

DIFFERENT STROKES FOR DIFFERENT FOLKS:
EXPERIMENTAL EVIDENCE ON THE EFFECTIVENESS OF INPUT AND
OUTPUT INCENTIVE CONTRACTS FOR HEALTH CARE PROVIDERS
WITH VARYING SKILLS

Manoj Mohanan
Katherine Donato
Grant Miller
Yulya Truskinovsky
Marcos Vera-Hernández

Working Paper 25499
<http://www.nber.org/papers/w25499>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge MA, 02138
January 2019

Mohanan is lead author. Donato, Miller, Truskinovsky, and Vera-Hernandez contributed equally to the manuscript. 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. Vera-Hernández was supported by European Research Council's Horizon 2020 research and innovation program (No. 695300 – HKADeC-ERC-2015-AdG)). 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, Neeraj Sood, Oriana Bandiera, Paul Gertler, Rohini Pande, Rob Garlick, Victoria Baranov, Xiao Yu Wang, and to audiences at AEA/ASSA 2017, ASHEcon 2016, Barcelona GSE 2016, BREAD/CEPR 2016, Duke, Erasmus, Harvard, iHEA Congress Milan, Tilburg, and U. of S. 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 RCT was registered in the American Economic Association Registry for randomized control trials under trial number AEARCTR-0000179. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2019 by Manoj Mohanan, Katherine Donato, Grant Miller, Yulya Truskinovsky, and Marcos Vera-Hernández. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Different Strokes for Different Folks: Experimental Evidence on the Effectiveness of Input and Output Incentive Contracts for Health Care Providers with Varying Skills
Manoj Mohanan, Katherine Donato, Grant Miller, Yulya Truskinovsky, and Marcos Vera-Hernández
NBER Working Paper No. 25499
January 2019
JEL No. D86,J41,O15

ABSTRACT

A central issue in designing performance incentive contracts is whether to reward the production of outputs versus use of inputs: the former rewards efficiency and innovation in production, while the latter imposes less risk on agents. Agents with varying levels of skill may perform better under different contracts as well – more skilled workers may be better able to innovate, for example. We study these issues empirically through an experiment enabling us to observe and verify outputs (health outcomes) and inputs (adherence to recommended medical treatment) in Indian maternity care. We find that both output and input incentive contracts achieved comparable reductions in post-partum hemorrhage rates, the dimension of maternity care most sensitive to provider behavior and the largest cause of maternal mortality. Interestingly, and in line with theory, providers with advanced qualifications performed better and used new strategies under output incentives, while under input incentives, providers with and without advanced qualifications performed equally.

Manoj Mohanan
Sanford School of Public Policy
201 Science Drive RH128
Durham NC 27708
manoj.mohanan@duke.edu

Yulya Truskinovsky
Wayne State University
2137 Faculty Administration Building
656 W. Kirby St.
Detroit, MI 48202
yulya.truskinovsky@wayne.edu

Katherine Donato
Boston Consulting Group
1900 16th St Mall #1150
Denver, CO 80202
donatoke@gmail.com

Marcos Vera-Hernández
University College London
Department of Economics
Gower St.
London WC1E 6BT
United Kingdom
m.vera@ucl.ac.uk

Grant Miller
CHP/PCOR
Stanford University
117 Encina Commons Stanford,
CA 94305-6019 and NBER
ngmiller@stanford.edu

A data appendix is available at <http://www.nber.org/data-appendix/w25499>
A randomized controlled trials registry entry is available at <https://www.socialscisceregistry.org/trials/179>

1. INTRODUCTION

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, Fernandopulle et al. 2004). A central issue in the design of performance incentives is whether to reward an agent's use of inputs or instead to reward outputs directly (Khalil and Lawarree 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 maternal health care providers to two treatment arms and a control arm. Providers in treatment arms were given incentive contracts rewarding

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 (Lazear 1986, Baker 2002, Prendergast 2002)

performance either for input use (adherence to guidelines for best practices) or production of outputs (good maternal and neonatal health – low levels of post-partum hemorrhage, sepsis, pre-eclampsia, 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 25 and 31 per 1,000 live births, respectively (NHM 2013, Mony, Jayanna et al. 2015). The top three causes of maternal mortality are post-partum hemorrhage, pre-eclampsia, and sepsis, and the major risk factors for neonatal mortality are infections (sepsis and tetanus, for example), pre-term 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% in 2015-16) (Government of India 2016), poor maternal and neonatal health outcomes persist because of low quality maternal health care in medical facilities (NRHM 2015).

The quality of public services such as health and education in developing countries is generally low (Chaudhury, Hammer et al. 2006, Das, Holla et al. 2012, Das and Hammer 2014, Mohanan, Vera-Hernández et al. 2015, Das, Holla 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 (Lavy 2002,

Glewwe, Ilias et al. 2010, Muralidharan and Sundararaman 2011, Behrman, Parker et al. 2015), while incentives based on service delivery indicators such as institutional deliveries, delivery of prenatal care, vaccinations, and healthcare utilization are typically used in the health sector (Basinga, Gertler et al. 2011, Soeters, Peerenboom et al. 2011, Miller, Luo et al. 2012, Gertler and Vermeersch 2013, Ashraf, Bandiera et al. 2014, Gertler, Giovagnoli et al. 2014, Miller and Babiarz 2014, Olken, Onishi et al. 2014, Celhay, Gertler et al. 2015, Dupas and Miguel 2016, Sherry, Bauhoff et al. 2017).^{2,3,4} The predominance of input incentive contracts in the health sector – an environment in which there is often considerable scope for innovation using local/contextual information⁵ – 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 post-partum 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 (pre-eclampsia, sepsis, and

² 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% by the end of the academic year (Miller, Luo et al. 2012, Luo, Miller 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 et al. 2014).

³ 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, Gertler et al. 2011, Peabody, Shimkhada et al. 2011, Soeters, Peerenboom et al. 2011, Van de Poel, Flores et al. 2016)

⁴ 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, Hirschhorn et al. 2017).

⁵ See <http://www.innovationsinhealthcare.org/> for examples of efforts that adopt novel approaches to improving access to care and improving quality of health care.

neonatal survival) did not change in either contract group relative to the control group.⁶ 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 for input and output contracts were INR 13,850 and 56,812 respectively (about US \$252 and \$1033 in 2010).⁷

Health providers with varying levels of skills responded very differently to input- and output-incentive contracts, however. High skill providers with advanced medical training in obstetrics and gynecology in output contracts stated that they had implemented new health delivery strategies and produced better health outcomes, reducing PPH rates by 9 percentage points relative to lower skilled providers with basic medical training but no advanced qualifications. In contrast, there were no significant differences in implementation of new strategies or health outcomes 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,

⁶ Note that of the four outcomes, two (pre-eclampsia and neonatal mortality) were included primarily to address concerns about multi-tasking (i.e., to minimize reductions in effort focused on these conditions). However, providers can do relatively little to prevent pre-eclampsia, and there are relatively many non-medical factors that contribute to neonatal mortality.

⁷ 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.

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 et al. 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.⁸

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.⁹ While there is a wide theoretical literature on this topic (Jensen and Murphy 1990, Holmstrom and Milgrom 1991, Khalil and Lawarree 1995, Hall and Liebman 1998, Prendergast 1999, Lazear 2000, Prendergast 2002, Laffont and Martimort 2009, Prendergast 2011), the empirical literature that explores the relative effectiveness of contracting on inputs vs. 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. 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

⁸ 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.

⁹ We collect detailed information on inputs, using 48 indicators for five key domains of medical care delivered to mothers and their infants throughout pregnancy, delivery, and post-natal care.

extend the growing literatures on the impact of performance incentives and optimal contracts (Callen, Gulzar et al. 2015).

The rest of the paper proceeds as follows: Section 2 provides a simple conceptual framework of input and output contracts, followed by details of the study design, data collection, and analysis in Section 3. Section 4 presents results, including mechanisms that might explain our findings, and Section 5 concludes.

2. 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 set-up, 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 that vary across providers; 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 . Low skilled providers believe that the health production function is $y_i = h(\theta_1^L e_1, \theta_2^L e_2, \varepsilon)$, with $\theta_1^L \neq \theta_1$ and $\theta_2^L \neq \theta_2$. In other words, low skill providers have incorrect beliefs about the productivity shifters. Alternatively, high skill

providers hold correct beliefs – that is, $\theta_1^H = \theta_1$ and $\theta_2^H = \theta_2$. We also assume that the distribution of ε as well as those of θ_1 and θ_2 are independent of provider skill.¹⁰

We assume that both input choices (e_1, e_2) and output (y) are verifiable. Principals know the distribution of productivity shifters, F_{θ_1, θ_2} , but they cannot make contracts with agents contingent on the values of these productivity shifters. In other words, principals cannot take advantage of local/contextual information, reflected in θ_1 and θ_2 , when writing contracts. 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 Lawarree 1995, Prendergast 2002, Prendergast 2011).

An input incentive contract is a function $w(e_1, e_2)$ that remunerates providers according to input levels. The principal will choose $w(e_1, e_2)$ such that

$$\begin{aligned} & \text{Max} \iint h(\theta_1 e_1, \theta_2 e_2, \varepsilon) \partial F_{\theta_1, \theta_2} \partial G_\varepsilon - w(e_1, e_2) \\ & \text{st: } \text{Max}_{\{e_1, e_2\}} [U(w(e_1, e_2)) - v_1(e_1) - v_2(e_2)] \geq \bar{U}, \end{aligned}$$

where \bar{U} is the provider's reservation utility. Implicitly, an input incentive contract is only feasible if the principal can observe input levels (e_1, e_2). Note that 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.¹¹

¹⁰ By way of example, our set-up assumes that the effect of a certain medicine is the same independently of being prescribed by a high or low skill provider, but the two types of providers might differ in how effective they perceive the medicine to be.

¹¹ 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.

Consequently, input-based payments allow the principal to circumvent low skill providers' incorrect beliefs about the productivity shifters. Under input incentive contracts, average health outcomes, $y = \iint h(\theta_1 e_1, \theta_2 e_2, \varepsilon) \partial F_{\theta_1 \theta_2} \partial G_\varepsilon$, are therefore also the same for high and low skill providers.

An output incentive contract is a function $w(y)$ that remunerates providers according to health outcomes produced. In this case, a provider of type $j \in \{H, L\}$ who wants to achieve average health outcome y and believe him/herself to have productivity shifters (θ_1^j, θ_2^j) will choose inputs (e_1, e_2) to:

$$\begin{aligned} \text{Max } \int U(w(h(\theta_1^j e_1, \theta_2^j e_2, \varepsilon))) - v_1(e_1) - v_2(e_2) \partial G_\varepsilon \\ \text{st: } y = \int h(\theta_1^j e_1, \theta_2^j e_2, \varepsilon) \partial G_\varepsilon, \end{aligned}$$

implying that provider input choices (e_1, e_2) depend on their beliefs about their productivity shifters (θ_1, θ_2) .

The model above assumes that both input choices and outputs are verifiable and allows us to consider trade-offs between input- and output incentive contracts. On one hand, provider remuneration under the output incentive contract, $w(y)$, is partly random and not completely under the control of the agent. This risk introduces a distortion in the output incentive contract, requiring principals to compensate agents for this risk. On the other hand, because principals (health authorities) cannot take advantage of local/contextual information (reflected in θ_1 and θ_2) when establishing contracts, an input incentive contract could lead some providers to choose inefficient combinations of inputs (e_1, e_2) . Output incentive contracts can circumvent this by allowing providers to choose (e_1, e_2) according to their own productivity shifters (θ_1, θ_2) .

In the output contract case, high skill providers, who hold correct beliefs about the productivity shifters (θ_1, θ_2) can make more efficient input choices than with input incentive contracts. The amount of inefficiency for low skill providers in our model depends on how incorrect their beliefs about θ_1^L , and θ_2^L are. It therefore remains possible that input incentive contracts deliver more efficient input choices among low skill providers.¹²

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.

3. STUDY DESIGN, INCENTIVE CONTRACT STRUCTURE, DATA COLLECTION, AND ESTIMATION

3.1. Design and Implementation of the Experiment

Our experiment and data collection activities spanned two years, from late 2012 to late 2014.¹³ 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.

3.1.a. 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

¹² Ultimately, the relative efficiency of input- or output incentive contracts depends on a variety of parameters including the amount of risk, providers' degree of risk aversion, the variability in productivity shifters, the proportion of low skill providers, and how misinformed low skill providers are.

¹³ This study was approved by Duke University Office of Human Subjects Research (Pro00031046).

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.¹⁴

3.1.b. 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 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.¹⁵

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

¹⁴ Providers working in large multi-specialty 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.

¹⁵ 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.

participate over the course of the study, and were classified as attriters from the study (2 from the input incentive group and 3 from the control group). Our final analytical sample thus includes 135 providers: 53 providers in outputs arm, 38 providers in inputs arm, and 44 providers in control arm.¹⁶ 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 to an MD in USA 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% level (Appendix Table A2).

3.2. 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).¹⁷ These letters also provided a broad overview of what participation

¹⁶ Further details on enrollment of providers and sample sizes at each stage are included in the pre-analysis plan (<https://www.socialscisearch.org/trials/179>).

¹⁷ 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.

in the study would entail, including future meetings and payments to compensate participating providers for their time to compile patient lists and complete surveys.

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 (Appendix 1 shows each type of contract and accompanying WHO guidelines). 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 INR 150,000 (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.¹⁸ Finally, providers in all three arms were offered INR 2,500 (about US \$45) at each visit as compensation for the time required to participate in the study. This small payment also aimed to develop credibility for future reward payments.

3.2.a. Output Contract Structure

Output incentive payments were offered for achieving low rates of four adverse health outcomes (post-partum hemorrhage (PPH), pre-eclampsia, sepsis, and neonatal mortality) during the study period among a provider's patients. PPH, pre-eclampsia, and sepsis are the three leading causes of maternal mortality globally, accounting for 27%, 14%, and 11% of all maternal deaths (respectively) between 2003-09 (Say, Chou et al. 2014). However, unlike PPH and sepsis, there is little that healthcare providers can do to prevent pre-eclampsia – a hypertensive disorder that occurs during pregnancy. Similarly, only 10.5% of neonatal mortality is attributed to complications during childbirth (Liu, Oza et al. 2015). As a result, we did not anticipate marked incentive effects on pre-eclampsia 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 policy maker, ensuring that total incentive payments do not exceed a fixed budget constraint.

¹⁸ 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.

For neonatal mortality, a provider would receive INR 15,000 unless one of their newborn patients died. For each of the other three maternal health outcomes (PPH, Pre-eclampsia, 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 pre-established incidence rate ceiling \bar{x}_i :

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

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 3 outputs equally. 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):¹⁹

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

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

¹⁹ For example, pre-intervention rates of post-partum hemorrhage (PPH) were estimated at 35 percent ($\bar{x}_{PPH} = 35$) in the study area. Providers could earn $\alpha_{PPH} = \text{Rs. } 850$ (equivalent to about \$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 \$170; if they were able to completely eliminate PPH in their patient population, they would earn \$595.

3.2.b. 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.²⁰ These inputs are categorized into five domains: pregnancy care, childbirth care, counseling for postnatal maternal care, newborn care, and counseling for postnatal newborn care.²¹ 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 pre-established performance floor \underline{x}_i %:

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

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%.²² The final reward payment for each provider was the sum of rewards earned for performance in each of the five domains of care.

3.2.c. 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

²⁰ These were the most up-to-date guidelines at the time of the intervention.

²¹ Details of the measurement of these health inputs are below and in Appendix 2: Calculation of Inputs and Outputs.

²² For example, pre-intervention 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% of appropriate childbirth care according to WHO guidelines. Providers earn $\alpha_{Childbirth\ Care} = \text{Rs. } 750$ (equivalent to about \$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 \$150, and if she were able to provide this level of care for 100% of her patients, she would earn \$525.

collect survey data from their patients and received the same follow-up visits as intervention arm providers.

3.3. 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 indicators. To minimize this concern, we also collected data from approximately 75 households (not used in this analysis) in areas surrounding the 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.²³

Using patient lists, we then aimed to sample 25 women who had recently given birth at the provider's facility.²⁴ Enumerators collected the list of patients and a study team member managing

²³ See page 5 of sample contracts in 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.

²⁴ Power calculations were conducted prior to the data collection. Estimated pre-intervention 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 intra-class 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, post-partum hemorrhage and pre-eclampsia, have at least 85 percent power to detect

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 outcomes²⁵, input use in the five domains of maternity care, and basic socio-demographic information. We aimed to interview every mother within approximately 2 weeks after she gave birth to minimize recall inaccuracy (Das, Hammer et al. 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.²⁶

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, Goudar 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, Ronsmans et al. 2000, Stanton, Rawlins et al. 2013). Second, prior to our study, we

improvements to the target levels. Note that these calculations do not take into account additional precision gained by including covariates.

²⁵ We collected data from household surveys about signs and symptoms for the health outcomes and used algorithms described in the appendix to establish whether a woman had each adverse health outcome or not.

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

conducted our own 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.²⁷

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, Ronsmans et al. 2000, Stanton, Rawlins 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.²⁸ 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).²⁹

²⁷ Results from this validation study to be published in a separate manuscript, and available upon request.

²⁸ 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 Appendix on Calculation of Inputs and Outputs, also available at <https://www.socialsciregistry.org/trials/179>.

²⁹ The Anderson index is calculated as a weighted mean of the standardized values of all inputs within each domain (with variables re-defined 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).

3.4. 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). 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:

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

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 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 re-sampling 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 pre-eclampsia, 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.³⁰

As section 2 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.³¹

4. 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.

4.1. Health Outputs

Table 3 reports estimates of how each incentive contract influences maternal and child health outcomes. Our preferred (pre-specified) 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. Appendix Table A3 shows the unadjusted as well as adjusted p-values for the main results.

³⁰ Our pre-analysis plan also included the evaluation of other potential mechanisms through which incentive contracts might influence outcomes. In a separate paper (Donato, Miller et al. 2017) we examined heterogeneity of program impact as a function of personality traits. We focus attention in this paper on the primary pre-specified analyses (on outputs and inputs).

³¹ Note that the vector X_p includes provider qualification.

In both incentive contract groups, post-partum hemorrhage (PPH) rates declined by nearly identical (and statistically indistinguishable) amounts relative to the control group.³² Column 2 shows that input contract providers reduced PPH incidence 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% and 20% 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 0.031 for the output group (p-values with and without multiple comparison corrections are reported in Appendix Table A3).

We do not find statistically significant changes for other health outcomes after adjusting for multiple comparisons.³³ 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 post-partum bleeding, for example, for which we find evidence in Section 4.2). Alternatively, among the four domains of health outcome, providers have the least control over pre-eclampsia because it is a hypertensive disorder developed earlier during pregnancy – and women generally seek antenatal care from other providers. Furthermore, the biological causes of pre-eclampsia 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, von Dadelszen et al. 2010, Mol, Roberts et al. 2016, Phipps, Prasanna et al. 2016). For sepsis, a key preventive strategy (wearing gloves during delivery) was already practiced among

³² Testing $\beta_{output} = \beta_{input}$, we fail to reject the null hypothesis (p=0.897).

³³ As mentioned earlier, although providers can do little to prevent pre-eclampsia (a hypertensive disorder that occurs during pregnancy) or neonatal mortality (almost 90% 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 pre-eclampsia and sepsis for input and output contracts, only the pre-eclampsia result is marginally significant (p = 0.07) when not adjusting for multiple comparisons.

99% of control group providers, and prophylactic antibiotics are commonly used at high (and inappropriate) rates in rural India, including Karnataka.³⁴

4.2. 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 to mothers shortly after delivery) rose by 0.0773 index points relative to the control group; this estimate is statistically significant (unadjusted p-value = 0.033), but not at conventional levels after correcting for multiple hypotheses testing (p=0.156) – see Appendix Table A4 for full adjusted and unadjusted p-values.³⁵ There were no improvements in the five composite domains of maternal and neonatal care in the input incentive contract group. (In Section 4.5, 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., ‘multi-tasking’ (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).³⁶ Although not pre-specified, we therefore

³⁴ 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, Arora et al. 2011).

³⁵ The magnitude of the increase (0.0773) is not directly interpretable because the weights used to compute the index change the scale (Anderson 2008).

³⁶ Active Management of Third Stage of Labor (AMTSL) recommended by WHO guidelines also includes early cord clamping, controlled traction of the umbilical cord, and trans-abdominal manual massage of the uterus (Urner, Zimmermann et al. 2014). Abdominal massage was included in the 2009 guidelines from Government of India (MOHFW 2009) and was also recommended by Am. Coll. of Obs. and Gyn. at the time (ACOG 2011). The 2012

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).³⁷

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 4.1, 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 (an increase of 25 percent relative to the control group mean of 0.29). 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).³⁸

Additionally, a key corrective clinical action to prevent PPH when the placenta is not delivered normally is manual placenta removal (Urner, Zimmermann et al. 2014). ATMSL, 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 (Begley, Gyte 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 (26

revised guidelines from WHO no longer recommend cord traction or abdominal massage as standard practice (Tunçalp, Souza et al. 2013).

³⁷ 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.

³⁸ 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.

percent reduction), suggesting fewer instances in which corrective action was needed.³⁹ The corresponding estimate in the input arm is less precise, but comparable in magnitude.

4.3. 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. Figures 2 and 3 show the distributions of incentive payments made to providers in treatment arm. *Ex post*, the average payment was much higher in the output contract group (INR56,812, or USD 1033) than in the input contract group (INR13,850, or USD 252).⁴⁰ In each figure, we also construct counterfactual distributions that reflect hypothetical input contract payments to output contract group providers (and vice versa). In general, for the specific contracts that we study, payments for outputs are nearly four times as expensive as payments for inputs. Importantly, this potentially reflects a substantial risk premium required by providers accepting output payment contracts to compensate them for the risk that they are not rewarded for additional effort exerted (because outputs are not fully under their control).

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 for the principal. However, it is important to note that this only applies to the input and output incentive contracts studied here, and it is not generalizable to comparing 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

³⁹ 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.

⁴⁰ Exchange rate 1USD = 55 INR in 2013.

to the principal.⁴¹ Although we were unable to 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 the relevant elasticity to be significant (Luo, Miller et al. 2015).

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

As our conceptual framework in section 2 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 as more skilled providers are better able to choose the optimal combination of inputs using local/contextual information (albeit with less control over contracted outcomes – and therefore more uncertainty about incentive payments).

We examine differences in providers’ behavioral responses to incentive contracts by level of skill, measuring skills according to whether or not providers have medical degrees with specific obstetric training (“MBBS plus” providers) qualifying them to provide maternity care.⁴² Table 6 shows that in the output contract group, “MBBS plus” providers (column 1) produced PPH rates that were 9 percentage points lower on average than providers without obstetric qualifications. In contrast, “MBBS plus” providers performed no better (or worse) than less qualified providers in the input contract group. These results are consistent with output incentive contracts leading

⁴¹ We are grateful to Oriana Bandiera and Paul Gertler for helpful discussions on this point.

⁴² 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 2 to 3 years of training after completing medical school) (Mahal and Mohanan 2006).

providers to use local/contextual information to improve care beyond simple guideline adherence – but only when they also have sufficient complementary skills to do so.

To explore high-skilled provider use of local/contextual information under output incentive contracts further, we also directly examine providers' reports of implementing new delivery strategies since our baseline survey. Table 7 shows that output contracts increased the probability that “MBBS plus” providers implemented new strategies by 0.364 ($0.364 = -0.165 + 0.529$; $se = 0.142$), which is statistically different from zero. In contrast, the input contract did not increase the use of new strategies among “MBBS plus” providers ($0.143 = -0.263 + 0.406$; $se = 0.167$). The first two rows also show that neither type of contract increased the probability that less qualified providers implemented new strategies.

We also investigate whether or not 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). As Appendix Table A5 shows, 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 these differences are imprecise.

Another explanation for differences in provider performance is that MBBS plus providers are not only better trained, but that they also work in better health care facilities. Indeed, facilities in which MBBS plus providers work are better equipped with screening tests for pregnancy disorders, infrastructure for intrapartum care and 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.^{43, 44}

4.5. Expectations and Multi-tasking

Although our incentive contracts generally cover all domains of maternity care provision, a natural concern with performance incentives is ‘multi-tasking’ (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% 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

⁴³ Consistent with our pre-analysis 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 (See Appendix Table A6).

⁴⁴ 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 (se) for outputxMBBS-Plus -0.074 (0.099)

patients' clinical needs, only 9% said neonatal mortality – while 75% said PPH (Figure 4). This pattern of beliefs is consistent with output contract providers diverting effort away from postnatal newborn care (and preventing neonatal mortality) and towards preventing and treating PPH. In contrast, Table 4, Column 10, 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 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.⁴⁵ We estimate the effect of incentive contracts at the 25th, median, and 75th 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 25th percentile (with t-statistic of 4.4) and 5 percentage points (although not statistically significant) at the 75th percentile (See Appendix Table A7). This pattern of results is consistent with an interpretation that providers with high levels of beliefs exert high effort even without incentive contracts with relatively little

⁴⁵ 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).

room for improvement, while providers with low levels of beliefs are able to exert higher effort with incentives.

4.6. 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.⁴⁶ 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 sub-sample to regress an indicator of whether or not any of the four major adverse health outcomes (PPH, pre-eclampsia, 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 A8 reports the means of these predicted probabilities for each study arm.

Both input and output contract providers had patients who were 6-9 percentage points more likely to experience any adverse health event than patients in the control group (a statistically

⁴⁶ The contract documents emphasized the importance of maintaining appropriate patient referral patterns; this was further reinforced in communication with providers during the visits.

significant difference).⁴⁷ 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), we propose that this finding may instead reflect a demand response: as providers in both treatment groups provided higher quality services, patients with greater underlying risk of adverse health outcomes were more likely to seek care from them. An 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. We also note that such a demand response could explain the positive effect of the incentive contracts on the incidence of pre-eclampsia, which women can observe ex ante during pregnancy, compared to PPH or sepsis.

5. CONCLUSION

The use of performance incentives in public service delivery has grown rapidly in developing countries in recent years (Finan, Olken et al. 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 (a) the trade-off between rewarding the production of outputs versus the use of inputs and (b) 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 pre-specified input combinations.

⁴⁷ T-statistics for comparing the input vs. control incentive group and the input incentive vs. control group are 3.94 and 5.11, respectively.

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%. 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 1000 live births in 2015) (World Health Organization 2015). Moreover, agents (health providers) responded very differently to the incentive contracts according to their underlying qualifications and skills. With output incentive contracts, those with advanced qualifications reduced PPH substantially, implementing new delivery strategies to do so – while those lacking appropriate qualifications failed to reduce PPH. Alternatively, those with varying qualifications performed equivalently under input incentive contracts, following guidelines in similar ways.

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 et al. 2014, Das, Gopalan et al. 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.

SUPPLEMENTARY MATERIAL:

Appendix materials for online publication are included at the end of this manuscript.

REFERENCES:

- Anderson, M. L. (2008). "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." JASA: Journal of the American Statistical Association **103**(484): 45.
- Ashraf, N., O. Bandiera and B. K. Jack (2014). "No margin, no mission? A field experiment on incentives for public service delivery." Journal of Public Economics **120**: 1-17.
- Basinga, P., P. Gertler, A. Binagwaho, A. Soucat, J. Sturdy and C. Vermeersch (2011). "Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation." The Lancet **377**(9775): 1421–1428.
- Begley, C. M., G. M. Gyte, D. Devane, W. McGuire and A. Weeks (2011). "Active versus expectant management for women in the third stage of labour." Cochrane Database Syst Rev(11): Cd007412.
- Behrman, J., S. Parker, P. Todd and K. Wolpin (2015). "Aligning Learning Incentives of Students and Teachers: Results from a Social Experiment in Mexican High Schools." Journal of Political Economy **123**(2): 325-364.
- Callen, M., S. Gulzar, A. Hasanain, Y. Khan and A. Rezaee (2015). "Personalities and Public Sector Performance: Evidence from a Health Experiment in Pakistan." National Bureau of Economic Research Working Paper Series **21180**(May).
- Celhay, P., P. Gertler, P. Giovagnoli and C. Vermeersch (2015). Temporary Incentives and Long-Run Medical Care Performance: Evidence from A Field Experiment. NBER Working Paper No. 21361.
- Chaudhury, N., J. Hammer, M. Kremer, K. Muralidharan and F. H. Rogers (2006). "Missing in Action: Teacher and Health Worker Absence in Developing Countries." The Journal of Economic Perspectives **20**(1): 91-116.
- Das, A., S. S. Gopalan and D. 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): 321.
- Das, J. and J. Hammer (2014). "Quality of Primary Care in Low-Income Countries: Facts and Economics." Annual Review of Economics **6**(1): 525-553.
- Das, J., J. Hammer and C. Sánchez-Paramo (2012). "The impact of recall periods on reported morbidity and health seeking behavior." Journal of Development Economics **98**(1): 76-88.
- Das, J., A. Holla, V. Das, M. Mohanan, D. Tabak and B. 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): 2774-2784.
- Das, J., A. Holla, A. Mohpal and K. Muralidharan (2016). "Quality and Accountability in Healthcare Delivery: Audit Evidence from Primary Care Providers in India." American Economic Review **106**(12): 3765-3799.

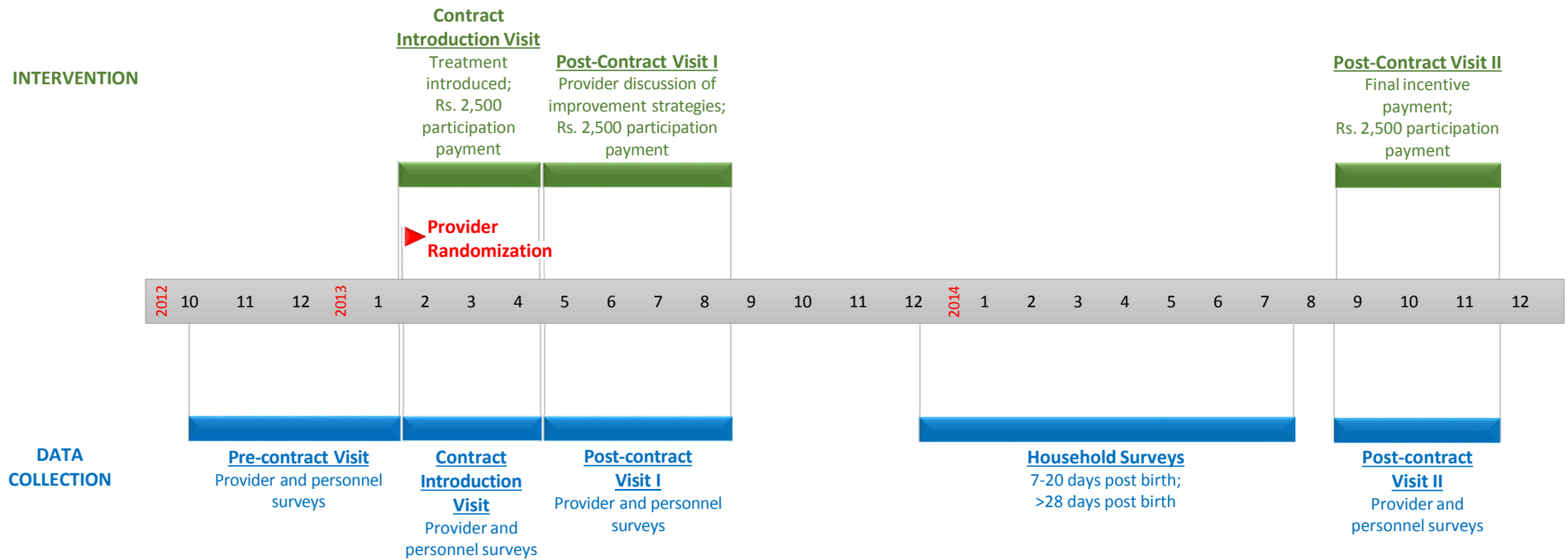
- Donato, K., G. Miller, M. Mohanan, Y. Truskinovsky and M. Vera-Hernández (2017). "Personality Traits and Performance Contracts: Evidence from a Field Experiment among Maternity Care Providers in India." American Economic Review **107**(5): 506-510.
- Dupas, P. and E. Miguel (2016). Impacts and Determinants of Health Levels in Low-Income Countries. NBER Working Paper No. 22235. Cambridge, MA.
- Filippi, V., C. Ronsmans, T. Gandaho and W. Graham (2000). "Women's reports of severe (near-miss) obstetric complications in Benin." Studies in family planning **31**(4): 309.
- Finan, F., B. A. Olken and R. Pande (2015). The Personnel Economics of the State. NBER Working Paper No. 21825.
- Fritsche, G. B., R. Soeters and B. Meessen (2014). Performance-Based Financing Toolkit. Washington, DC, World Bank.
- Ganguly, N., N. Arora, S. Chandy, M. Fairoze, J. P. Gill, U. Gupta, S. Hossain, S. Joglekar, P. Joshi, M. Kakkar, A. Kotwani, A. Rattan, H. Sudarshan, K. Thomas, C. Wattal, A. Easton and R. Laxminarayan (2011). "Rationalizing antibiotic use to limit antibiotic resistance in India [sup]+." Indian Journal of Medical Research **134**: 281.
- Gertler, P., P. Giovagnoli and S. Martinez (2014). Rewarding Performance to Enable a Healthy Start to Life: The Impact of Plan Nacer on Birth Outcomes. Policy Research Working Paper; No. 6884. World Bank, Washington DC.
- Gertler, P. and C. Vermeersch (2013). Using Performance Incentives to Improve Medical Care Productivity and Health Outcomes. NBER Working Paper 19046.
- Glewwe, P., N. Ilias and M. Kremer (2010). "Teacher Incentives." American Economic Journal: Applied Economics **2**(3): 205-227.
- Government of India, M. (2016). State Factsheet - Karnataka, National Family Health Survey - 4. New Delhi: http://rchiips.org/NFHS/pdf/NFHS4/KA_FactSheet.pdf.
- Hall, B. J. and J. B. Liebman (1998). "Are CEOs really paid like bureaucrats?" The Quarterly Journal of Economics **113**(3): 653-691.
- Holmstrom, B. and P. Milgrom (1991). "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." Journal of Law, Economics, and Organization **7**(special issue): 24-52.
- Jensen, M. C. and K. J. Murphy (1990). "Performance Pay and Top-Management Incentives." The Journal of Political Economy **98**(2): 225.
- Khalil, F. and J. Lawarree (1995). "Input versus output monitoring: Who is the residual claimant?" Journal of Economic Theory **66**(1): 139.
- Laffont, J.-J. and D. Martimort (2009). The theory of incentives: the principal-agent model, Princeton University Press.
- Lavy, V. (2002). "Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement." Journal of Political Economy **110**(6): 1286-1317.
- Lazear, E. P. (2000). "The power of incentives." The American Economic Review **90**(2): 410-414.

- Liu, L., S. Oza, D. Hogan, J. Perin, I. Rudan, J. E. Lawn, S. Cousens, C. Mathers and R. E. Black (2015). "Global, regional, and national causes of child mortality in 2000–13, with projections to inform post-2015 priorities: an updated systematic analysis." The Lancet **385**(9966): 430-440.
- Luo, R., G. Miller, S. Rozelle, S. Sylvia and M. Vera-Hernández (2015). Can Bureaucrats Really Be Paid Like CEOs? School Administrator Incentives for Anemia Reduction in Rural China. NBER Working Paper No. 21302.
- Mahal, A. and M. 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. Encyclopedia of Health Economics. A. J. Culyer. San Diego, Elsevier: 457-466.
- Miller, G., R. Luo, L. Zhang, S. Sylvia, Y. Shi, P. Foo, Q. Zhao, R. Martorell, A. Medina and S. Rozelle (2012). "Effectiveness of provider incentives for anaemia reduction in rural China: A cluster randomised trial." BMJ **345**(7870): e4809-e4809.
- Mohanan, M., M. Vera-Hernández, V. Das and et al. (2015). "The know-do gap in quality of health care for childhood diarrhea and pneumonia in rural india." JAMA Pediatrics **169**(4): 349-357.
- Mol, B. W. J., C. T. Roberts, S. Thangaratinam, L. A. Magee, C. J. M. de Groot and G. J. Hofmeyr (2016). "Pre-eclampsia." The Lancet **387**(10022): 999-1011.
- Mony, P. K., K. Jayanna, S. Bhat, S. V. Rao, M. Crockett, L. Avery, B. M. Ramesh, S. Moses and J. Blanchard (2015). "Availability of emergency neonatal care in eight districts of Karnataka state, southern India: a cross-sectional study." BMC Health Services Research **15**(1): 1-9.
- Mullen, K. J., R. G. Frank and M. B. Rosenthal (2010). "Can you get what you pay for? Pay-for-performance and the quality of healthcare providers." The Rand Journal of Economics **41**(1): 64-91.
- Muralidharan, K. and V. Sundararaman (2011). "Teacher Performance Pay: Experimental Evidence from India." Journal of Political Economy **119**(1): 39-77.
- NHM (2013). Statewise Information: Karnataka. <http://nrhm.gov.in/nrhm-in-state/state-wise-information/karnataka.html>, Government of Karnataka. **2016**.
- NRHM (2015). RMNCH+A: Maternal Health Background National Health Mission, MOHFW, Government of India. **2016**.
- Olken, B. A., J. Onishi and S. 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.
- 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 vs. visual assessment for estimating postpartum hemorrhage." International Journal of Gynecology and Obstetrics **93**(3): 220-224.

- Peabody, J., R. Shimkhada, S. Quimbo, J. Florentino, M. Bacate, C. E. McCulloch and O. Solon (2011). "Financial Incentives And Measurement Improved Physicians' Quality Of Care In The Philippines." Health Affairs **30**(4): 773-781.
- Phipps, E., D. Prasanna, W. Brima and B. Jim (2016). "Preeclampsia: Updates in Pathogenesis, Definitions, and Guidelines." Clinical Journal of the American Society of Nephrology **11**(6): 1102-1113.
- Prendergast, C. (1999). "The provision of incentives in firms." Journal of economic literature **37**(1): 7-63.
- Prendergast, C. (2002). "The tenuous trade-off between risk and incentives." Journal of Political Economy **110**(5): 1071-1102.
- Prendergast, C. (2011). "What have we learnt about pay for performance?: Geary Lecture winter 2010." Economic and Social Review **42**(2): 113-134.
- Raith, M. (2008). "Specific Knowledge and Performance Measurement." The Rand journal of economics **39**(4): 1059-1079.
- Roland, M. (2004). "Linking Physicians' Pay to the Quality of Care — A Major Experiment in the United Kingdom." The New England journal of medicine **351**(14): 1448-1454.
- Rosenthal, M. B., R. Fernandopulle, H. R. Song and B. Landon (2004). "Paying For Quality: Providers' Incentives For Quality Improvement." Health affairs **23**(2): 127-141.
- Say, L., D. Chou, A. Gemmill, Ö. Tunçalp, A.-B. Moller, J. Daniels, A. M. Gülmezoglu, M. Temmerman and L. Alkema (2014). "Global causes of maternal death: a WHO systematic analysis." The Lancet Global Health **2**(6): e323-e333.
- Semrau, K. E. A., L. R. Hirschhorn, M. Marx Delaney, V. P. Singh, R. Saurastri, N. Sharma, D. E. Tuller, R. Firestone, S. Lipsitz, N. Dhingra-Kumar, B. S. Kodkany, V. Kumar and A. A. Gawande (2017). "Outcomes of a Coaching-Based WHO Safe Childbirth Checklist Program in India." New England Journal of Medicine **377**(24): 2313-2324.
- Sherry, T. B., S. Bauhoff and M. Mohanan (2017). "Multitasking and Heterogeneous Treatment Effects in Pay-for-Performance in Health Care: Evidence from Rwanda." American Journal of Health Economics: 1-35.
- Singh, P. (2015). "Performance pay and information: Reducing child undernutrition in India." Journal of Economic Behavior & Organization **112**: 141-163.
- Soeters, R., P. B. Peerenboom, P. Mushagalusa and C. Kimanuka (2011). "Performance-Based Financing Experiment Improved Health Care In The Democratic Republic Of Congo." Health Affairs **30**(8): 1518-1527.
- Stanton, C. K., B. Rawlins, M. Drake, M. dos Anjos, D. Cantor, L. Chongo, L. Chavane, M. da Luz Vaz and J. Ricca (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): e60694.
- Stegers, E. A. P., P. von Dadelszen, J. J. Duvekot and R. Pijnenborg (2010). "Pre-eclampsia." The Lancet **376**(9741): 631-644.

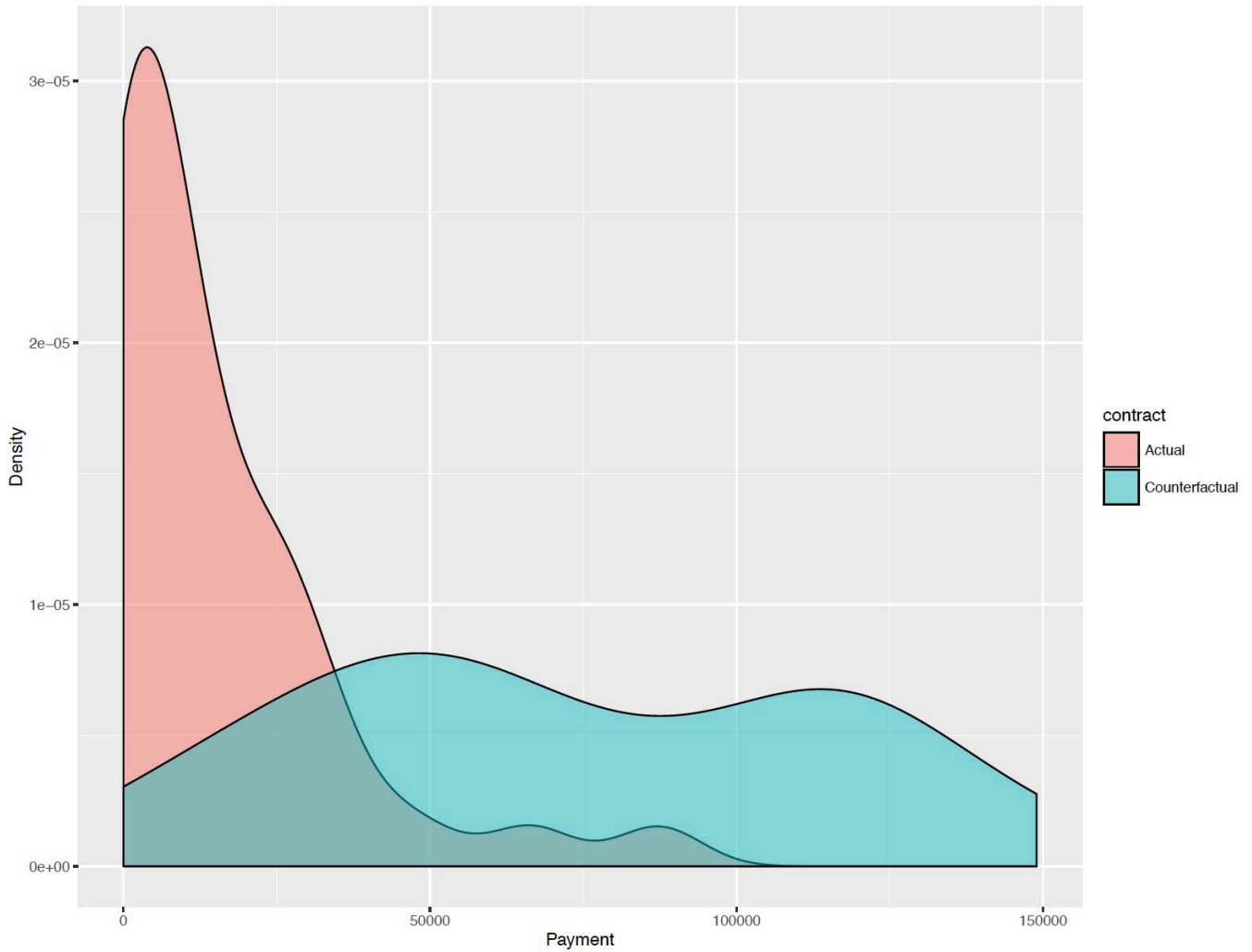
- Stewart, M. K. and M. Festin (1995). "Validation study of women's reporting and recall of major obstetric complications treated at the Philippine General Hospital." International Journal of Gynecology and Obstetrics **48**: S53-S66.
- Tunçalp, Ö., J. P. Souza and M. Gülmezoglu (2013). "New WHO recommendations on prevention and treatment of postpartum hemorrhage." International Journal of Gynecology & Obstetrics **123**(3): 254-256.
- Urner, F., R. Zimmermann and A. Krafft (2014). "Manual Removal of the Placenta after Vaginal Delivery: An Unsolved Problem in Obstetrics." Journal of Pregnancy **2014**: 5.
- Van de Poel, E., G. Flores, P. Ir and O. 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, A. (2015). Financial incentives in health: the magic bullet we were hoping for? W. B. Chief Economist. <https://blogs.worldbank.org/developmenttalk/financial-incentives-health-magic-bullet-we-were-hoping>, World Bank. **2017**.
- Westfall, P. H. and S. S. Young (1993). Resampling-based multiple testing: Examples and methods for p-value adjustment, John Wiley & Sons.
- World Bank (2016). Results Based Financing Health Projects. Washington DC, World Bank. **2016**.
- World Health Organization (2015). Maternal mortality in 1990-2015: India. U. WHO, UNFPA, World Bank Group, and United Nations Population Division and M. M. E. I.-A. Group. http://www.who.int/gho/maternal_health/countries/ind.pdf, World Health Organization.

Figure 1: Timeline of Interventions and Data Collection



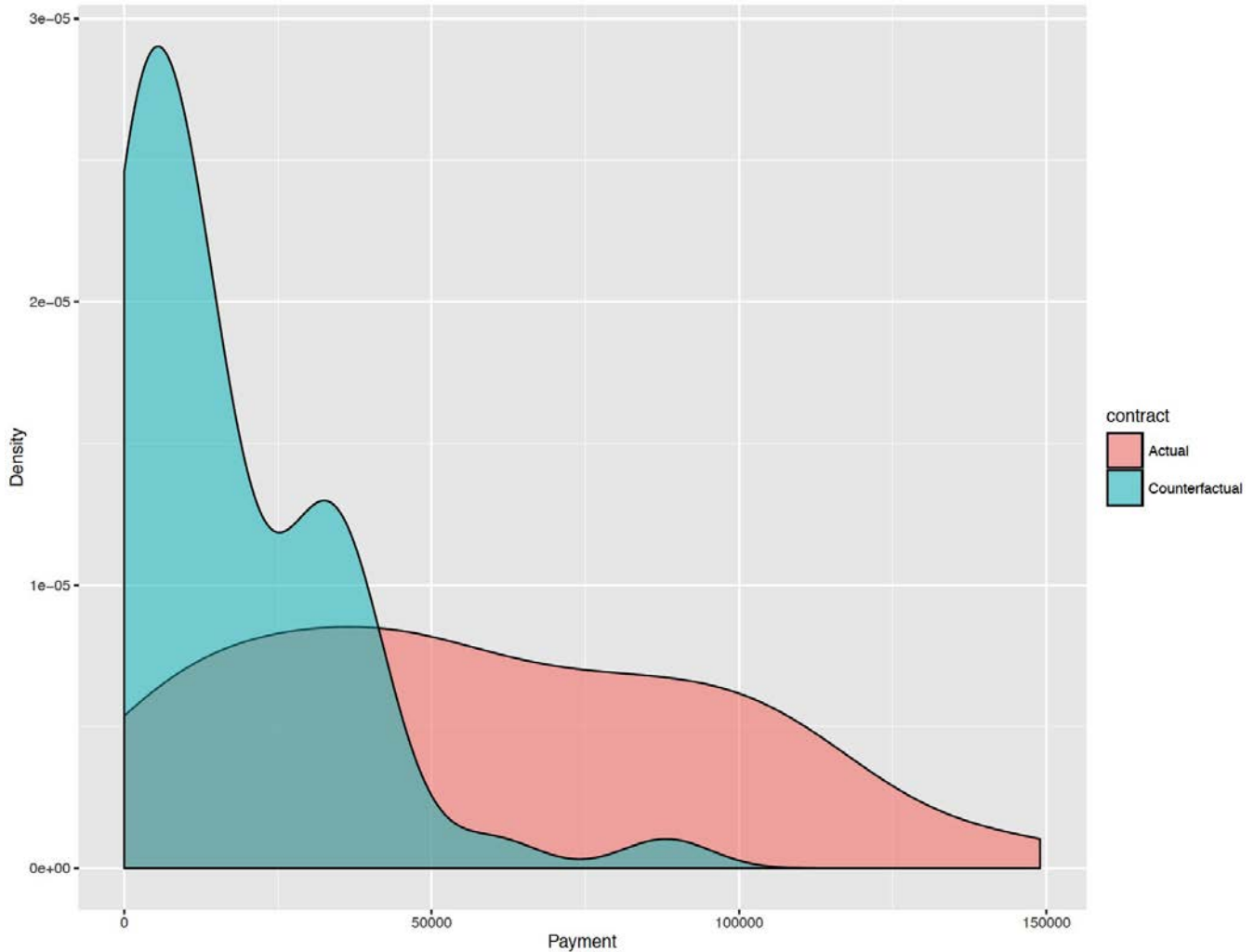
Notes: The timeline shows study implementation period from October 2012 to November 2014. The timing of interventions are labeled (in green) above the timeline, and all data collection and surveys are labeled (in blue) below the timeline. Providers were randomized into treatment arm in early 2014, and 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 - 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.

Figure 2: Distribution of Actual and Counterfactual Payments for Inputs Group



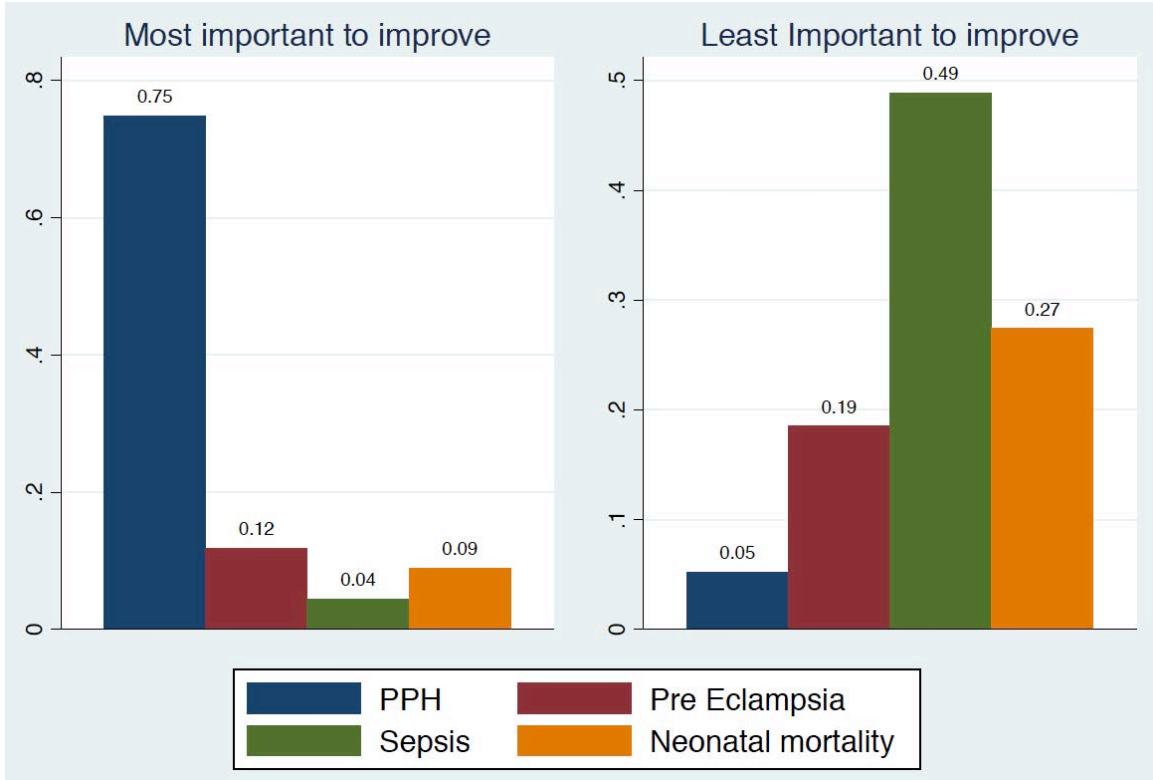
Notes: The distributions show payments made to providers in the input contracts arm. Actual payments are amounts paid out to providers at the end of the experiment based on levels of inputs provided. The distribution labeled “counterfactual” show the payments that might have been made to the same providers if they had been paid based on outcomes instead.

Figure 3: Distribution of Actual and Counterfactual Payments for Outputs Group



Notes: The distributions show payments made to providers in the output contracts arm. Actual payments are amounts paid out to providers at the end of the experiment, based on their performance on contracted outputs. The distribution labeled “counterfactual” show the payments that might have been made to the same providers if they had been paid based on inputs provided instead.

Figure 4: Provider Expectations about Improvements in Outcomes



Notes: Figure on the left shows providers' response to 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 figure on the right shows providers' responses indicating outcomes that they thought were least important to improve among their patients.

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 (percent)	0.56 (0.5)	0.55 (0.5)	0.57 (0.5)	0.55 (0.5)	0.98
MBBS plus (percent)	0.59 (0.49)	0.45 (0.5)	0.64 (0.48)	0.64 (0.49)	0.14
MBBS (percent)	0.21 (0.41)	0.26 (0.45)	0.19 (0.39)	0.2 (0.41)	0.71
BAMS (percent)	0.16 (0.37)	0.26 (0.45)	0.15 (0.36)	0.09 (0.29)	0.13
Other qualification (percent)	0.04 (0.19)	0.03 (0.16)	0.02 (0.14)	0.07 (0.25)	0.52
Provider Age (mean)	47.01 (10.29)	46.42 (9.14)	47.45 (11.33)	46.98 (10.12)	0.89
Years practicing (mean)	19.93 (10.68)	19.68 (9.95)	20.98 (11)	18.89 (11.04)	0.64
Years clinic operating (mean)	17.32 (11.84)	15.5 (11.04)	19.28 (12.78)	16.52 (11.24)	0.3
N	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.

Table 3: Impact of Provider Incentives on Outputs

	Postpartum Hemorrhage		Pre-eclampsia		Sepsis		Neonatal Death	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Input incentives	-0.0842*** (0.0297)	-0.0843*** (0.0284)	0.0312 (0.0450)	0.0573 (0.0434)	0.0333 (0.0228)	0.0369 (0.0253)	-0.0073 (0.0087)	0.0032 (0.0051)
Output incentives	-0.0622** (0.0286)	-0.0742** (0.0294)	0.0466 (0.0325)	0.0611 (0.0328)	0.0065 (0.0198)	0.0208 (0.0225)	-0.0091 (0.0111)	0.0079 (0.0067)
District & 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	2890	2748	2894	2748	2891	2748	2894	2748
R ²	0.266	0.279	0.255	0.271	0.106	0.119	0.0582	0.0565

Notes: Estimates obtained through OLS. Robust standard errors, clustered at the provider level, are reported in parentheses. *, **, and *** denote statistical significance based on p-values less than 0.1, 0.5 and 0.01, 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 (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 appendix for details of measurement.

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 (0.0455)	0.0029 (0.0458)	-0.0203 (0.0338)	0.0146 (0.0284)	0.0380 (0.0390)	0.0422 (0.0392)	-0.0545 (0.0350)	-0.0288 (0.0371)	-0.0650 (0.0576)	-0.0065 (0.0577)
Output incentives	-0.0529 (0.0373)	-0.0551 (0.0401)	-0.0311 (0.0268)	-0.0191 (0.0250)	0.0674 (0.0354)	0.0773 (0.0358)	-0.0285 (0.0322)	-0.0146 (0.0360)	-0.1610*** (0.0435)	-0.1386*** (0.0437)
District & 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	-0.0621	-0.0621	-0.00480	-0.00480	-0.0876	-0.0876	-0.00203	-0.00203	-0.0680	-0.0680
Observations	2890	2748	2894	2748	2891	2748	2894	2748	2894	2748
R ²	0.355	0.361	0.356	0.382	0.406	0.422	0.427	0.447	0.471	0.490

Notes: Estimates obtained through OLS. Robust standard errors, clustered at the provider level, are reported in parentheses. *, **, and *** denote statistical significance based on p-values less than 0.1, 0.5 and 0.01, 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 (houseless, kutch, 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 appendix for details of measurement.

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* (0.0415)	0.0760* (0.0443)	0.0636** (0.0322)	0.0305 (0.0290)	0.0518 (0.0322)	0.0718* (0.0427)	-0.0786 (0.0483)	-0.0504 (0.0437)
Output incentives	0.0730* (0.0422)	0.0694* (0.0417)	0.0623** (0.0286)	0.0382 (0.0266)	0.00517 (0.0289)	-0.0106 (0.0353)	-0.0666* (0.0386)	-0.0722* (0.0381)
District & 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	2791	2656	1707	1610	1665	1571
R ²	0.260	0.292	0.322	0.340	0.372	0.393	0.266	0.277

Notes: Estimates obtained through OLS. Robust standard errors, clustered at the provider level, are reported in parentheses. *, **, and *** denote statistical significance based on p-values less than 0.1, 0.5 and 0.01. 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 (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 appendix for details of measurement. Dependent variable for columns 7 & 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.

Table 6: Impact of Incentives on PPH, by Provider Qualifications

	(1)	(2)
MBBS Plus	-0.002 (0.052)	-0.019 (0.064)
Input incentives	-0.052 (0.043)	-0.079 (0.078)
Output incentives	-0.007 (0.044)	-0.010 (0.065)
Input X MBBS-Plus	-0.054 (0.054)	-0.006 (0.076)
Output X MBBS-Plus	-0.094* (0.052)	-0.093 (0.069)
District & Enumerator fixed effects	Yes	Yes
Household- and provider-level controls	Yes	Yes
Facility Characteristics	No	Yes
Observations	2748	2748
R ²	0.280	0.281

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 (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). Column 2 includes additional controls for availability of emergency obstetric care, screening tests, intrapartum management, human resources and drugs, and their interactions with the treatment arms. The dependent variable (PPH) is measured through household surveys fielded between November 2013 and July 2014; see appendix for details of measurement.

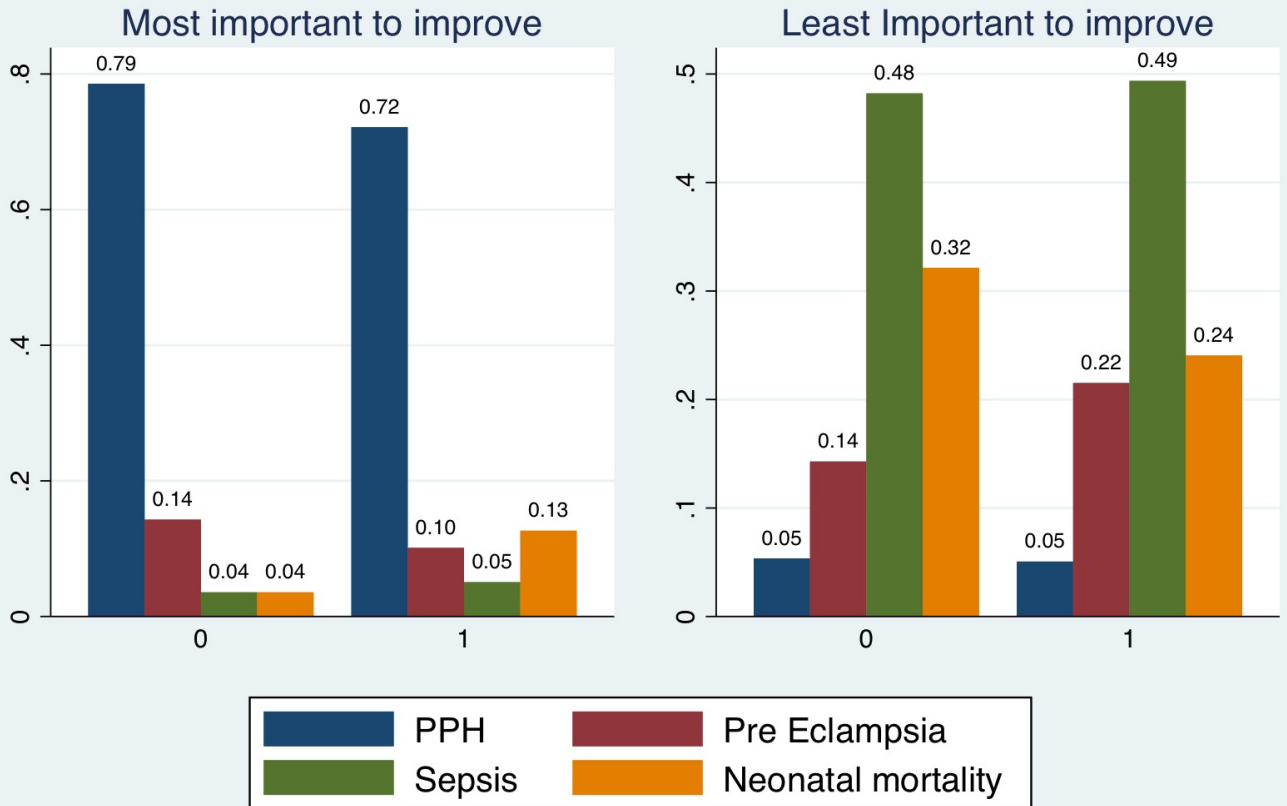
Table 7: Provider Qualifications and Association with Implementing New Strategies

Panel A: Regressions	Implement New Strategies (1)
Input incentives	-0.263 (0.168)
Output incentives	-0.165 (0.158)
Input incentives X MBBS plus	0.406* (0.244)
Output incentives X MBBS plus	0.529** (0.218)
MBBS plus	-0.446*** (0.150)
Panel B: Results from Linear Combinations	
Effect of Input Contracts on MBBS plus	0.143 (0.167)
Effect of Output Contracts on MBBS plus	0.364*** (0.142)
District fixed effects	Yes
Provider-level controls	Yes
Observations	135
R-squared	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.

Appendix Figure 1

Provider expectations about improvements in outcomes by provider qualifications



Appendix Table A1: Joint Test of Orthogonality

	Treatment Group	Input Group	Output Group
	(1)	(2)	(3)
Female provider	0.034 (0.088)	-0.013 (0.113)	0.061 (0.112)
MBBS plus	-0.082 (0.110)	-0.214 (0.136)	-0.003 (0.139)
MBBS	0.357 (0.251)	0.299 (0.261)	0.378 (0.277)
BAMS	0.402 (0.254)	0.362 (0.278)	0.436 (0.293)
Years practicing	0.003 (0.004)	0.004 (0.006)	0.004 (0.005)
Years clinic operating	0.002 (0.003)	-0.004 (0.007)	0.005 (0.004)
Constant	0.258 (0.262)	0.275 (0.298)	-0.021 (0.292)
Observations	135	82	97
R-squared	0.033	0.068	0.042
F stat	0.679	1.093	0.709
p-value	0.667	0.374	0.643

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 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 Arm

	Total (N)	Input (N)	Output (N)	Control (N)	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: Impact of Provider Incentives on Outputs - Correcting for Multiple Hypothesis Testing

	Postpartum Hemorrhage	Sepsis	Neonatal Death	Pre-eclampsia
	(1)	(2)	(3)	(4)
Input incentives	-0.0843 (0.0284)	0.0369 (0.0253)	0.0032 (0.0051)	0.0573 (0.0434)
Unadjusted p-value	0.004	0.147	0.532	0.189
<i>Adjusted p-value</i>	<i>0.010</i>	<i>0.441</i>	<i>0.611</i>	<i>0.188</i>
Output incentives	-0.0742 (0.0294)	0.0208 (0.0225)	0.0079 (0.0067)	0.0611 (0.0328)
Unadjusted p-value	0.013	0.356	0.237	0.065
<i>Adjusted p-value</i>	<i>0.031</i>	<i>0.611</i>	<i>0.566</i>	<i>0.116</i>
District fixed effects	Yes	Yes	Yes	Yes
Enumerator fixed effects	Yes	Yes	Yes	Yes
Household- and provider-level controls	Yes	Yes	Yes	Yes
Control mean	0.365	0.0651	0.0121	0.179
Observations	2748	2748	2748	2748
R ²	0.280	0.119	0.0576	0.271

Notes: Each column reports estimates obtained through an OLS regression; robust standard errors, clustered at the provider level, are reported in parentheses and the associated p-value is reported below. The adjusted p-values are calculated (*in italics*) using the free step-down resampling method and implemented using code from Soledad Giardili and Marcos Vera Hernandez, accounting for the grouping of PPH, Sepsis and NNM into outputs that are primary influenced by care at the time of delivery across two treatment arms. Each specification includes district and enumerator fixed effects and household-level controls (mother's age and education; household's caste and house type (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 are measured through household surveys fielded between November 2013 and July 2014; see appendix for details of measurement.

Appendix Table A4: Impact of Provider Incentives on Inputs - Correcting for Multiple Hypothesis Testing

	Childbirth Care	Postnatal Maternal Care	Newborn Care	Pregnancy Care	Postnatal Newborn Care
	(1)	(2)	(3)		
Input incentives	0.0146 (0.0284)	0.0422 (0.0392)	-0.0288 (0.0371)	0.0029 (0.0458)	-0.0065 (0.0577)
Unadjusted p-value	0.608	0.283	0.439	0.949	0.910
<i>Adjusted p-value</i>	<i>0.879</i>	<i>0.781</i>	<i>0.879</i>	<i>0.948</i>	<i>0.910</i>
Output incentives	-0.0191 (0.0250)	0.0773 (0.0358)	-0.0146 (0.0360)	-0.0551 (0.0401)	-0.1386 (0.0437)
Unadjusted p-value	0.447	0.033	0.686	0.172	0.002
<i>Adjusted p-value</i>	<i>0.879</i>	<i>0.156</i>	<i>0.879</i>	<i>0.276</i>	<i>0.003</i>
District fixed effects	Yes	Yes	Yes	Yes	Yes
Enumerator fixed effects	Yes	Yes	Yes	Yes	Yes
Household- and provider-level controls	Yes	Yes	Yes	Yes	Yes
Control mean	-0.00480	-0.0876	-0.00203	-0.0621	-0.0680
Observations	2739	2739	2740	2748	2748
R ²	0.383	0.423	0.449	0.361	0.490

Notes: Each column reports estimates obtained through an OLS regression; robust standard errors, clustered at the provider level, are reported in parentheses and the associated p-value is reported below. The adjusted p-values (*in italics*) are calculated using the free step-down resampling method and implemented using code from Soledad Giardili and Marcos Vera Hernandez, accounting for the grouping of childbirth care, postnatal maternal care, and newborn care into inputs that are primarily influenced by care at the time of delivery across two treatment arms. Each specification includes district and enumerator fixed effects and household-level controls (mother's age and education; household's caste and house type (houseless, kutcha, semi-pucca, or pucca); head of household's religion; mother's history of hypertension, diabetes, asthma, hyper- of 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 are 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 appendix for details of measurement.

Appendix Table A5: Interaction of Provider Incentives with MBBSPlus 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)	(9)	(10)	(11)	(12)
Input incentives	0.0722* (0.0415)	0.0789* (0.0446)	0.0600 (0.0677)	0.0636** (0.0322)	0.0297 (0.0289)	0.106 (0.0755)	0.0518 (0.0322)	0.0739* (0.0422)	0.112 (0.0823)	-0.0786 (0.0483)	-0.0475 (0.0438)	-0.00615 (0.0608)
Output incentives	0.0730* (0.0422)	0.0728* (0.0421)	0.0438 (0.0581)	0.0623** (0.0286)	0.0407 (0.0272)	0.0770 (0.0802)	0.00517 (0.0289)	-0.0130 (0.0351)	0.00272 (0.0731)	-0.0666* (0.0386)	-0.0763** (0.0372)	-0.0530 (0.0610)
MBBSplus		0.0634 (0.0590)	0.0371 (0.0926)		-0.0376 (0.0296)	-0.0218 (0.0689)		0.168*** (0.0479)	0.0584 (0.0803)		0.00195 (0.0789)	0.0331 (0.0840)
Input X MBBSplus			0.0289 (0.0808)			-0.0742 (0.114)			-0.0164 (0.114)			-0.0125 (0.126)
Output X MBBSplus			0.0471 (0.0682)			0.0101 (0.102)			0.0725 (0.101)			0.00381 (0.0918)
District & enumerator fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household- and provider-level controls	No	Yes	Yes	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	0.289	0.289	0.290
Observations	135	135	135	2791	2656	2656	1707	1610	1610	1665	1571	1572
R ²	0.260	0.300	0.303	0.322	0.340	0.164	0.372	0.396	0.208	0.266	0.277	0.0962

Notes: Estimates obtained through OLS. Robust standard errors, clustered at the provider level, are reported in parentheses. *, **, and *** denote statistical significance based on p-values less than 0.1, 0.5 and 0.01. 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 (houseless, kutchra, 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 appendix for details of measurement. Dependent variable for columns 7 & 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.

Appendix Table A6: Heterogeneity of impact of incentive contracts on PPH
by gender, time preference and risk aversion

	(1) Gender	(2) Time Preference	(3) Risk Aversion
Interaction with Input arm	-0.016 (0.055)	0.054 (0.075)	0.057 (0.063)
Interaction with Output arm	0.034 (0.072)	0.004 (0.064)	0.006 (0.055)
District & Enumerator fixed effects	Yes	Yes	Yes
Household- and provider-level controls	Yes	Yes	Yes
Fully interacted with controls for Facility Characteristics	Yes	Yes	Yes
Observations	2748	2748	2632
R ²	0.280	0.281	0.270

Notes: Column 1 shows coefficient from interaction of providers' gender (female=1) with incentive contract arms. Columns 2 and 3 show interactions with dichotomized variables for time (patient=1) and risk aversion (averse =1). Each specification includes district and enumerator fixed effects, household-level controls (mother's age and education; household's caste and house type (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 appendix for details of measurement.

Appendix Table A7: Heterogeneity of effect of incentive contracts on PPH reduction by expectations

	Inputs Contract	Outputs Contract
<i>Estimated effects from linear combinations</i>		
Treatment + interaction with expectations at 25th percentile	-0.131 (0.030)	-0.125 (0.028)
Treatment + interaction with expectations at mean	-0.096 (0.027)	-0.079 (0.032)
Treatment + interaction with expectations at median	-0.091 (0.028)	-0.073 (0.033)
Treatment + interaction with expectations at 75th percentile	-0.051 (0.035)	-0.021 (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 25th percentile, mean, median and 75th 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 (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 appendix for details of measurement. Standard errors, clustered at the level of the provider, reported in parenthesis.

Appendix Table A8: 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 sub-sample to regress an indicator for whether or not any of the four major adverse health outcomes that we study occurred (PPH, pre-eclampsia, 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 that there is no difference in means.