performance-Based Financing on Illness, Care-Seeking and Treatment among Children: An Impact Evaluation in Rwanda." BMC Health Services Research 15 (1). https://doi.org/10.1186/s12913-015-1033-7.	2015	Rwanda	PBF to facilities in different districts	Care received, diarrhea, fever or acute respiratory infections among children	DD
Soeters, Robert, Peter Bob Peerenboom, Pacifique Mushagalusa, and Célestin Kimanuka. 2011. "Performance-Based Financing Experiment Improved Health Care In The Democratic Republic Of Congo." Health Affairs 30 (8): 1518–27. https://doi.org/10.1377/hlthaff.2009.0019.	2011	DRC	Performance-based subsidies for health facilities	Patient payments to health centers	RCT
Sylvia, Sean, Renfu Luo, Linxiu Zhang, Yaojiang Shi, Alexis Medina, and Scott Rozelle. 2013. "Do You Get What You Pay for with School-Based Health Programs? Evidence from a Child Nutrition Experiment in Rural China." Economics of Education Review 37: 1–12.	2013	China	Incentives for school principals to reduce anemia	Student anemia prevalence, test score	RCT

Van De Poel, Ellenvan, de P., Gabriela Flores, Por Ir, and Owen O'Donnell. 2016. "Impact of Performance-Based Financing in a Low-Resource Setting: A Decade of Experience in Cambodia." Health Economics 25 (6): 688–705.	2016	Cambodia	Contracting services to NGO with subsidies based on maternal/child services	Institutional births, child vaccination, birth spacing, antenatal care.	DD
Wu, Bingxiao. 2014. Mismeasurement in Pay-for-Performance: Evidence from an Intervention to Reduce Health Care Spending in China. https://search.proquest.com/docview/1566806965?accountid=14524.	2014	China	Financial penalty to hospitals with high drug expenditure	Drug/non-drug expenditures.	DD
Zeng, Wu, Marion Cros, Katherine D. Wright, and Donald S. Shepard. 2013. "Impact of Performance-Based Financing on Primary Health Care Services in Haiti." Health Policy and Planning 28 (6): 596–605.	2013	Haiti	Performance-based subsidies with training	Service provision for children and pregnant women	DD
Zeng, Wu, Donald S Shepard, Ha Nguyen, Collins Chansa, Ashis Kumar Das, Jumana Qamruddin, and Jed Friedman. 2018. "Cost–Effectiveness of Results-Based Financing, Zambia: A Cluster Randomized Trial." Bulletin of the World Health Organization 96 (11): 760–771. https://doi.org/10.2471/BLT.17.207100.	2018	Zambia	Results-based financing (increased funding tied to performance on pre-agreed indicators), input-based financing (increased funding not tied to performance)	Financial costs for programme implementation and verification, consumables and supervision. Quality of maternal/child services.	RCT
Zeng, Wu, Donald S. Shepard, Jean d. D. Rusatira, Aaron P. Blaakman, and Bernice M. Nsitou. 2018. "Evaluation of Results-Based Financing in the Republic of the Congo: A Comparison Group Pre-Post Study." Health Policy and Planning 33 (3): 392–400.	2018	ROC	Performance-based subsidies for health facilities	Curative visits, patient referrals, children receiving vitamin A, HIV testing of pregnant women and assisted deliveries. Third diphtheria, pertussis, and tetanus immunization.	DD

*

Appendix

Evidence on cost-effectiveness

The review has thus far summarized the causal impacts of P4P on child health outcomes and provider behavior. Policy-relevant comparisons of different P4P schemes against each other and against alternative child health interventions ultimately require a cost-effectiveness analysis. This section summarizes current knowledge on cost-effectiveness and draws on the few studies in our evidence base that directly compares interventions on cost-effectiveness and comparisons across studies which calculate cost-effectiveness of evaluated programs. Our review of cost-effectiveness is (further) limited in that it only includes studies that observe health outcomes and that comparisons are made difficult by contextual differences. Nevertheless, the evidence points to the cost-effectiveness of individual P4P over a demand-side intervention improving access to healthcare. At the healthcare facility level, increasing operational budgets may be more cost-effective than PBF tied to service provision.

The best evidence we have on cost-effectiveness of P4P reports quality-adjusted life years (QALYs) gained or disability-adjusted life years (DALYs) reduced per dollar spent on programs. Peabody et al. (2017) estimates cost-effectiveness of two policies implemented by the QIDS experiment in the Philippines: universal health coverage for children under 5 in treated areas and P4P for physicians treating children under 5 based on quality of care. They estimate that both programs reduce wasting by about 9% but that P4P results in a larger absolute effect due to only a 32% utilization rate among covered households. With comparable programmatic costs, the study estimates 1.52 and 0.5 DALYs averted per incremental dollar due to P4P and coverage, respectively, and concludes that performance pay was more cost-effective than providing universal health coverage. This highlights that cost-effectiveness alone is not a sufficient metric since it doesn't reflect the heterogeneous impacts of a program or that one program compared to another might reach different populations. Gertler, Giovagnoli, and Martinez (2014) translates improvements in in-hospital neonatal mortality and low birth weight from Argentina's *Plan Nacer* PBF program into roughly 0.001 DALYs averted per incremental dollar (USD 814 per additional DALY averted). Zeng, Shepard, et al. (2018a) tests PBF in Zambia against funding bonuses not tied to performance. They estimate 0.001 QALYs gained per incremental dollar (USD 809 per QALY gained) and 0.002 QALYs per dollar (USD 413 per QALY gained) for PBF and unconditional bonuses, respectively. While there were greater health benefits in PBF districts, they estimate that funding for inputs not tied to performance remain more cost-effective.