American Economic Journal: Applied Economics 2021, 13(4): 34–69
https://doi.org/10.1257/app.20190220

Different Strokes for Different Folks? Experimental Evidence on the Effectiveness of Input and Output Incentive Contracts for Health Care Providers with Varying Skills†

By Manoj Mohanan, Katherine Donato, Grant Miller, Yulya Truskinovsky, and Marcos Vera-Hernández*

A central issue in designing incentive contracts is the decision to reward agents' input use versus outputs. The trade-off between risk and return to innovation in production can also lead agents with varying skill levels to perform differentially under different contracts. We study this issue experimentally, observing and verifying inputs and outputs in Indian maternity care. We find that both contract types achieve comparable reductions in postpartum hemorrhage rates, but payments for outputs were four times that of inputs. Providers with varying qualifications performed equivalently under input incentives, while providers with advanced qualifications may have performed better under output contracts. (*JEL* D82, D86, I12, J13, J16, J41, O15)

Performance incentives have long been used to correct a range of principal-agent problems (Jensen and Murphy 1990, Hall and Liebman 1998, Lazear 2000, Roland 2004, Rosenthal et al. 2004). A central issue in the design of performance

*Mohanan: Sanford School of Public Policy, Duke University (manoj.mohanan@duke.edu); Donato: Boston Consulting Group (donatoke@gmail.com); Miller: Stanford University and NBER (ngmiller@stanford.edu); Truskinovsky: Department of Economics, Wayne State University (yulya.truskinovsky@wayne.edu); Vera-Hernández: Department of Economics, University College London and IFS (m.vera@ucl.ac.uk). Donato, Miller, Truskinovsky, and Vera-Hernandez contributed equally to the manuscript. Neale Mahoney was coeditor for this article. This research was made possible by funding and support from 3ie and DFID-India (OW2: 205 co-PIs: Mohanan and Miller), World Bank HRITF (TF099435: PI Mohanan), and Government of Karnataka. We are grateful, for comments and suggestions, to Alessandra Voena, Alessandro Tarozzi, Amar Hamoudi, Duncan Thomas, Erica Field, Jerry La Forgia, Jishnu Das, Imran Rasul, Meredith Rosenthal, Michael Callen, Nava Ashraf, Neale Mahoney, Neeraj Sood, Oriana Bandiera, Paul Gertler, Rohini Pande, Rob Garlick, Seema Jayachandran, Victoria Baranov, Xiao Yu Wang, and anonymous reviewers as well as to audiences at AEA/ASSA 2017, ASHEcon 2016, Barcelona GSE 2016, BREAD/CEPR 2016, Duke, Erasmus, FLAME, Harvard, iHEA Congress Milan, Tilburg, and University of Southern California. Manveen Kohli provided excellent project management. We are thankful to Kultar Singh, Swapnil Shekhar, and Anil Lobo from Sambodhi, as well as the field team for project implementation and data collection. We gratefully acknowledge the support we received from World Bank (Paolo Belli, Patrick Mullen, and Vikram Rajan) and the Government of Karnataka (Vandita Sharma, Selva Kumar, Suresh Mohammed, Raghavendra Jannu, Atul Tiwari, Dr. Nagaraj, Dr. Sridhar, Dr. Prakash Kumar, Dr. Amruteshwari, and several others). We are especially grateful to the many doctors and clinical experts who provided valuable guidance and feedback, including Matthews Mathai, Dinesh Agarwal, Ayaba Worjolah, Vinod Paul, Sharad Iyengar, Kirti Iyengar, Amarjit Singh, Suneeta Mittal, Lalit Baveja, and Sunesh Kumar. This study was approved by Duke University Office of Human Subjects Research (Pro00031046) and registered in the American Economic Association Registry for randomized control trials under trial number AEARCTR-0000179.

†Go to https://doi.org/10.1257/app.20190220 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

incentives is whether to reward an agent's use of inputs or instead to reward outputs directly (Khalil and Lawarrée 1995, Prendergast 2002, Raith 2008, Prendergast 2011).[1] A key underlying assumption in models of output incentive contracts is that workers know the production function, and if correctly incentivized, can find the optimal combination of inputs to produce a given level of output. However, if workers have low levels of human capital or skills, this assumption might be untenable, leading to suboptimal input choices. Alternatively, agents' knowledge of the production function is less relevant for them to be rewarded according to input use (following explicit guidelines given by better-informed principals). But an important drawback of input incentive contracts is that highly skilled agents might be penalized if they take advantage of local contextual information that might be relevant when choosing optimal input combinations to achieve desired outputs.

In this paper, we study input and output incentive contracts for maternal health care in India—and the ways in which health providers with varying levels of skill respond to them—through a field experiment. The production of maternal and neonatal health is complex, and in India, there is considerable scope for improvement in performance and substantial variation in health providers' human capital. We randomly assign private maternal health care providers to two treatment arms and a control arm. Providers in treatment arms were given incentive contracts rewarding performance either for input use (adherence to guidelines for best practices) or production of outputs (good maternal and neonatal health—low levels of postpartum hemorrhage, sepsis, preeclampsia, or neonatal death). We also study how responses to performance contracts vary by levels of skills by comparing performance of providers with advanced medical training to those with basic medical training. We focus on providers' implementation of new strategies in the two contracts and how the effect of these innovations varies by providers' skill level.

We conducted the experiment in rural areas of Karnataka, an Indian state with poor levels of maternal and neonatal health in 2012–2014. Karnataka's maternal mortality rate (MMR) was 144 deaths per 100,000 live births, and its neonatal and infant mortality rates were 22 and 31 per 1,000 live births, respectively (Government of Karnataka 2016, Mony et al. 2015).[2] The top three causes of maternal mortality are postpartum hemorrhage, preeclampsia, and sepsis, and the major risk factors for neonatal mortality are infections (sepsis and tetanus, for example), preterm births, and birth asphyxia. Policy efforts to improve maternal and neonatal health outcomes have long focused on promoting childbirth in medical facilities (rather than in private homes), where many of these causes can—in principle—be prevented or managed. However, despite rapidly rising institutional delivery rates (reaching 94.3 percent in 2015–2016) (Paswan et al. 2016), poor maternal and neonatal health outcomes persist because of low quality maternal health care in medical facilities (Nair and Panda 2011).

[1] The decision depends on a variety of considerations, including the costs of monitoring inputs relative to outputs; the degree of risk aversion of agents relative to principals; the riskiness of the output measure; the degree to which productivity is heterogeneous across agents; the extent of distortions due to multi-tasking; and the feasibility of principals dictating inputs to agents.

[2] For comparison, India's average MMR in 2013 was estimated to be 190, and the infant and neonatal mortality rates were estimated at 28 and 40, respectively (United Nations 2019)

The quality of health services (both public and private) in developing countries is generally low (Das et al. 2012, Das and Hammer 2014, Mohanan et al. 2015, Das et al. 2016), and the use of performance incentives is increasingly widespread (see Finan, Olken, and Pande 2015 and Miller and Babiarz 2014 for reviews). Output incentives are more common in the education sector,[3] while incentives based on service delivery indicators[4] such as institutional deliveries, delivery of prenatal care, vaccinations, and healthcare utilization are typically used in the health sector.[5,6,7] The predominance of input incentive contracts in the health sector—an environment in which there is often considerable scope for innovation using local/contextual information[8]—underscores the importance of empirical research comparing contractual bases in health.

On average, we find that providers in both the input and output contract arms achieved similar improvements in maternal health, reducing rates of postpartum hemorrhage (PPH—the leading cause of maternal mortality both in India and globally) by approximately 21 percent. Performance on other dimensions of maternal and neonatal care (preeclampsia, sepsis, and neonatal survival) did not change in either contract group relative to the control group.[9] In achieving PPH reductions, providers in both groups used similar strategies (and similar input combinations), focusing on stocking medicines that reduce bleeding after delivery, for example. We also find little evidence that output contract providers developed or implemented novel strategies on average to improve outcomes despite having the flexibility to do so. Moreover, despite equivalent PPH reductions in both contract groups, input contract payments were substantially smaller than output ones: average payments

[3] Examples of empirical research on output incentives in education include Lavy (2002); Glewwe, Ilias, and Kremer (2010); Fryer (2011); Muralidharan and Sundararaman (2011); Behrman et al. (2015).

[4] Examples of research on service delivery indicators in health include Basinga et al. (2011); Soeters et al. (2011); Miller et al. (2012); Gertler and Vermeersch (2013); Ashraf, Bandiera, and Jack (2014); Gertler, Giovagnoli, and Martinez (2014); Miller and Babiarz (2014); Olken, Onishi, and Wong (2014); Celhay et al. (2015); Dupas and Miguel (2016); Sherry, Bauhoff, and Mohanan (2017).

[5] There have been few efforts to directly reward health outcomes in developing countries. Two recent exceptions in China and India study interventions outside the medical care system, focusing on childhood malnutrition. Primary school principals in China, who were offered performance incentives for reducing anemia, were able to reduce anemia prevalence by 25 percent by the end of the academic year (Miller et al. 2012, Luo et al. 2015). In India, Singh (2015) found that frontline workers in India's Integrated Child Development Services (ICDS) program who were offered high levels of incentives were able to reduce severe malnutrition by 6.3 percentage points. The Plan Nacer program in Argentina introduced performance incentives based on 10 indicators, of which two were outcomes (birth weight and APGAR scores) and the remaining 8 were self-reported/administrative service delivery indicators (Gertler, Giovagnoli, and Martinez 2014).

[6] Fritsche, Soeters, and Meessen (2014) report that the World Bank's health results trust fund, which supports performance based financing programs in health, had over 60 projects at various stages of development. Other examples of performance incentives in developing countries include Basinga et al. 2011, Peabody et al. 2011, Soeters et al. 2011, and Van de Poel et al. 2016.

[7] A recent high-profile health sector example is the Better Birth trial—a RCT promoting adherence to the "Safe Childbirth Checklist" in rural India, finding some improvement in provision of inputs on the checklist but no significant improvement in health outcomes (Semrau et al. 2017).

[8] See http://www.innovationsinhealthcare.org/ for examples of efforts that adopt novel approaches to improving access to care and improving quality of health care.

[9] Note that of the four outcomes, two (preeclampsia and neonatal mortality) were included primarily to address concerns about multitasking (i.e., to minimize reductions in effort focused on these conditions). However, providers can do relatively little to prevent preeclampsia, and there are relatively many nonmedical factors that contribute to neonatal mortality.

for input and output contracts were 13,850 and 56,812 rupees, respectively (about US$252 and $1,033 in 2010).[10,11]

The response of providers with varying levels of skills to input- and output-incentive contracts differed in magnitude although this difference was estimated imprecisely. High-skill providers with advanced medical training in obstetrics and gynecology in output contracts stated that they had implemented new health delivery strategies. They also produced better health outcomes, reducing PPH rates by 11 percentage points relative to lower-skilled providers with basic medical training who reduced it by less than 1 percentage point; the difference, however, was just shy of statistical significance ($p = 0.057$). There were no large observed differences in implementation of new strategies or health outcomes (PPH reduction of 5.6 percentage points, not statistically significant) between high- and low-skilled providers in the input contracts group.

We also investigate two potential concerns with our study. First, because we reward providers according to contracted outcomes among their patients, providers could potentially manipulate the composition of their patients rather than improving their performance (selecting patients more likely to experience good health outcomes, for example). To address this concern, incentive contracts were explicitly structured to be nullified if providers diverted risky patients, and we collected population surveillance data to test for patient selection; we do not find evidence of providers in treatment arms referring high risk patients away to other hospitals. Second, a natural concern with performance incentives is the possibility of "multitasking," or the diversion of effort from unrewarded outcomes to rewarded ones (Holmstrom and Milgrom 1991; Prendergast 1999; Mullen, Frank, and Rosenthal 2010). To minimize the possibility of effort diversion, our incentive contracts covered all major inputs and outputs involved in maternity care including neonatal health, and maternity care was deliberately chosen as a relatively narrow area of medical practice.[12]

Our paper makes two key contributions to existing literature. First, because we purposefully designed our study to observe and verify both input use (beyond what is ordinarily possible in real-world settings) and outputs, we are able to test the effectiveness of input and output incentive contracts.[13] While there is a wide theoretical literature on this topic (Jensen and Murphy 1990, Holmstrom and Milgrom 1991, Khalil and Lawarrée 1995, Hall and Liebman 1998, Prendergast 1999, Lazear 2000, Prendergast 2002, Laffont and Martimort 2002, Prendergast

[10] Our incentive contracts were not specifically designed to achieve identical levels of outcomes, since the underlying production function was unknown. The identical levels of performance in the two treatment arms is only a convenient accident that now enables us to directly compare the cost to the principal of these two types of contracts. However, since we do not observe providers' responses across a full range of rates for rewards, we are unable to draw inferences about efficiency of the two contract structures.

[11] The contracts offered providers the potential to earn up to approximately 150,000 rupees (about US$2,700 at the time of the contract—slightly more than 15 percent of a specialist doctor's salary in Karnataka). Details in Section II.

[12] The restricted scope of pregnancy and maternity care was also a rationale for selecting obstetric providers for our study. Although obstetric care providers typically refer neonatal care to pediatricians, many of the providers in our sample are the only healthcare providers in their area. Hence we include neonatal health outcomes in the contracts to minimize concerns of multitasking.

[13] We collect detailed information on inputs, using 48 indicators for 5 key domains of medical care delivered to mothers and their infants throughout pregnancy, delivery, and postnatal care.

2011), the empirical literature that explores the relative effectiveness of contracting on inputs versus outputs remains thin. To the best of our knowledge, this paper is the first to empirically compare the performance of agents under input and output contracts in a health care setting.[14] Second, we study the important role of provider training and skill in differential behavioral responses to each type of contract. By focusing on how agents' performance varies by level of human capital, we extend the growing literatures on the impact of performance incentives and optimal contracts (Callen et al. 2015).

The rest of the paper proceeds as follows: Section I provides a simple conceptual framework of input and output contracts, followed by details of the study design, data collection, and analysis in Section II. Section III presents results, including mechanisms that might explain our findings, and Section IV concludes.

I. Conceptual Framework

In this section, we outline a basic principal-agent framework to elucidate the trade-offs between input and output contracts and the role played by agent skill. In our setup, a principal (health authority) hires an agent (health care provider) to maximize health, y, net of monetary costs paid to the agent, w. A health care provider produces health according to $y = h(\theta_1 e_1, \theta_2 e_2, \varepsilon)$, where e_1 and e_2 are inputs chosen by the provider; θ_1 and θ_2 are productivity shifters; and ε is a random component with cumulative distribution function G_ε, which is strictly positive for all values of ε. The health production function $h(\cdot)$ is increasing in all of its arguments and strictly concave. The provider's objective is to maximize utility from payments, $U(w)$, net of input costs, $v_1(e_1) + v_2(e_2)$, where $U'(\cdot) > 0$, $U''(\cdot) < 0$, $v_1'(\cdot) > 0$, $v_2'(\cdot) > 0$, $v_1''(\cdot) > 0$, $v_2''(\cdot) > 0$.

We assume that there are two type of providers, H and L, with high and low levels of medical training (qualifications) corresponding to high and low levels of (clinical) skills. The proportion of low-skill providers among all providers is π. All providers have the same value of θ_1, but there is a distribution on the values that θ_2 can take, which is given by F_{θ_2}. However, H and L perceive a different value: $\theta_2^k = \rho_k \theta_2$, $k \in \{H, L\}$. We assume the support of θ_2 to be bounded: $\theta_2^L \in [\underline{\theta}_2^L, \overline{\theta}_2^L]$, and $\theta_2^H \in [\underline{\theta}_2^H, \overline{\theta}_2^H]$.

We assume that both input choices (e_1, e_2) and output (y) are verifiable. The modelling challenge is then why inputs contracts are not always optimal. We achieve this by restricting the principal's ability to make contracts contingent on the productivity shifter, θ_2. That is, principals know the distribution of productivity shifter θ_2, F_{θ_2}, but they cannot make contracts with agents contingent on the values of this productivity shifter. In other words, principals cannot take advantage of local/contextual

[14] While our paper is, to our knowledge, the first to test the effect of input and output incentives contracts in health, our work is also similar in spirit to Sarojini Hirshleifer's (2017) paper on incentives offered to students for inputs or outputs in the context of education. Hirshleifer finds larger improvements in the inputs arm, and that the input incentives were almost twice as cost effective. In addition to the differences between health and education settings, another key difference is that we study the effect of rewarding agents (doctors) for performance related to outcomes of patients, while Hirshleifer studies the effect of rewards offered to students based on their own inputs or outputs.

information, reflected in θ_2, when writing the contracts because it would be too costly to observe and verify each agent's θ_2. Unlike the standard setting in which a contract can be made fully contingent on agents' production functions, this restriction implies that input incentive contracts will not necessarily be optimal (even if inputs are verifiable) and output incentive contracts can be more efficient (Khalil and Lawarrée 1995; Prendergast 2002, 2011).

An input incentive contract is a function $w(e_1, e_2)$ that remunerates providers according to input levels. To obtain the optimal input incentive contract, the principal will choose $w(e_1, e_2)$ to solve

$$\max \iint h\left(\theta_1 e_1^*, \theta_2 e_2^*, \varepsilon\right) \partial F_{\theta_2} \partial G_{\varepsilon} - w\left(e_1^*, e_2^*\right)$$

subject to

$$\left\{e_1^*, e_2^*\right\} \in \max_{\{e_1, e_2\}} \left[U\left(w\left(e_1, e_2\right)\right) - v_1\left(e_1\right) - v_2\left(e_2\right)\right],$$
$$U\left(w\left(e_1^*, e_2^*\right)\right) - v_1\left(e_1^*\right) - v_2\left(e_2^*\right) \geq \bar{U},$$

where $\bar{U}$ is the provider's reservation utility. Note that, conditional on θ the provider does not bear any financial risk because payment is only contingent on input levels, which are completely under his/her control. Also, both high- and low-skill providers will choose the same input levels because both maximize the same function, $U(w(e_1, e_2)) - v_1(e_1) - v_2(e_2)$, which is independent of health outcomes produced—and hence their beliefs about the health production function.[15] Under input incentive contracts, average health outcomes, $y = \iint h(\theta_1 e_1, \theta_2 e_2, \varepsilon) \partial F_{\theta_2} \partial G_{\varepsilon}$, are therefore also the same for high- and low-skill providers.

Note that our model does not consider that intrinsic or prosocial motivation might be crowded out (Deci and Ryan 1985, Fehr and Falk 2002, Bénabou and Tirole 2006) because we cannot study them empirically. However, there are a number of papers that have shown their relevance in the provision of care, especially when incentives are lower powered (Kolstad 2013; Ashraf, Bandiera, and Jack 2014; Luo et al. 2020; Ashraf et al. 2020).

An output incentive contract is a function $w(y)$ that remunerates providers according to health outcomes produced. For the optimal output incentive contract, the Principal will choose $w(y)$ such that it maximizes

$$\pi \iint \left[h\left(\theta_1 e_1^*\left(\rho_L \theta_2\right), \theta_2 e_2^*\left(\rho_L \theta_2\right), \varepsilon\right) - w\left(h\left(\theta_1 e_1^*\left(\rho_L \theta_2\right), \theta_2 e_2^*\left(\rho_L \theta_2\right), \varepsilon\right)\right)\right] \partial F_{\theta_2} \partial G_{\varepsilon}$$
$$+ (1-\pi) \iint \left[h\left(\theta_1 e_1^*\left(\rho_H \theta_2\right), \theta_2 e_2^*\left(\rho_H \theta_2\right), \varepsilon\right) - w\left(h\left(\theta_1 e_1^*\left(\rho_H \theta_2\right), \theta_2 e_2^*\left(\rho_H \theta_2\right), \varepsilon\right)\right)\right] \partial F_{\theta_2} \partial G_{\varepsilon},$$

[15] This is true because we are assuming that providers are not altruistic. In other words, they will not provide additional, unrewarded inputs that they know to be beneficial if not compensated for doing so.

subject to $\forall K$ in $\{L,H\}$,

$$\left\{e_1^*\left(\theta_2^K\right), e_2^*\left(\theta_2^K\right)\right\} \in \arg\max \int U\Big(w\Big(h\big(\theta_1 e_1, \theta_2^K e_2, \varepsilon\big)\Big)\Big)\,\partial G_\varepsilon - v_1(e_1) - v_2(e_2),$$

$$\int U\Big(w\Big(h\Big(\theta_1 e_1^*\big(\theta_2^K\big), \theta_2^K e_2^*\big(\theta_2^K\big), \varepsilon\Big)\Big)\Big)\partial G_\varepsilon - v_1\Big(e_1^*\big(\theta_2^K\big)\Big) - v_2\Big(e_2^*\big(\theta_2^K\big)\Big) \geq \bar{U},$$

$$\theta_2^K \in \left[\underline{\theta}_2^K, \bar{\theta}_2^K\right].$$

The first set of constraints represent the decision problem of the agents. They choose effort levels in order to maximize their expected utility net of effort, taking into account what they perceive the productivity shifter to be (θ_2^L and θ_2^H for L and H respectively). Hence, the agents' optimal effort choices are a function of their perceived productivity shifters ($\{e_1^*(\theta_2^L),\ e_2^*(\theta_2^L)\}$ and $\{e_1^*(\theta_2^H),\ e_2^*(\theta_2^H)\}$). The two last constraints ensure that the agent's optimal effort choices provide them at least with their reservation utility. The principal maximizes the expected health output net of payments to the agents across the continuous of agents (first two lines of the problem). Note that the principal computes the expected health outcome using the correct productivity shifter (θ_2 multiplies e_2^*), but that agents' effort choices are a function of what they perceive the productivity shifter to be (because $\rho_H\theta_2 = \theta_2^H$ and $\rho_L\theta_2 = \theta_2^L$, we represent $e_1^*(\theta_2^L)$ as $e_1^*(\rho_L\theta_2), e_1^*(\theta_2^H)$ as $e_1^*(\rho_H\theta_2)$, $e_2^*(\theta_2^L)$ as $e_2^*(\rho_L\theta_2)$, and $e_2^*(\theta_2^H)$ as $e_2^*(\rho_H\theta_2)$).

It is useful to compare agents' effort choices under an input and output contract. Under an input contract, the agents' choices do not depend on what they perceive the productivity to be. The principal circumvents agent's choice of inputs by paying directly for inputs. On the contrary, under an output incentive contract, agents' choice of inputs do depend on their perceived value of the productivity shifter. Intuitively, as they consider what input mix to choose to maximize their surplus (which depends on the output achieved through the incentive contract), they must consider what output they will obtain with different input levels. This will be a key implication that we will be testing in our empirical analysis.

This framework is also useful to discuss whether an input or output contract would be optimal depending on the information held by the principal and agent.[16] Whilst they impose risk on risk-averse agents, output contracts have the advantage that agents can use information on their own θ_2 to make optimal input choices that best suit them (the principal cannot make agent-specific optimal input choices because she cannot make the contract contingent on each agent's value of θ_2, which is a limitation of the input contract). On the contrary, if both agents have sufficiently incorrect information on θ_2, (both ρ_L and ρ_H are sufficiently different from one), then an input contract would tend to be optimal because it is less risky for the agents, and it prevents the agents from making wrong input choices led by their wrong perceptions of the productivity shifter, θ_2. Hence, the principal might be better off

[16] We do not consider in our model the case where input and output contracts could be used to screen providers. It is possible that, when used as a screening device, such contracts might differentially attract providers with different levels of capacity and information. However, our study was not designed to study such a screening mechanism. Instead, we focus on relative effectiveness of input or output contracts that could be implemented by the government.

making input choices on behalf of the agents (through an input contract) which are optimal "on average," although it disregards improvements that can achieved by tailoring the inputs used to the agent's specific θ_2. If the H agent has correct information ($\rho_H = 1$) and the L agent has sufficiently incorrect information (ρ_L sufficiently different from 1), then whether an input or output contract is optimal will depend, among other things, on the shares of each type of agents. Our model examines the effect of performance incentive contracts conditional on average levels of intrinsic motivation. Because we had not collected data on intrinsic motivation of providers, we are unable to test heterogeneity of effects on productivity.

A testable implication of our conceptual framework is that health outcomes will depend on provider skills under output incentive contracts (with better health outcomes for more skilled providers), but that health outcomes will be independent of provider skill with input incentive contracts. Moreover, we expect higher-skilled providers under output contracts to tailor their input choices to their local/contextual information.

II. Study Design, Incentive Contract Structure, Data Collection, and Estimation

A. *Design and Implementation of the Experiment*

Our experiment and data collection activities spanned two years, from late 2012 to late 2014.[17] The timeline of the project is shown in Figure 1, with details about when data were collected indicated at the bottom, and timing of the intervention visits indicated at the top.

Eligibility of Providers.—Using multiple data sources, we identified the potential universe of private obstetric care providers for inclusion in our study. The first source was data collected by the Karnataka state government on all private sector doctors who provided obstetric care (i.e., those who cared for pregnant mothers and conducted deliveries) in rural areas—at least 10 km away from district headquarters. Second, during field visits by our enumerators to verify these providers, our field teams located additional providers who were inadvertently missed in the government survey and conducted interviews with them to confirm eligibility. Further eligibility for providers' inclusion in our study was based on conducting at least two deliveries per month, practicing primarily in OBGYN clinics, willingness to participate in the study (including responding to surveys and signing the incentive contracts), and continuing to practice in the same location over the study period. Providers working in large multispecialty hospitals were not included in our sample. We targeted smaller facilities in order to ensure that providers would have sufficient agency over their facilities' health provision.

[17] This study was approved by Duke University Office of Human Subjects Research (Pro00031046). Details of the study design and analysis plan (American Economic Association Registry trial number AEARCTR-0000179) were registered on the AEA registry (Mohanan and Miller 2016).

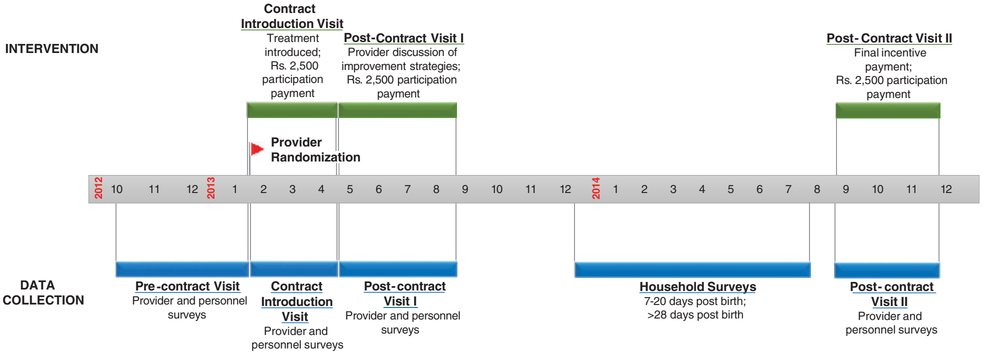

Figure 1. Timeline of Interventions and Data Collection

Notes: The timeline shows study implementation period from October 2012 to November 2014. The timing of interventions are labeled (in green) above the timeline, and all data collection and surveys are labeled (in blue) below the timeline. Providers were randomized into treatment arm in early 2014, and contracts signed during January–April 2013. Providers were visited again during May–August 2013 to discuss strategies and collect provider data. Household surveys (of mothers who delivered babies at study providers' facilities) were conducted between December 2013 and July 2014. The providers were visited again at the end of the study to make the incentive payments as specified in contracts and collect data.

Randomization.—The set of providers that we randomize come from the two different sources mentioned above. Of the 120 eligible providers in the data from the state government, using simple randomization, 38 providers were assigned to the input group, 40 to the output group, and 42 to the control group. Other eligible providers, who were inadvertently left out in the government-funded survey and identified by our field team during fieldwork, were randomized as follows: once the provider was confirmed to meet all eligibility criteria, the field team would call our project office to assign the provider to a study arm. This allocation was done according to a list of sequential unique identifiers, which were randomized prior to fieldwork (this list was unknown to field enumerators). Using this procedure, 2 providers were allocated to the input group, 13 to the output group, and 5 to the control.[18]

In all, 140 providers met all eligibility criteria and signed the incentive contracts in our study (note that the control group also signed a contract). Of these, 5 providers declined to participate over the course of the study, and were classified as attritors from the study (2 from the input incentive group and 3 from the control group). Our final analytical sample thus includes 135 providers: 38 providers in inputs arm, 53 providers in outputs arm, and 44 providers in control arm.[19] Table 1 shows the number of providers who were identified in sampling and the attrition.

Table 2 reports summary statistics for our final sample of providers used for analysis. Just over half of providers were female. Nearly 60 percent had advanced qualifications in obstetrics or a related field—we refer to this group as "MBBS plus." Of the remaining, over half had either basic training in allopathic medicine, equivalent

[18] Note that we could not ensure an equal number of providers across arms because we did not know how many providers the field team would find, and we did not want to have a predictable sequence so that our field enumerators could anticipate the treatment allocation of a potential provider.

[19] Further details on enrollment of providers and sample sizes at each stage are included in the preanalysis plan (https://www.socialscienceregistry.org/trials/179).

Table 1—Provider Sampling and Attrition

	Control	Input contract	Output contract
A. Providers identified from government survey data	42	38	40
B. Additional eligible providers identified during fieldwork for verification	5	2	13
C. Attrited from survey	3	2	0
Final Analytical Sample (A + B − C)	44	38	53

Notes: This table reports counts of the universe of providers identified as eligible for the study by randomly assigned treatment arm. Because providers identified during fieldwork were assigned to study arms based on a randomized list of sequence numbers (unknown to field enumerators, and the sequence was not predictable) it was not possible to ensure an equal number of providers across arms. Providers identified as attritors in row C declined to participate in the study during or after signing the contract. The last row includes the final sample of providers used in the analysis.

Table 2—Summary Statistics and Balance

Variables	All	Input group	Output group	Control group	Test of equality (*p*-value)
Female provider (share)	0.56	0.55	0.57	0.55	0.98
	(0.5)	(0.5)	(0.5)	(0.5)	
MBBS plus (share)	0.59	0.45	0.64	0.64	0.14
	(0.49)	(0.5)	(0.48)	(0.49)	
MBBS (share)	0.21	0.26	0.19	0.2	0.71
	(0.41)	(0.45)	(0.39)	(0.41)	
BAMS (share)	0.2	0.29	0.17	0.16	0.33
	(0.4)	(0.46)	(0.38)	(0.37)	
Other qualification (share)	0.04	0.03	0.02	0.07	0.52
	(0.19)	(0.16)	(0.14)	(0.25)	
Provider age (mean)	47.01	46.42	47.45	46.98	0.89
	(10.29)	(9.14)	(11.33)	(10.12)	
Years practicing (mean)	19.93	19.68	20.96	18.89	0.64
	(10.64)	(9.95)	(10.89)	(11.04)	
Years clinic operating (mean)	17.32	15.5	19.28	16.52	0.3
	(11.84)	(11.04)	(12.78)	(11.24)	
Observations	135	38	53	44	

Notes: This table reports mean provider characteristics by study group. Provider characteristics are self-reported and measured through interviews with the provider or with a staff member. Rows 2–4 refer to provider training: "MBBS plus" is a Bachelor of Medicine degree with a specialization such as obstetrics, "MBBS" is a Bachelor of Medicine degree with no additional specialization, "BAMS" is a degree in Ayurveda medicine. Standard deviations are reported in parentheses. *p*-values in the final column are associated with *F*-tests of joint equality across the three study groups.

to an MD in the United States or comparable training in Ayurvedic medicine—corresponding to MBBS and BAMS degrees, respectively (Mahal and Mohanan 2006). The average provider had been practicing for nearly two decades. Joint tests of orthogonality show there are no significant differences in provider demographics between the three study arms (Appendix Table A1). The attrition of five providers across the three study groups was not statistically different at the 5 percent level (Appendix Table A2).

B. *Study Arms/Contract Types*

The three contracts (control, input incentive contract, and output incentive contract) were designed to be as comparable as possible other than the basis of payment. Providers were first introduced to the contracts during visits between February and April 2013 (Figure 1 shows our study timeline). During these initial visits, all providers (including those in the control group) were given copies of letters of support from the state government and a full set of reference materials including guidelines for maternity care from the World Health Organization (WHO) and Government of India (GoI).[20] These letters also provided a broad overview of what participation in the study would entail, including future meetings and compensation to participating providers for their time to compile patient lists and complete surveys (participation payments).

Each provider was also given a copy of his/her randomly assigned contract. Each treatment group contract explained the specific basis by which the provider would be rewarded at the end of the study period, including details of reward calculations and payments (online Appendix 1 shows each type of contract and accompanying WHO guidelines).[21] The contracts specified that the final payment will be made only at the end, and there were no interim incentive payments.

Input and output incentive contracts were designed to have equal maximum level of payments. Payment levels were also set to ensure that the project could meet payment obligations in the event that all providers achieved the maximum performance level. The resulting contracts offered providers the potential to earn up to approximately 150,000 rupees (about US$2,700 at the time of the contract—slightly more than 15 percent of a specialist doctor's salary in Karnataka).

The control arm contract was designed to inform providers about our study of maternal and child health, to provide the same WHO and GoI guidelines, and to require control providers to sign an "agreement" confirming their willingness to participate in a study of maternal and neonatal health. The control contract did not mention reward payments made to other providers in the study.

Enumerators were trained to ensure that the providers fully understood their contracts, including basis and structure of incentive payments, the potential reward payments possible for strong performance, and the fact that providers would not lose money by participating in the study, regardless of their performance. Contracts also specified that providers' performance on rewarded outcomes would be evaluated using data collected from household surveys with their patient population.[22] Finally, providers in all three arms were offered 2,500 rupees (about US$45) at each visit as

[20] A complete set of guidelines was also provided to the providers on a CD. If a provider was unable to access the materials on the CD, she was offered the option of having the hard copy versions sent to her at no charge.

[21] The contracts also provide benchmark information in specifying floors for incentive payments. Note that the control group contract did not include this information, and therefore our treatment effects include both the effect of incentives plus any effect of providing benchmark information (a design choice made for policy relevance given sample size constraints).

[22] To avoid possible collusion or gaming, information about specific survey questions used to calculate rewards was not shared with anyone outside of the study team, including the enumerators when they first met providers to implement the contracts.

compensation for the time required to participate in the study. This small payment also aimed to develop credibility for future reward payments.

Output Contract Structure.—Output incentive payments were offered for achieving low rates of four adverse health outcomes (postpartum hemorrhage (PPH), preeclampsia, sepsis, and neonatal mortality) during the study period among a provider's patients. PPH, preeclampsia, and sepsis are the three leading causes of maternal mortality globally, accounting for 27 percent, 14 percent, and 11 percent of all maternal deaths (respectively) between 2003 and 2009 (Say et al. 2014). However, unlike PPH and sepsis, there is little that healthcare providers can do to prevent preeclampsia—a hypertensive disorder that occurs during pregnancy. Similarly, only 10.5 percent of neonatal mortality is attributed to complications during childbirth (Liu et al. 2015). As a result, we did not anticipate marked incentive effects on preeclampsia or neonatal mortality (but included them to minimize reductions in effort focused on them—i.e., to mitigate multitasking).

Ideally, we would have set the reward levels for each health outcome optimally: the rewards that maximize the principal's utility subject to the participation constraint of the provider. However, this requires detailed knowledge of the production, utility, and cost functions, which were unknown when designing the study. Our approach (described below) therefore resembles one of a cautious policymaker, ensuring that total incentive payments do not exceed a fixed budget constraint.

For neonatal mortality, a provider would receive 15,000 rupees unless one of their newborn patients died. For each of the other three maternal health outcomes (PPH, preeclampsia, and sepsis), the reward payment for output i, $P(x_i)$, was a decreasing linear function of incidence rate x_i, with payment increment α_i for incidence rates below a preestablished incidence rate ceiling $\bar{x}_i$:

$$P(x_i) = \begin{cases} \alpha_i(\bar{x}_i - x_i), & \text{for } x_i \leq \bar{x}_i; \\ 0, & \text{for } x_i > \bar{x}_i. \end{cases}$$

We set $\bar{x}_i$ equal to the pre-intervention average rates, which we estimated using existing data from government surveys. To set levels of α_i, we first allocated the remaining available budget for output contracts (after deducting payment for neonatal mortality) to each of the three outputs equally. The value of α_i for each output was then determined by dividing the available budget for that output by the potential improvement for that output (i.e., the difference between the pre-intervention average level of $\bar{x}_i$ and 0.05, which assumes providers would, on average, not be able to eliminate negative health outcomes completely):[23]

$$\alpha_{i=OUTPUT} = \frac{(Budget\,for\,output\,contracts - NMR\,payment)/3}{(\bar{x}_i - 0.05)}.$$

[23] For example, preintervention rates of postpartum hemorrhage (PPH) were estimated at 35 percent ($\bar{x}_{PPH} = 35$) in the study area. Providers could earn $\alpha_{PPH} = 850$ rupees (equivalent to about US$17 at the time of the contract) for every percentage point below 35 percent incidence of PPH in their patient population. If the rate of PPH measured in their patient population over the study period was 25 percent, they would earn US$170; if they were able to completely eliminate PPH in their patient population, they would earn US$595.

The final reward payment for providers in the output group was then the sum of rewards for each of the four outputs.

Input Contract Structure.—Providers assigned to the input treatment arm were offered incentive payments for health inputs provided to patients according to 2009 World Health Organization (WHO) guidelines.[24] These inputs are categorized into five domains: pregnancy care, childbirth care, counseling for postnatal maternal care, newborn care, and counseling for postnatal newborn care.[25] Analogous to the structure of output incentives, for each domain i, the input reward payment $P(x_i)$ was structured as an increasing linear function of the input level x_i—the share of measurable inputs for appropriate care for domain i, averaged over the provider's patients—with incremental payment α_i above a preestablished performance floor $\underline{x}_i$ percent:

$$P(x_i) \;=\; \begin{cases} \alpha_i(x_i - \underline{x}_i), & \text{for } x_i \;\geq\; \underline{x}_i; \\ 0, & \text{for } x_i \;<\; \underline{x}_i. \end{cases}$$

As in the output contract case, α_i for inputs was calculated by dividing the available budget by the projected range of improvements from the pre-intervention average rates to an average of 90 percent.[26] The final reward payment for each provider was the sum of rewards earned for performance in each of the five domains of care.

Control Arm Contracts.—Providers assigned to the control arm received contract agreements that provided the same information, guidelines, and participation payments as in the two incentive contract arms—but had no payments related to performance. Control providers were also told that the project team would collect survey data from their patients and received the same follow-up visits as intervention arm providers.

C. *Data Collection, Household Sampling, and Measurement*

We collected data from providers through multiple interviews over the study period and from households at end of the study period (Figure 1 shows details of timing of data collection and intervention visits to providers). Through our provider surveys, we collected information about providers' medical practices, staffing, and infrastructure, as well as intended strategies for improving quality of care and health outcomes.

Additionally, we collected patient lists from providers to create our primary patient sampling frame. A natural concern with this approach is that providers would have incentives to selectively report only patients with relatively good performance

[24] These were the most up-to-date guidelines at the time of the intervention.

[25] Details of the measurement of these health inputs are below and in online Appendix 2: Calculation of Inputs and Outputs.

[26] For example, preintervention coverage of the inputs in the Childbirth Care domain was estimated at about 65 percent ($\underline{x}_{Childbirth\,Care} = 65$) in the study area: patients receive 65 percent of appropriate childbirth care according to WHO guidelines. Providers earn $\alpha_{Childbirth\,Care} = 750$ rupees (equivalent to about US$15 at the time of the contract) for every percentage point in coverage of these inputs above 65 percent. If 75 percent of a provider's patients had received appropriate level of inputs for the Childbirth Care domain, she would earn US$150, and if she were able to provide this level of care for 100 percent of her patients, she would earn US$525.

indicators. To minimize this concern, we also collected data from approximately 75 households (not used in this analysis) in areas surrounding each clinic to ensure there were no cases with negative outcomes at the providers' facilities but were not reported by providers, or that were inappropriately referred away. The incentive contracts also clearly explained that any instances of patient list manipulation, either through selective referrals or reporting, would nullify the contracts.[27]

Using patient lists, we then aimed to sample 25 women who had recently given birth at the provider's facility.[28] Enumerators collected the list of patients and a study team member managing the field project conducted random sampling of patients. In instances where there were fewer than 25 deliveries over the timespan of data collection, all listed patients were surveyed. These surveys measured the four major health outcomes,[29] input use in the five domains of maternity care, and basic sociodemographic information. We aimed to interview every mother within approximately two weeks after she gave birth to minimize recall inaccuracy (Das, Hammer, and Sánchez-Paramo 2012). In practice, we conducted surveys with new mothers between 7–20 days after delivery, and also did a very brief follow up with these mothers after 28 days after birth to assess the infant's status. In total, we interviewed 2,895 new mothers.[30]

Measurement of health input use and outputs poses important challenges, especially in developing country contexts where reliable administrative data on input use are not available. Using providers' reports of outcomes leads to concerns of gaming when incentives are tied to performance. Furthermore, providers may not always be able to accurately identify some health outcomes. For example, in the case of maternal health, evidence from studies comparing actual blood loss to providers' visual estimates show that providers tend to underestimate the amount of blood loss by one-third (Patel et al. 2006).

Given that we chose to measure health outcomes and health input use through household surveys, we relied on two general criteria for selecting our specific measures (which we use both for calculating incentive payments as well as for our empirical analysis). First, we chose questions previously validated through past research published in the clinical literature (Stewart and Festin 1995, Filippi et al. 2000, Stanton et al. 2013). Second, prior to our study, we conducted our own

[27] See page 5 of sample contracts in online Appendix 1 for exact language on selective referrals that would nullify contracts. Using data collected from communities around the provider, we verified that there were no unusual patterns of referral suggesting providers did not respond by selecting patients with better outcomes or selectively reporting by providers.

[28] Power calculations were conducted prior to the data collection. Estimated preintervention performance rates and feasible improvement levels (i.e., target levels) were determined using existing data from government surveys and calibrated through piloting with doctors in Karnataka and Delhi to ensure that they were locally appropriate. We assumed 25 mothers per provider and an intraclass correlation coefficient of 0.05. At the individual level, all five categories for quality of care have at least 85 percent power to detect improvements that reach the target levels, with the "Childbirth Care," "Postnatal Maternal Care," and "Postnatal Newborn Care" categories having at least 95 percent power. Two of the four outputs, postpartum hemorrhage and preeclampsia, have at least 85 percent power to detect improvements to the target levels. Note that these calculations do not take into account additional precision gained by including covariates.

[29] We collected data from household surveys about signs and symptoms for the health outcomes and used algorithms described in the online Appendix to establish whether a woman had each adverse health outcome or not.

[30] Some providers conducted fewer than 25 deliveries over the data collection period, resulting in fewer than the targeted 3,375 mothers (135 providers $\times$ 25 mothers). On average, we have data from 21.4 mothers per provider, with an interquartile range of 17 to 26 mothers per provider.

validation exercise. Specifically, we trained nurse enumerators to observe and code health input use in real-time during labor and delivery for 150 deliveries in rural Karnataka. Within two weeks after delivery, we then visited these new mothers and administered a set of survey questions intended to measure the same health input use, as reported by the mother. We then chose measures that performed well in our validation exercise as additional survey questions for the project.[31]

Mothers in our sample were classified as having an adverse health outcome based on a combination of her responses to relevant questions, following previous studies of the sensitivity and specificity of responses to these questions for clinical evaluation of the incidence of these outcomes (Stewart and Festin 1995, Filippi et al. 2000, Stanton et al. 2013). We evaluate inputs provided by each provider by measuring each provider's adherence to WHO guidelines. Given the criteria described above, we generated household survey questions that women could plausibly answer and that related to the guidelines. The responses to these questions were assigned a score of 1 if they adhered to the guidelines, and 0 otherwise.[32] A provider's performance in a particular domain was then the mean of these scores for all mothers who received care from the provider, where higher scores reflect greater adherence to the guidelines and better performance. For analysis of inputs within each domain, we aggregate the multiple measures into a summary index following Anderson (2008).[33,34]

D. *Analysis*

We use the estimation strategy that we specified in our pre-analysis plan published in the AEA RCT registry in December 2013, prior to collecting any household-level data (Mohanan and Miller 2016). To estimate the effect of each type of incentive contract on health outputs and health input use, we regress outcomes on dummy variables indicating treatment status with the following estimating equation:

$$(1) \qquad y_{ipde} = \alpha + \beta T_p + \theta X_p + \gamma Z_i + s_d + \lambda_e + u_{ipde},$$

where y_{ipde} is an outcome of interest (i.e., level of care—inputs—received or health outcomes) for woman i who received care from provider p, located in district d, and was interviewed by enumerator e, T_p is a vector of provider-level treatment indicators, X_p is a vector of baseline (pre-contract) provider characteristics, Z_i is a vector of time-invariant household characteristics (such as mother's age, education status, religion, and birth history), and s_d and λ_e represent district and enumerator fixed effects (respectively). We also show estimates that do not condition on household

[31] Results from this validation study will be published in a separate manuscript and are available upon request.

[32] For example, if a woman answered affirmatively to the question, "Was your blood pressure checked during labor?" the question was assigned a "1." Details about the specific questions used for each domain and how responses were coded are included in the online Appendix on Calculation of Inputs and Outputs, also available at https://www.socialscienceregistry.org/trials/179.

[33] The Anderson index is calculated as a weighted mean of the standardized values of all inputs within each domain (with variables redefined so that higher values imply a better, more desirable outcome). The weights are calculated to maximize the amount of information captured in the index, with highly correlated variables receiving less weight (Anderson 2008).

[34] Data for replication is available online (Mohanan et al. 2020).

or provider characteristics, but only include enumerator and district fixed effects, as specified in our pre-analysis plan. In all cases, we cluster standard errors at the provider level.

Given that we test multiple hypotheses across two treatment arms, we report p-values adjusted for multiple comparisons within each pre-specified family of hypotheses to control for the Familywise Error Rate (using the free step-down resampling method described in Westfall and Young 1993) and across the two types of contracts. Following our pre-analysis plan, we consider PPH, sepsis, and neonatal death as one family of health outcomes influenced by medical care provided around the time of delivery (as opposed to care throughout pregnancy for preeclampsia, which we test across two types of contracts). Similarly, for input use, we consider three domains (childbirth care, postnatal maternal care, and newborn care) to be a family of outcomes because these are all inputs provided at the time of delivery.[35]

As Section I indicated, we expect health outcomes to vary according to a provider's skills under output incentive contracts, but to be independent of them under an input incentive contract. To test this hypothesis, we augment regression (1) with an indicator for higher provider qualification multiplied by each provider contract arm.[36]

III. Results

In this section, we first report how our incentive contracts influenced the production of health outputs and the provision of health inputs, investigate the mechanisms underlying these results, and examine the relative costs of the two types of contracts. We then study how providers with varying levels of qualifications and skills responded differently to each type of contract.

A. *Health Outputs*

Table 3 reports estimates of how each incentive contract influences maternal and child health outcomes. Our preferred (prespecified) estimates from equation (1), shown in even-numbered columns, condition on provider and patient characteristics as well as district and enumerator fixed effects (odd-numbered columns report estimates that condition only on district and enumerator fixed effects). The levels of statistical significance indicated reflect p-values adjusted for multiple comparisons within each family of hypotheses to control for the Familywise Error Rate. The rows in italics in Table 3 report the adjusted p-values for the main results.

In both incentive contract groups, postpartum hemorrhage (PPH) rates declined by nearly identical (and statistically indistinguishable) amounts relative to the control group.[37] Column 2 shows that input contract providers reduced PPH incidence

[35] Our preanalysis plan also included the evaluation of other potential mechanisms through which incentive contracts might influence outcomes. In a separate paper (Donato et al. 2017), we examine heterogeneity of program impact as a function of personality traits. We focus attention in this paper on the primary prespecified analyses (on outputs and inputs).

[36] Note that the vector X_p includes provider qualification.

[37] Testing $\beta_{output} = \beta_{input}$, we fail to reject the null hypothesis ($p = 0.72$).

Table 3—Impact of Provider Incentives on Outputs

	Postpartum hemorrhage		Preeclampsia		Sepsis		Neonatal death	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Input incentives	−0.0842	−0.0845	0.0312	0.0573	0.0333	0.0371	−0.0073	0.0032
(SE)	(0.0297)	(0.0284)	(0.0450)	(0.0434)	(0.0228)	(0.0253)	(0.0087)	(0.0051)
Adjusted *p*-value	*0.0244*	*0.012*	*0.4756*	*0.185*	*0.4284*	*0.437*	*0.7649*	*0.605*
Output incentives	−0.0622	−0.0741	0.0466	0.0612	0.0065	0.0209	−0.0091	0.0079
(SE)	(0.0286)	(0.0294)	(0.0325)	(0.0329)	(0.0198)	0.0224	(0.0111)	(0.0067)
Adjusted *p*-value	*0.1253*	*0.032*	*0.2585*	*0.112*	*0.7649*	*0.605*	*0.7649*	*0.557*
p-value for difference of output versus input	0.3765	0.7209	0.7022	0.9235	0.0769	0.3127	0.7882	0.2881
District and enumerator fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household- and provider-level controls	No	Yes	No	Yes	No	Yes	No	Yes
Control mean	0.365	0.365	0.179	0.179	0.0651	0.0651	0.0121	0.0121
Observations	2,890	2,748	2,894	2,748	2,891	2,748	2,894	2,748
R^2	0.266	0.279	0.255	0.270	0.106	0.119	0.0582	0.0532

Notes: Estimates obtained through OLS. Robust standard errors, clustered at the provider level, are reported in parentheses. *p*-values in *italics* are adjusted for multiple hypotheses tested and calculated using the free step-down resampling method. Each specification includes district and enumerator fixed effects; even columns additionally include household-level controls (mother's age and education; household's caste and house type, i.e., houseless, kutcha, semi-pucca, or pucca; head of household's religion; mother's history of hypertension, diabetes, asthma, hyper- or hypothyroidism, and convulsions; whether the mother has had a previous stomach surgery; whether it is the mother's first pregnancy, number of previous pregnancies, whether the mother has had a stillbirth or abortion, and number of previous children birthed; whether the household owns land, has no literate adults, and owns a Below Poverty Line card) as well as provider-level controls (primary provider's gender, professional qualifications, number of years in practice, and number of years that the facility has been in operation). All dependent variables measured through household surveys fielded between November 2013 and July 2014; see online Appendix for details of measurement.

among their patients by 8.4 percentage points, while output contract providers reduced PPH incidence by 7.4 percentage points. Compared to the control group mean (0.365), these reductions correspond to a 23 percent and 20 percent decline, respectively. Both are also statistically significant after correcting for multiple comparisons: adjusted *p*-values using the Westfall and Young (1993) step-down resampling method are 0.01 for the input group and 0.03 for the output group.

We do not find statistically significant changes for other health outcomes after adjusting for multiple comparisons.[38] This pattern of results is reasonable—in rural India, PPH is most amenable to improvement through changes in provider behavior at the time of delivery (with the use of drugs to control postpartum bleeding, for example, for which we find evidence in Section IIIB). Alternatively, among the four domains of health outcome, providers have the least control over preeclampsia because it is a hypertensive disorder developed earlier during pregnancy—and

[38] As mentioned earlier, although providers can do little to prevent preeclampsia (a hypertensive disorder that occurs during pregnancy) or neonatal mortality (almost 90 percent of which is attributed to causes other than childbirth complications), we included them in output contracts to mitigate concerns of potential multitasking. Among the results for preeclampsia and sepsis for input and output contracts, only the preeclampsia result is marginally significant ($p = 0.07$) when not adjusting for multiple comparisons.

women generally seek antenatal care from other providers. Furthermore, the biological causes of preeclampsia remain scientifically unclear, essentially making it impossible for providers to predict and prevent this condition, but it can be better managed if detected earlier in the pregnancy (Steegers et al. 2010, Mol et al. 2016, Phipps et al. 2016). For sepsis, a key preventive strategy (wearing gloves during delivery) was already practiced among 99 percent of control group providers, and prophylactic antibiotics are commonly used at high (and inappropriate) rates in rural India, including Karnataka.[39]

B. *Health Input Use and Underlying Mechanisms*

Table 4 then reports estimates from equation (1) for provision of health inputs. Because we only find significant health improvements for PPH, we do not expect substantial improvements in input use across all five domains of maternal and neonatal care. Column 6 shows that in the output contract group, the postnatal maternity care index (which primarily reflects postnatal health counseling provided to mothers shortly after delivery) rose by 0.077 index points relative to the control group; this estimate is statistically significant (unadjusted p-value $= 0.032$), but not at conventional levels after correcting for multiple hypotheses testing (adjusted p-value $= 0.154$).[40] There were no improvements in the five composite domains of maternal and neonatal care in the input incentive contract group. In Section IIIE, we discuss the 0.14 point decline in the postnatal newborn care counseling index shown for the output contract group in column 10, which we believe reflects a reduction in effort devoted to newborn care (i.e., "multitasking"; see Holmstrom and Milgrom 1991, Prendergast 2011).

However, other than in postnatal maternity care, we do not observe significant improvements for indices in other domains of care. This is probably because the indices aggregate many inputs, only a subset of which directly influence PPH (those included in Active Management of Third Stage of Labor (AMTSL), for example).[41] Although not prespecified, we therefore directly examine changes in two inputs most closely related to PPH: parenteral oxytocic drugs (whose administration is recommended universally for all mothers) and manual removal of placenta (which reflects complications that could potentially be avoided with better care).[42]

[39] The other clinical action listed in the guidelines given to providers is handwashing, but provider handwashing behavior is not reliably observed by mothers or accompanying caregivers. Antibiotics are routinely overused in clinical settings in India (Ganguly et al. 2011).

[40] The magnitude of the increase (0.0773) is not directly interpretable because the weights used to compute the index change the scale (Anderson 2008).

[41] Active Management of Third Stage of Labor (AMTSL) recommended by WHO guidelines also includes early cord clamping, controlled traction of the umbilical cord, and transabdominal manual massage of the uterus (Urner, Zimmermann, and Krafft 2014). Abdominal massage was included in the 2009 guidelines from Government of India (MOHFW 2009) and was also recommended by American College of Obstetricians and Gynecologists at the time (ACOG 2011). The 2012 revised guidelines from WHO no longer recommends cord traction or abdominal massage as standard practice (Tunçalp, Souza, and Gülmezoglu 2013).

[42] Within the WHO guidelines that our input contracts reward, a clinical action closely related to the prevention of PPH—and recommended universally for all mothers—is the administration of medicines (parenteral oxytocic drugs), which are effective in stopping post-delivery bleeding. Clinical actions not universally recommended—ones that are clinically appropriate conditional on presence of a risk factor or manifestation of an adverse outcome, for example—are more difficult to interpret if the conditions requiring them are preventable.

Table 4—Impact of Provider Incentives on Inputs

	Pregnancy care		Childbirth care		Postnatal maternal care counseling		Newborn care		Postnatal newborn care counseling	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Input incentives	−0.0106	0.0028	−0.0203	0.0142	0.0380	0.0426	−0.055	−0.029	−0.0650	−0.0059
(SE)	(0.0455)	(0.0460)	(0.0338)	(0.0285)	(0.0390)	(0.0392)	(0.040)	(0.0372)	(0.0576)	(0.0578)
Adjusted *p*-value	*0.819*	*0.880*	*0.692*	*0.875*	*0.698*	*0.766*	*0.289*	*0.875*	*0.262*	*0.910*
Output incentives	−0.0529	−0.0548	−0.0311	−0.0191	0.0674	0.077	−0.029	−0.015	−0.1610	−0.138
(SE)	(0.0373)	(0.0402)	(0.0268)	(0.0250)	(0.0354)	(0.0358)	(0.032)	(0.0360)	(0.0435)	(0.0437)
Adjusted *p*-value	*0.263*	*0.276*	*0.665*	*0.875*	*0.273*	*0.154*	*0.692*	*0.875*	*0.001*	*0.003*
p-value for difference of output versus input	0.4112	0.2774	0.5842	0.1466	0.3853	0.3046	0.4038	0.6534	0.0530	0.0090
District and enumerator fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household- and provider-level controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Control mean	−0.062	−0.062	−0.005	−0.005	−0.088	−0.088	−0.002	−0.002	−0.068	−0.068
Observations	2,893	2,748	2,892	2,747	2,890	2,747	2,890	2,748	2,890	2,747
R^2	0.355	0.3621	0.356	0.382	0.406	0.421	0.427	0.447	0.471	0.489

Notes: Estimates obtained through OLS. Robust standard errors, clustered at the provider level, are reported in parentheses. *p*-values in *italics* are adjusted for multiple hypotheses tested and calculated using the free step-down resampling method. Each specification includes district and enumerator fixed effects; even columns additionally include household-level controls (mother's age and education; household's caste and house type, i.e., houseless, kutcha, semi-pucca, or pucca; head of household's religion; mother's history of hypertension, diabetes, asthma, hyper- or hypothyroidism, and convulsions; whether the mother has had a previous stomach surgery; whether it is the mother's first pregnancy, number of previous pregnancies, whether the mother has had a stillbirth or abortion, and number of previous children birthed; whether the household owns land, has no literate adults, and owns a Below Poverty Line card) as well as provider-level controls (primary provider's gender, professional qualifications, number of years in practice, and number of years that the facility has been in operation). All dependent variables measured through household surveys fielded between November 2013 and July 2014 and are based on WHO Guidelines (available at http://whqlibdoc.who.int/hq/2007/who_mps_07.05_eng.pdf); see online Appendix for details of measurement.

The first two columns of Table 5 show estimates for providers' stocking of parenteral oxytocic drugs at their clinics. Consistent with our PPH results in Section IIIA, we find that providers in both output and input contract groups were approximately 7 percentage points more likely to maintain stocks of parenteral oxytocic drugs in their clinics (relative to the control group mean of 0.93). Consistent with this finding, columns 3 and 4 also show estimates of patients' reported use of medicines to prevent bleeding, which are 6 percentage points higher in both incentive contract groups relative to the control group (estimates are statistically indistinguishable from each other with and without conditioning on various control variables, but only statistically different from zero in column 3).[43]

Additionally, a key corrective clinical action to prevent PPH when the placenta is not delivered normally is manual placenta removal (Urner, Zimmermann, and Krafft 2014). AMTSL, which is recommended by WHO guidelines, minimizes the time required for normal delivery of an intact placenta, so reductions in manual placenta removal can be interpreted as improvements in maternity care related to PPH

[43] This particular input is possibly measured with greater error than others because mothers and those accompanying them during childbirth are unable to observe the specific types of drugs administered.

Table 5—Impact of Provider Incentives on PPH Prevention and Management

	Parenteral oxytocic drugs available		Medicine use to reduce bleeding after delivery		Massage abdomen after delivery		Placenta manually removed	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Input incentives	0.0722	0.076	0.0636	0.0305	0.0518	0.0718	−0.0786	−0.0504
SE	(0.0415)	(0.0443)	(0.0322)	(0.029)	(0.0322)	(0.0427)	(0.0483)	(0.0436)
p-value	*0.085*	*0.089*	*0.05*	*0.029*	*0.11*	*0.095*	*0.106*	*0.251*
Output incentives	0.0730	0.0694	0.0623	0.0382	0.00517	−0.0105	−0.0666	−0.072
SE	(0.0422)	(0.0417)	(0.0286)	(0.0265)	(0.0289)	(0.0354)	(0.0386)	(0.0380)
p-value	*0.087*	*0.099*	*0.031*	*0.154*	*0.858*	*0.764*	*0.087*	*0.06*
District and enumerator fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household- and provider-level controls	No	Yes	No	Yes	No	Yes	No	Yes
Control mean	0.932	0.932	0.460	0.460	0.517	0.517	0.289	0.289
Observations	135	135	2,791	2,656	1,707	1,610	1,665	1,571
R^2	0.260	0.270	0.322	0.340	0.372	0.396	0.266	0.276

Notes: Estimates obtained through OLS. Robust standard errors, clustered at the provider level, are reported in parentheses. All specifications include district and enumerator fixed effects; even columns additionally include household-level controls (mother's age and education; household's caste and house type, i.e., houseless, kutcha, semi-pucca, or pucca; head of household's religion; mother's history of hypertension, diabetes, asthma, hyper- or hypothyroidism, and convulsions; whether the mother has had a previous stomach surgery; whether it is the mother's first pregnancy, number of previous pregnancies, whether the mother has had a stillbirth or abortion, and number of previous children birthed; whether the household owns land, has no literate adults, and owns a Below Poverty Line card) as well as provider-level controls (primary provider's gender, professional qualifications, number of years in practice, and number of years that the facility has been in operation). Dependent variables for columns 1–6 are measured through household surveys fielded between November 2013 and July 2014; see online Appendix for details of measurement. Dependent variable for columns 7 and 8 measured through interviews with a member of the hospital personnel and is a binary indicator for whether the provider's facility had any parenteral oxytocic drugs available at the time of the survey at the end of the study period.

(Begley et al. 2011). Column 8 of Table 5 shows a statistically significant 7 percentage point decline in manual placenta removal in the output contract arm (25 percent reduction relative to control), suggesting fewer instances in which corrective action was needed.[44] The corresponding estimate in the input arm is less precise, but comparable in magnitude.

C. *Relative Costs of Input and Output Contracts*

Given that our input and output incentive contracts produced statistically indistinguishable improvements in maternal health, we next briefly compare the costs required to produce these health benefits. Panel A of Figure 2 shows the distributions of incentive payments made to providers in treatment arm. Ex post, the average payment was much higher in the output contract group (56,812 rupees or US$1,033)

[44] Although abdominal massaging is no longer a recommended best practice as per revised WHO guidelines, we also see in Table 5 that providers in input contract arm were 7 percentage points (18 percent) more likely to massage the mother's abdomen relative to control arm, while providers in output contracts arm had no significant change.

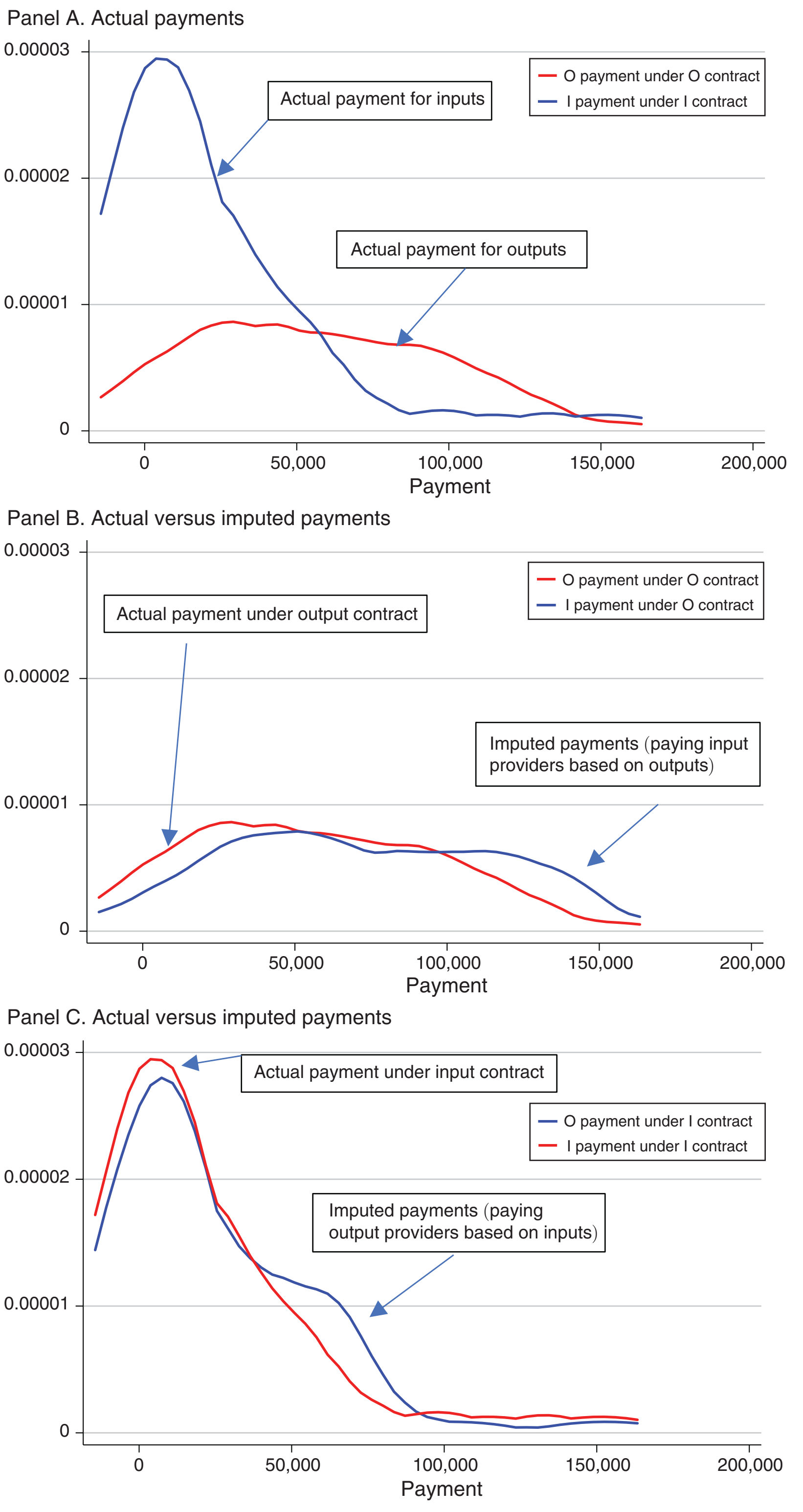

Figure 2. Comparing Actual and Imputed Payments under Input and Output Contracts

than in the input contract group (13,850 rupees or US$252).[45] In panels B and C, we also construct imputed distributions that reflect hypothetical payments that output contract group providers might have received if paid mechanically based on inputs (and vice versa). Note that this is only a mechanical effect, holding behavior constant. In general, for the specific contracts that we study, payments realized ex post for outputs are nearly four times as large as payments for inputs. A potential explanation is that some rewarded outputs (such as low neonatal mortality) might have occurred anyway, but we nonetheless rewarded them to ensure that providers did not divert effort away from these health outcomes.

In the setting of our study, the input contract was more efficient than the output incentive contract because it delivered the same health outcomes at much lower cost to the principal. However, it is important to note that this only applies to the input and output incentive contracts that we study and is not generalizable to other input and output contracts. For instance, if the elasticity of an output with respect to the reward in the output contract is very low, a much less generous output contract could have delivered similar health improvements at a smaller cost to the principal.[46] Although we were unable to conduct the experiment over a range of payment rates for inputs and outputs, we note that other studies that experimented with different payment rates for output contracts have found this elasticity to be significant and sizeable (Luo et al. 2015).

D. *The Role of Skills in Provider Responses to Output and Input Incentive Contracts*

As our conceptual framework in Section I suggests, we expect provider skills to play an important role in determining the effectiveness (and relative effectiveness) of output and input incentive contracts. With input incentive contracts, providers are paid to use explicitly specified inputs ("follow orders"); hence, provider skill may be less relevant. Alternatively, with output incentive contracts, provider skill may play a much more important role because more-skilled providers are better able to choose the optimal combination of inputs in innovative ways using local/contextual information (albeit with less control over contracted outcomes and therefore more uncertainty about incentive payments).

In Table 6, we examine differences in providers' behavioral responses to incentive contracts by level of skill, measuring skills based on whether or not providers have medical degrees with specific obstetric training ("MBBS plus" providers) qualifying them to provide maternity care.[47] Our results suggest that in the output contract group, "MBBS plus" providers produced PPH rates that were 11 percentage points lower on average than providers without obstetric qualifications, although this result is just shy of conventional statistical significance. The table also shows that

[45] Exchange rate 1 USD = 55 INR in 2013.

[46] We are grateful to Oriana Bandiera and Paul Gertler for helpful discussions on this point.

[47] The basic medical education at the level of MBBS and BAMS includes a few months of training in obstetrics that gives only introductory level of skills. Such providers are able to conduct normal deliveries but do not have training in management of complications or the surgical skills that are acquired as part of advanced obstetric training programs (typically two to three years of training after completing medical school) (Mahal and Mohanan 2006).

Table 6—Impact of Incentives on Post Partum Hemorrhage by Provider Qualifications

	(1)
MBBS plus	0.01
(SE)	(0.054)
p-values	*0.866*
Input incentives	−0.037
(SE)	(0.045)
p-values	*0.397*
Output incentives	−0.001
(SE)	(0.047)
p-values	*0.976*
Input × MBBS plus	−0.059
(SE)	(0.056)
p-values	*0.318*
Output × MBBS plus	−0.111
(SE)	(0.057)
p-values	*0.057*
p-value for difference of input/output × MBBS plus	*0.494*
District and enumerator fixed effects	Yes
Household- and provider-level controls	Yes
Observations	2,748
R^2	0.280

Notes: Estimates from OLS regression on PPH as a function of provider qualification category. The MBBS plus variable takes value 1 if the provider holds an MBBS degree (Bachelor of Medicine, Bachelor of Surgery) with advanced medical training in obstetrics and gynecology, 0 otherwise. Robust standard errors, clustered at the provider level, are reported in parentheses. Each specification includes district and enumerator fixed effects, household-level controls (mother's age and education; household's caste and house type, i.e., houseless, kutcha, semi-pucca, or pucca; head of household's religion; mother's history of hypertension, diabetes, asthma, hyper- or hypothyroidism, and convulsions; whether the mother has had a previous stomach surgery; whether it is the mother's first pregnancy, number of previous pregnancies, whether the mother has had a stillbirth or abortion, and number of previous children birthed; whether the household owns land, has no literate adults, and owns a Below Poverty Line card) as well as provider-level controls (primary provider's gender, number of years in practice, and number of years that the facility has been in operation). The dependent variable (PPH) is measured through household surveys fielded between November 2013 and July 2014; see online Appendix for details of measurement.

differences in the interaction terms are not statistically significant (*p*-value of 0.49). For the output contract group, we can reject (with a *p*-value of 0.057) that "MBBS plus" providers performed the same as less-qualified providers. We are unable to reject the corresponding relationship under input contracts. Considering both interactions, we are also unable to reject that "MBBS plus" providers outperformed their less-qualified peers as much under input contracts as the output contracts. A potential mechanism through which "MBBS plus" providers might do better with output incentive contracts would be if they use local/contextual information to improve

Table 7—Provider Qualifications and Relationship with Implementing New Strategies

	Implement new strategies	(SE)	*p*-value
Panel A. Regressions			
Input incentives	−0.263	(0.168)	*0.120*
Output incentives	−0.165	(0.158)	*0.299*
Input incentives × MBBS plus	0.406	(0.244)	*0.099*
Output incentives × MBBS plus	0.529	(0.218)	*0.017*
MBBS plus	−0.446	(0.299)	*0.004*
Panel B. Results from linear combinations			
Effect of input contracts on MBBS plus	0.143	(0.167)	*0.395*
Effect of output contracts on MBBS plus	0.364	(0.142)	*0.012*
District fixed effects	Yes		
Provider-level controls	Yes		
Observations	135		
R^2	0.378		

Notes: Estimates obtained through OLS. The dependent variable is an indicator for if the provider reported implementing any new strategies since signing the contract, measured through a survey at the first post-contract provider visit. The MBBS plus variable takes value 1 if the provider holds an MBBS degree (Bachelor of Medicine, Bachelor of Surgery) with advanced medical training in obstetrics and gynecology, 0 otherwise. The specification also includes district fixed effects as well as provider-level controls (primary provider's gender, number of years in practice, and number of years that the facility has been in operation). Robust standard errors, clustered at the provider level, are reported in parentheses.

care beyond simple guideline adherence—but only when they also have sufficient complementary skills to do so.

To further explore the possibility that high-skilled providers use local/contextual information in innovative ways under output incentive contracts, we directly examine providers' reports of implementing new delivery strategies since our baseline survey. Table 7 shows that among "MBBS plus" providers, the input and output contracts increased the probability of implementing new strategies by 0.41 and 0.53. Although "MBBS plus" providers in the output contract group were nearly 13 percentage points more likely to implement new strategies, the standard errors of the estimated effects are large (and the difference is not statistically significant at conventional levels). As a thought exercise, we also compare the performance of "MBBS plus" in the two incentive contract arms to "MBBS only" providers in the control arm. Output contracts increased the probability that "MBBS plus" providers implemented new strategies by 0.364 ($0.364 = -0.165 + 0.529$; standard error $= 0.142$) relative to less-qualified providers in the control arm, which is statistically different from zero. In contrast, the input contract did not increase the use of new strategies among "MBBS plus" providers relative to less-qualified providers in control arm, ($0.143 = -0.263 + 0.406$; standard error $= 0.167$). The first two rows also show that neither type of contract increased the probability that less qualified ("MBBS only") providers implemented new strategies.

We also investigate if "MBBS plus" providers differed in their use of clinical inputs relevant to the prevention and control of PPH (use of parenteral oxytocic drugs and manual removal of placenta); see Appendix Tables A3 and A4. Output incentive providers with "MBBS plus" degrees performed better than their counterparts with input incentives both in the availability of medicines to reduce bleeding and in massaging the abdomen, but the differences are imprecise.

Finally, a potential explanation for differences in provider performance by qualifications is that "MBBS plus" providers are not only better trained but that they also work in better-equipped health facilities. Indeed, facilities in which "MBBS plus" providers work are more likely to have screening tests for pregnancy disorders, infrastructure for intrapartum care, and adequate supply of drugs. However, Table 6 (column 2) shows that controlling for these facility characteristics does not change the estimates for output contracts among MBBS plus providers.[48,49]

E. *Expectations and Multitasking*

Although our incentive contracts generally cover all domains of maternity care provision, a natural concern with performance incentives is "multitasking" (or the reduction of effort on unrewarded margins, or those for which expected net benefits are lower) (Holmstrom and Milgrom 1991, Prendergast 1999). Without knowing the underlying production function and cost functions, it was not possible to know ex ante if the contracts rewarded some outcomes more generously (net of the full cost of providing them) than others. Importantly, this depends on providers' expectations about their ability to improve outcomes (in both absolute and relative terms).

In Table 4, column 10, we find a 0.14 point decline in the postnatal newborn care counseling index among output contract group providers ($p < 0.01$), which may reflect a reduction in effort devoted towards newborn care. To explore this possibility further, we use measures of provider beliefs about their ability to improve each of the four major health outcomes (i.e., outputs) that we collected prior to introducing incentive contracts. About 35 percent of providers rated neonatal mortality as the most difficult one to improve among the four outcomes. Instead, providers generally attributed neonatal mortality to the actions of caregivers at home (driven by traditional beliefs that colostrum is "witch's milk," for example) and beyond providers' control. Moreover, when asked which of the four major health outcomes was most important to improve based on patients' clinical needs, only 9 percent said neonatal mortality, while 75 percent said PPH (Figure 3). This pattern of beliefs is consistent with output contract providers diverting effort away from postnatal newborn care (and preventing neonatal mortality) and toward preventing and treating PPH. In contrast, column 10 of Table 4 shows no commensurate reduction in postnatal newborn care counseling delivered by providers in the input contract group. Because postnatal newborn care counseling largely comprised of giving information to mothers about how to care for the newborns and detect birth-related complications at home, it is reasonable that input contract providers responded to

[48] Consistent with our preanalysis plan, we investigate heterogenous incentive effects on PPH given that this is the primary health outcome that improved. In addition to the role of qualifications, we also examine other provider characteristics, including gender of provider, time preferences, and risk aversion. We do not generally find significant heterogeneous effects along these dimensions.

[49] Results from a fully interacted model, that interacts treatment variables with all provider facility controls are statistically similar to results seen in Table 6, column 2: coefficient (standard error) for output × MBBS-Plus −0.074 (0.099).

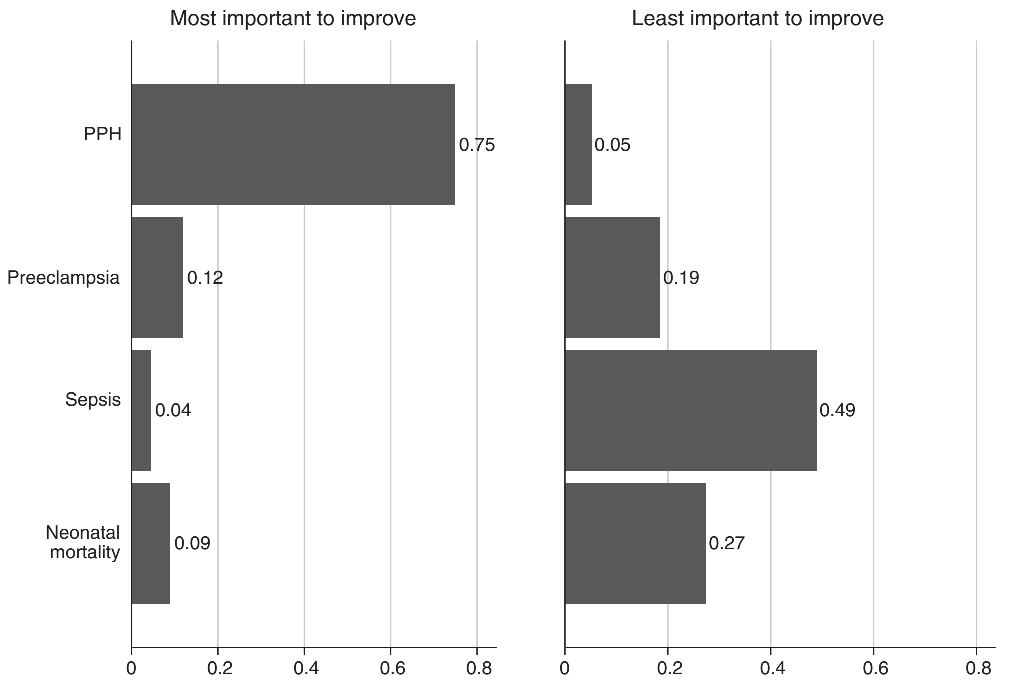

Figure 3. Provider Expectations about Improvements in Outcomes

Notes: Panel on the left shows providers' response to a question asking them to rank the four outcomes based on which one was most important to improve among their own patients. Bars indicate percentage of providers who responded that a given outcome was most important. The bars in the panel on the right shows providers' responses indicating outcomes that they thought were least important to improve among their patients.

performance incentives to deliver this counseling despite believing that it would have little effect on mothers' care for their babies at home.

Taken together, our results suggest that improvements in PPH under incentive contracts may have come at the expense of some reduction in newborn care—and did so only under circumstances in which providers believed that effort on newborn care was particularly unlikely to be rewarded (i.e., output incentive contracts).

We also investigate if providers' beliefs about their ability to reduce the incidence of PPH are a source of heterogeneity in providers' responses to incentives.[50] We estimate the effect of incentive contracts at the twenty-fifth, median, and seventy-fifth percentile of beliefs. Across both types of contracts, providers with lower levels of prior beliefs demonstrated larger reductions in PPH with incentive contracts—approximately 13 percentage point reduction at the twenty-fifth percentile (with t-statistic of 4.4) and 5 percentage points (although not statistically significant) at the seventy-fifth percentile in the inputs contract group (see Appendix Table A5). This pattern of results is consistent with an interpretation that providers

[50] Expectations are reported on provider surveys conducted prior to introduction of contracts. Providers were asked, "In general, do you think doctors can reduce the incidence rates of PPH?" with responses ranging from 1 (least likely) to 10 (most likely).

with high levels of beliefs exert high effort even without incentive contracts with relatively little room for improvement, while providers with low levels of beliefs are able to exert higher effort with incentives.

F. *Demand Response and Patient Selection*

An important issue in interpreting our results is the extent to which they reflect changes in patient composition rather than clinical actions taken by providers. There are two primary ways that patient composition might change: patient demand could change in response to improvements in quality of care, or providers could manipulate the composition of patients that they treat (by selectively referring some patients to other providers, for example). Although we are unable to distinguish between these two channels directly, we analyze their net effect. We also note that we deliberately constructed our incentive contracts to minimize provider manipulation of the types of patients that they treat, explicitly indicating that any evidence of patient selection would nullify their incentive contract.[51] We also collected data from 75 women who had babies in the past year in communities around each provider to assess whether the provider had engaged in gaming patient composition.

To investigate changes in patient composition, we first use our control group subsample to regress an indicator of whether or not any of the four major adverse health outcomes (PPH, preeclampsia, sepsis, and neonatal mortality) occurred on the individual characteristics that we use as controls in equation (1). We then use the resulting parameter estimates to predict the probability of an adverse health event for each mother in the full sample. Appendix Table A6 reports the means of these predicted probabilities for each study arm.

Both input and output contract providers had patients who were 6 to 9 percentage points more likely to experience any adverse health event than patients in the control group (a statistically significant difference).[52] Because it seems unlikely that providers in either treatment group would purposefully select patients with greater risk of health complications (especially those with output contracts), this finding may possibly reflect a demand response: if providers in both treatment groups provided higher quality services, patients with greater underlying risk of adverse health outcomes might be more likely to seek care from them. A potential implication of this finding is that our main results in Tables 3 and 4 may underestimate the effect of the incentive contracts on provider performance. Such a demand response could also plausibly explain the positive effect of the incentive contracts on the incidence of preeclampsia, which women can observe ex ante during pregnancy, compared to PPH or sepsis.

[51] The contract documents emphasized the importance of maintaining appropriate patient referral patterns; this was further reinforced in communication with providers during the visits.

[52] The t-statistics for comparing the input versus control incentive group and the input incentive versus control group are 3.94 and 5.11, respectively.

IV. Conclusion

The use of performance incentives in public service delivery has grown rapidly in developing countries in recent years (Finan, Olken, and Pande 2015; Wagstaff 2015). The World Bank alone currently supports more than 40 such large-scale programs in the health sector (World Bank 2016). However, very little empirical research examines key contract design issues that should guide these programs (Miller and Babiarz 2014). Theory suggests that two central considerations are (i) the trade-off between rewarding the production of outputs versus the use of inputs and (ii) how this trade-off may vary with worker/agent skill. While performance incentives rewarding outputs may encourage innovation and efficiency in context-dependent input choices, they also impose more risk on agents. Moreover, suitable skills may be necessary for agents to innovate or deviate efficiently from prespecified input combinations.

Through a maternity care experiment in India, our paper provides empirical evidence that output and input incentive contracts produced comparable health gains—a reduction in postpartum hemorrhage (PPH) exceeding 20 percent. This result is important given that PPH is the leading cause of maternal mortality worldwide, and India's maternal mortality ratio continues to be very high (174 per 100,000 live births in 2015) (World Health Organization 2015). Moreover, our results suggest that agents (health providers) might have responded differently to the incentive contracts according to their underlying qualifications and skills. Our results indicate that under output incentive contracts, more-qualified providers performed better than less-qualified providers. Under input contracts, our results are less conclusive as we are unable to reject the possibility that providers performed differently depending on their qualification.

We note that our research design assumes that intrinsic motivation is balanced across the treatment arms and our results do not address the question of how incentives interact with intrinsic motivation of providers. Our findings also point to future directions for research on how contract structure might be used to screen different types of providers—for instance, output contracts that reward innovation might be able to attract providers with higher levels of human capital. This is an important topic for future research.

Overall, our findings suggest that the focus on input incentives among many "pay-for-performance" programs in developing country health sectors may be appropriate despite the lack of previous empirical evidence on the underlying rationale (Fritsche, Soeters, and Meessen 2014; Das, Gopalan, and Chandramohan 2016). In particular, health providers in low-income countries often have relatively little training, and our results suggest that output incentives may be particularly ineffective in improving their performance, but that incentives for adherence to established clinical guidelines may be an appropriate strategy.

Appendix

Appendix Table A1—Joint Test of Orthogonality

	Treatment group (1)	Input group (2)	Output group (3)
Female provider	0.039	0.004	0.055
	(0.086)	(0.112)	(0.111)
MBBS plus	−0.069	−0.213	0.018
	(0.084)	(0.114)	(0.111)
Years practicing	0.003	0.005	0.003
	(0.004)	(0.006)	(0.005)
Years clinic operating	0.001	−0.006	0.004
	(0.004)	(0.006)	(0.004)
Constant	0.623	0.592	0.367
	(0.116)	(0.161)	(0.164)
Observations	135	82	97
R^2	0.033	0.068	0.042
F-statistic	0.304	1.021	0.452
p-value	0.875	0.402	0.77

Notes: Robust standard errors are reported in parentheses. The dependent variable in the first specification is an indicator for being in the treatment group; in the second specification it is an indicator for being in the input treatment group (excluding those in the output group); and in the third it is an indicator for being in the output group (excluding those in the input group). Provider characteristics are self-reported and measured through interviews with the provider or with a staff member. The following variables measure provider training: "MBBS plus" is a Bachelor of Medicine degree with a specialization such as obstetrics, "MBBS" is a Bachelor of Medicine degree with no additional specialization, "BAMS" is a degree in Ayurveda medicine. The last two rows report the F-statistic and associated p-value associated with a test that all coefficients jointly equal zero.

Appendix Table A2— Number of Providers by Treatment Group

	Total (observations)	Input (observations)	Output (observations)	Control (observations)	Test of equality (p-value)
In final sample	135	38	53	44	0.078
Attrition	5	2	0	3	
Total	140	40	53	47	

Notes: This table reports counts of the universe of providers identified as eligible for the study by randomly assigned treatment arm. Because providers identified during fieldwork were assigned to study arms based on a randomized list of sequence numbers (unknown to field enumerators, and the sequence was not predictable) it was not possible to ensure an equal number of providers across arms. Providers identified as attritors declined to participate in the study during or after signing the contract. The p-value in the final column is associated with F-tests of joint equality from a regression of treatment indicators on a binary indicator for refusing to participate.

Appendix Table A3—Interaction of Provider Incentives with MBBS Plus on PPH Prevention and Management

	Parenteral oxytocic drugs available			Medicine use to reduce bleeding after delivery			Massage abdomen after delivery		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Input incentives	0.0722	0.0789	0.0600	0.0636	0.0297	0.106	0.0518	0.0739	0.112
	(0.0415)	(0.0446)	(0.0677)	(0.0322)	(0.0289)	(0.0755)	(0.0322)	(0.0422)	(0.0822)
Output incentives	0.0730	0.0728	0.0438	0.0623	0.0407	0.0770	0.00517	−0.0130	0.00272
	(0.0422)	(0.0421)	(0.0581)	(0.0286)	(0.0272)	(0.0802)	(0.0289)	(0.0351)	(0.0731)
MBBS plus		0.0634	0.0371		−0.0026	0.174		0.130	0.0787
		(0.0590)	(0.0926)		(0.0520)	(0.0844)		(0.0639)	(0.0973)
Input × MBBS plus			0.0289			−0.0742			−0.0164
			(0.0808)			(0.114)			(0.114)
Output × MBBS plus			0.0471			0.0101			0.0725
			(0.0682)			(0.102)			(0.101)
District and enumerator fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household- and provider-level controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Control mean	0.932	0.932	0.932	0.460	0.459	0.460	0.517	0.516	0.517
Observations	135	135	135	2,791	2,656	2,656	1,707	1,610	1,610
R^2	0.260	0.300	0.303	0.322	0.340	0.164	0.372	0.396	0.208

Notes: Estimates obtained through OLS. Robust standard errors, clustered at the provider level, are reported in parentheses. All specifications include district and enumerator fixed effects; even columns additionally include household-level controls (mother's age and education; household's caste and house type, i.e., houseless, kutcha, semi-pucca, or pucca; head of household's religion; mother's history of hypertension, diabetes, asthma, hyper- or hypothyroidism, and convulsions; whether the mother has had a previous stomach surgery; whether it is the mother's first pregnancy, number of previous pregnancies, whether the mother has had a stillbirth or abortion, and number of previous children birthed; whether the household owns land, has no literate adults, and owns a Below Poverty Line card) as well as provider-level controls (primary provider's gender, professional qualifications, number of years in practice, and number of years that the facility has been in operation). Dependent variables for columns 4–9 are measured through household surveys fielded between November 2013 and July 2014; see online Appendix for details of measurement. Dependent variable for columns 1–3 measured through interviews with a member of the hospital personnel and is a binary indicator for whether the provider's facility had any parenteral oxytocic drugs available at the time of the survey at the end of the study period.

Appendix Table A4—Interaction of Provider Incentives with MBBS Plus on Manual Removal of Placenta

	Placenta manually removed		
	(1)	(2)	(3)
Input incentives	−0.0786	−0.0475	−0.00615
	(0.0483)	(0.0438)	(0.0608)
Output incentives	−0.0666	−0.0763	−0.0530
	(0.0386)	(0.0372)	(0.0610)
MBBS plus		0.00195	0.0331
		(0.0789)	(0.0840)
Input × MBBS plus			−0.0125
			(0.126)
Output × MBBS plus			0.00381
			(0.0918)
District and enumerator fixed effects	Yes	Yes	Yes
Household- and provider-level controls	No	Yes	Yes
Control mean	0.289	0.289	0.290
Observations	1,665	1,571	1,572
R^2	0.266	0.277	0.0962

Notes: Estimates obtained through OLS. Robust standard errors, clustered at the provider level, are reported in parentheses. All specifications include district and enumerator fixed effects; even columns additionally include household-level controls (mother's age and education; household's caste and house type, i.e., houseless, kutcha, semi-pucca, or pucca; head of household's religion; mother's history of hypertension, diabetes, asthma, hyper- or hypothyroidism, and convulsions; whether the mother has had a previous stomach surgery; whether it is the mother's first pregnancy, number of previous pregnancies, whether the mother has had a stillbirth or abortion, and number of previous children birthed; whether the household owns land, has no literate adults, and owns a Below Poverty Line card) as well as provider-level controls (primary provider's gender, professional qualifications, number of years in practice, and number of years that the facility has been in operation).

Appendix Table A5—Heterogeneity of Effect of Incentive Contracts on PPH Reduction Based on Expectations

	Inputs contract	Outputs contract
Estimated effects from linear combinations		
Treatment + interaction with expectations at twenty-fifth percentile	−0.131	−0.125
	(0.030)	(0.028)
Treatment + interaction with expectations at mean	−0.095	−0.078
	(0.027)	(0.032)
Treatment + interaction with expectations at median	−0.091	−0.073
	(0.028)	(0.033)
Treatment + interaction with expectations at seventy-fifth percentile	−0.051	−0.021
	(0.035)	(0.045)

Notes: Expectations are reported on provider surveys in response to question, "In general, do you think doctors can reduce the incidence rates of PPH?" with responses ranging from 1 (least likely) to 10 (most likely). Linear combination above report estimated incentive effects at twenty-fifth percentile, mean, median, and seventy-fifth percentile of reported expectations. All specifications includes district and enumerator fixed effects, household-level controls (mother's age and education; household's caste and house type, i.e., houseless, kutcha, semi-pucca, or pucca; head of household's religion; mother's history of hypertension, diabetes, asthma, hyper- or hypothyroidism, and convulsions; whether the mother has had a previous stomach surgery; whether it is the mother's first pregnancy, number of previous pregnancies, whether the mother has had a stillbirth or abortion, and number of previous children birthed; whether the household owns land, has no literate adults, and owns a Below Poverty Line card) as well as provider-level controls. The dependent variable (PPH) is measured through household surveys fielded between November 2013 and July 2014; see online Appendix for details of measurement. Standard errors, clustered at the level of the provider, are reported in parentheses.

Appendix Table A6—Predicted Adverse Health Events in Providers' Patient Population

	Input incentive group	Output incentive group	Control group
Mean predicted probability of complications	0.52	0.55	0.46
t-statistic for comparison to input	—	−1.44	−3.94
t-statistic for comparison to output	1.44	—	−5.11

Notes: Mean rates of predicted adverse health events are generated by using the control group subsample to regress an indicator for whether or not any of the four major adverse health outcomes that we study occurred (PPH, preeclampsia, sepsis, and neonatal mortality) on the individual characteristics that we use as individual-level controls as well as district fixed effects. For each woman in our full sample, we then use the resulting parameter estimates to predict the probability of an adverse health event for each woman. *t*-statistics are reported for tests for which there is no difference in means.

REFERENCES

American College of Obstetrics and Gynecology (ACOG). 2011. "Active Management of Third-Stage Labor." American College of Obstetrics and Gynecology. http://www.acog.org/-/media/Districts/District-VIII/ActiveManagementThirdStageLabor.pdf.

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103 (484): 1481–95.

Ashraf, Nava, Oriana Bandiera, Edward Davenport, and Scott S. Lee. 2020. "Losing Prosociality in the Quest for Talent? Sorting, Selection, and Productivity in the Delivery of Public Services." *American Economic Review* 110 (5): 1355–94.

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: 1–17.

Basinga, Paulin, Paul J. Gertler, Agnes Binagwaho, Agnes L.B. Soucat, Jennifer Sturdy, and Christel M.J. Vermeersch. 2011. "Effect on Maternal and Child Health Services in Rwanda of Payment to Primary Health-Care Providers for Performance: An Impact Evaluation." *Lancet* 377 (9775): 1421–28.

Begley, Cecily M., Gillian M.I. Gyte, Declan Devane, William McGuire, and Andrew Weeks. 2011. "Active versus Expectant Management for Women in the Third Stage of Labour." *Cochrane Database of Systematic Reviews* (11) CD007412.

Behrman, Jere R., Susan W. Parker, Petra E. Todd, and Kenneth I. Wolpin. 2015. "Aligning Learning Incentives of Students and Teachers: Results from a Social Experiment in Mexican High Schools." *Journal of Political Economy* 123 (2): 325–64.

Bénabou, Roland, and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review* 96 (5): 1652–78.

Callen, Michael, Saad Gulzar, Ali Hasanain, Muhammad Yasir Khan, and Arman Rezaee. 2015. "Personalities and Public Sector Performance: Evidence from a Health Experiment in Pakistan." NBER Working Paper 21180.

Celhay, Pablo, Paul Gertler, Paula Giovagnoli, and Christel Vermeersch. 2015. "Long Run Effects of Temporary Incentives on Medical Care Productivity." NBER Working Paper 21361.

Das, Ashis, Saji S. Gopalan, and Daniel Chandramohan. 2016. "Effect of Pay for Performance to Improve Quality of Maternal and Child Care in Low- and Middle-Income Countries: A Systematic Review." *BMC Public Health* 16 (1).

Das, Jishnu, and Jeffrey Hammer. 2014. "Quality of Primary Care in Low-Income Countries: Facts and Economics." *Annual Review of Economics* 6: 525–53.

Das, Jishnu, Jeffrey Hammer, and Carolina Sánchez-Paramo. 2012. "The Impact of Recall Periods on Reported Morbidity and Health Seeking Behavior." *Journal of Development Economics* 98 (1): 76–88.

Das, Jishnu, Alaka Holla, Veena Das, Manoj Mohanan, Diana Tabak, and Brian Chan. 2012. "In Urban and Rural India, A Standardized Patient Study Showed Low Levels of Provider Training and Huge Quality Gaps." *Health Affairs* 31 (12): 2774–84.

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

Deci, Edward L., and Richard M. Ryan. 1985. *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum Press.

Donato, Katherine, Grant Miller, Manoj Mohanan, Yulya Truskinovsky, and Marcos Vera-Hernández. 2017. "Personality Traits and Performance Contracts: Evidence from a Field Experiment among Maternity Care Providers in India." *American Economic Review* 107 (5): 506–10.

Dupas, Pascaline, and Edward Miguel. 2016. "Impacts and Determinants of Health Levels in Low-Income Countries." NBER Working Paper 22235.

Fehr, Ernst, and Armin Falk. 2002. "Psychological Foundations of Incentives." *European Economic Review* 46 (4–5): 687–724.

Filippi, V., C. Ronsmans, T. Gandaho, W. Graham, E. Alihonou, and P. Santos. 2000. "Women's Reports of Severe (Near-Miss) Obstetric Complications in Benin." *Studies in Family Planning* 31 (4): 309–24.

Finan, Frederico, Benjamin A. Olken, and Rohini Pande. 2015. "The Personnel Economics of the State." NBER Working Paper 21825.

Fritsche, György Bèla, Robert Soeters, and Bruno Meessen. 2014. *Performance-Based Financing Toolkit*. Washington, DC: World Bank.

Fryer, Roland G., Jr. 2011. "Financial Incentives and Student Achievement: Evidence from Randomized Trials." *Quarterly Journal of Economics* 126 (4): 1755–98.

Ganguly, Nirmal K., N.K. Arora, Sujith J. Chandy, Mohamed Nadeem Fairoze, J.P.S. Gill, Usha Gupta, Shah Hossain, et al. 2011. "Rationalizing Antibiotic Use to Limit Antibiotic Resistance in India." *Indian Journal of Medical Research* 134 (3): 281–94.

Gertler, Paul, Paula Giovagnoli, and Sebastian Martinez. 2014. "Rewarding Provider Performance to Enable a Healthy Start to Life: Evidence from Argentina's Plan Nacer." World Bank Policy Research Working Paper 6884.

Gertler, Paul, and Christel Vermeersch. 2013. "Using Performance Incentives to Improve Medical Care Productivity and Health Outcomes." NBER Working Paper 19046.

Glewwe, Paul, Nauman Ilias, and Michael Kremer. 2010. "Teacher Incentives." *American Economic Journal: Applied Economics* 2 (3): 205–27.

Government of Karnataka. 2016. "Home-Based Newborn Care." National Health Mission. https://karunadu.karnataka.gov.in/hfw/nhm/pages/nbch_hbnc.aspx (accessed July 5, 2021).

Hall, Brain J., and Jeffrey B. Liebman. 1998. "Are CEOs Really Paid Like Bureaucrats?" *Quarterly Journal of Economics* 113 (3): 653–91.

Hirshleifer, Sarojini. 2017. "Incentives for Effort or Outputs? A Field Experiment to Improve Student Performance." https://www.semanticscholar.org/paper/Incentives-for-Effort-or-Outputs-A-Field-Experiment-Hirshleifer/a1b0c68bc2fc49518c5e2d04d89ae3fd226ded78.

Holmstrom, Bengt, and Paul Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, and Organization* 7: 24–52.

Jensen, Michael C., and Kevin J. Murphy. 1990. "Performance Pay and Top-Management Incentives." *Journal of Political Economy* 98 (2): 225–64.

Khalil, Fahad, and Jacques Lawarrée. 1995. "Input versus Output Monitoring: Who Is the Residual Claimant?" *Journal of Economic Theory* 66 (1): 139–57.

Kolstad, Jonathan T. 2013. "Information and Quality When Motivation Is Intrinsic: Evidence from Surgeon Report Cards." *American Economic Review* 103 (7): 2875–2910.

Laffont, Jean-Jacques, and David Martimort. 2002. *The Theory of Incentives: The Principal-Agent Model*. Princeton: Princeton University Press.

Lavy, Victor. 2002. "Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement." *Journal of Political Economy* 110 (6): 1286–1317.

Lazear, Edward P. 2000. "The Power of Incentives." *American Economic Review* 90 (2): 410–14.

Liu, Li, Shefali Oza, Daniel Hogan, Jamie Perin, Igor Rudan, Joy E. Lawn, Simon Cousens, et al. 2015. "Global, Regional, and National Causes of Child Mortality in 2000–13, with Projections to Inform Post-2015 Priorities: An Updated Systematic Analysis." *Lancet* 385 (9966): 430–40.

Luo, Renfu, Grant Miller, Scott Rozelle, Sean Sylvia, and Marcos Vera-Hernández. 2015. "Can Bureaucrats Really Be Paid Like CEOs? School Administrator Incentives for Anemia Reduction in Rural China." NBER Working Paper 21302.

Luo, Renfu, Grant Miller, Scott Rozelle, Sean Sylvia, and Marcos Vera-Hernández. 2020. "Can Bureaucrats Really Be Paid Like CEOs? Substitution between Incentives and Resources among School Administrators in China." *Journal of the European Economic Association* 18 (1): 165–201.

Mahal, Ajay, and Manoj Mohanan. 2006. "Growth of Private Medical Education in India." *Medical Education* 40 (10): 1009–1011.

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 Athony J. Culyer, 457–66. North Holland: Elsevier.

Miller, Grant, Renfu Luo, Linxiu Zhang, Sean Sylvia, Yaojiang Shi, Patricia Foo, Qiran Zhao, Reynaldo Martorell, Alexis Medina and Scott Rozelle. 2012. "Effectiveness of Provider Incentives for Anaemia Reduction in Rural China: A Cluster Randomised Trial." *British Medical Journal* 345 (7870): e4809.

Ministry of Health and Family Welfare (MOHFW). 2009. *Trainees' Handbook for Training of Medical Officers in Pregnancy Care and Management of Common Obstetric Complications.* New Delhi: Government of India.

Mohanan, Manoj, Katherine Donato, Grant Miller, Yulya Truskinovsky, and Marcos Vera-Hérnandez. 2021. "Replication data for: Different Strokes for Different Folks? Experimental Evidence on the Effectiveness of Input and Output Incentive Contracts for Health Care Providers with Varying Skills." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E119741V1.

Mohanan, Manoj, and Grant Miller. 2016. Performance-Based Contracts in Healthcare: Experimental Evaluation of Contracting Based on Inputs and Outcomes. AEA RCT Registry. https://www.socialscienceregistry.org/trials/179/history/9312.

Mohanan, Manoj, Marocs Vera-Hernández, Veena Das, Soledad Giardili, Jeremy D. Goldhaber-Fiebert, Tracy L. Rabin, Sunil S. Raj, Jeremy I. Schwartz, and Aparna Seth. 2015. "The Know-Do Gap in Quality of Health Care for Childhood Diarrhea and Pneumonia in Rural India." *Journal of the American Medical Association Pediatrics* 169 (4): 349–57.

Mol, Ben. W.J., Claire T. Roberts, Shakila Thangaratinam, Laura A. Magee, Christianne J.M. de Groot, and G. Justus Hofmeyr. 2016. "Pre-eclampsia." *Lancet* 387 (10022): 999–1011.

Mony, Prem K., Krishnamurthy Jayanna, Swarnarekha Bhat, Suman V. Rao, Maryann Crockett, Lisa Avery, B.M. Ramesh, Stephen Moses, and James Blanchard. 2015. "Availability of Emergency Neonatal Care in Eight Districts of Karnataka State, Southern India: A Cross-Sectional Study." *BioMed Central Health Services Research* 15.

Mullen, Kathleen J., Richard G. Frank, and Meredith B. Rosenthal. 2010. "Can You Get What You Pay For? Pay-for-Performance and the Quality of Healthcare Providers." *Rand Journal of Economics* 41 (1): 64–91.

Muralidharan, Karthik, and Venkatesh Sundararaman. 2011. "Teacher Performance Pay: Experimental Evidence from India." *Journal of Political Economy* 119 (1): 39–77.

Nair, H., and R. Panda. 2011. "Quality of Maternal Healthcare in India: Has the National Rural Health Mission Made a Difference?" *Journal of Global Health* 1 (1): 79–86.

Olken, Benjamin A., Junko Onishi, and Susan Wong. 2014. "Should Aid Reward Performance? Evidence from a Field Experiment on Health and Education in Indonesia." *American Economic Journal: Applied Economics* 6 (4): 1–34.

Paswan, Balram, S.K. Singh, Hemkhothang Lhungdim, Chander Shekhar, Fred Arnold, Sunita Kishor, Abhishek Singh, et al. 2017. *National Family Health Survey 4: 2015–16*. Mumbai: International Institute for Population Sciences.

Patel, A., S.S. Goudar, S.E. Geller, B.S. Kodkany, S.A. Edlavitch, K. Wagh, S.S. Patted, V.A. Naik, N. Moss, and R.J. Derman. 2006. "Drape Estimation versus Visual Assessment for Estimating Postpartum Hemorrhage." *International Journal of Gynecology and Obstetrics* 93 (3): 220–24.

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.

Phipps, Elizabeth, Devika Prasanna, Wunnie Brima, and Belinda Jim. 2016. "Preeclampsia: Updates in Pathogenesis, Definitions, and Guidelines." *Clinical Journal of the American Society of Nephrology* 11 (6): 1102–13.

Prendergast, Canice. 1999. "The Provision of Incentives in Firms." *Journal of Economic Literature* 37 (1): 7–63.

Prendergast, Canice. 2002. "The Tenuous Trade-Off between Risk and Incentives." *Journal of Political Economy* 110 (5): 1071–1102.

Prendergast, Canice. 2011. "What Have We Learnt about Pay for Performance?" *Economic and Social Review* 42 (2): 113–34.

Raith, Michael. 2008. "Specific Knowledge and Performance Measurement." *Rand Journal of Economics* 39 (4): 1059–79.

Roland, Martin. 2004. "Linking Physicians' Pay to the Quality of Care—A Major Experiment in the United Kingdom." *New England Journal of Medicine* 351 (14): 1448–54.

Rosenthal, Meredith B., Rushika Fernandopulle, HyunSook Ryu Song, and Bruce Landon. 2004. "Paying for Quality: Providers' Incentives for Quality Improvement." *Health Affairs* 23 (2): 127–41.

Say, Lale, Doris Chou, Alison Gemmill, Özge Tunçalp, Ann-Beth Moller, Jane Daniels, A. Metin Gülmezoglu, Marleen Temmerman, and Leontine Alkema. 2014. "Global Causes of Maternal Death: A WHO Systematic Analysis." *Lancet Global Health* 2 (6): e323–33.

Semrau, Katherine E.A., Lisa R. Hirschhorn, Megan Marx Delaney, Vinay P. Singh, Rajiv Saurastri, Narender Sharma, Danielle E. Tuller, et al. 2017. "Outcomes of a Coaching-Based WHO Safe Childbirth Checklist Program in India." *New England Journal of Medicine* 377 (24): 2313–24.

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).

Singh, Prakarsh. 2015. "Performance Pay and Information: Reducing Child Undernutrition in India." *Journal of Economic Behavior and Organization* 112: 141–63.

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.

Stanton, Cynthia K., Barbara Rawlins, Mary Drake, Matias Dos Anjos, David Cantor, Lidia Chongo, Leonardo Chavane, et al. 2013. "Measuring Coverage in MNCH: Testing the Validity of Women's Self-Report of Key Maternal and Newborn Health Interventions during the Peripartum Period in Mozambique." *PLoS ONE* 8 (5).

Steegers, Eric A.P., Peter von Dadelszen, Johannes J. Duvekot, and Robert Pijnenborg. 2010. "Pre-eclampsia." *Lancet* 376 (9741): 631–44.

Stewart, M. Kathryn, and Mario Festin. 1995. "Validation Study of Women's Reporting and Recall of Major Obstetric Complications Treated at the Philippine General Hospital." *International Journal of Gynecology and Obstetrics* 48 (Supplement): S53–66.

Tunçalp, Özge, João Paulo Souza, and Metin Gülmezoglu. 2013. "New WHO Recommendations on Prevention and Treatment of Postpartum Hemorrhage." *International Journal of Gynecology and Obstetrics* 123 (3): 254–56.

United Nations. 2019. *Levels and Trends in Child Mortality: Report 2019.* New York: United Nations Children's Fund. https://www.unicef.org/media/60561/file/UN-IGME-child-mortality-report-2019.pdf (accessed January 31, 2020).

Urner, Fiona, Roland Zimmermann, and Alexander Krafft. 2014. "Manual Removal of the Placenta after Vaginal Delivery: An Unsolved Problem in Obstetrics." *Journal of Pregnancy* 2014.

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.

Wagstaff, Adam. 2015. "Financial Incentives in Health: The Magic Bullet We Were Hoping For?" *World Bank Blogs*, September 9. https://blogs.worldbank.org/developmenttalk/financial-incentives-health-magic-bullet-we-were-hoping.

Westfall, Peter H., and S. Stanley Young. 1993. *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*. New York: John Wiley & Sons.

World Health Organization. 2015. *Trends in Maternal Mortality: 1990–2015*. Geneva: World Health Organization. http://apps.who.int/iris/bitstream/handle/10665/194254/9789241565141_eng.pdf;jsessionid=6B2805F744F1A122CDC6F7072F6576CA?sequence=1.