

DIFFERENT STROKES FOR DIFFERENT FOLKS:

EXPERIMENTAL EVIDENCE ON PERFORMANCE INCENTIVE CONTRACTS

* with
Katherine Donato (Harvard),
Grant Miller (Stanford & NBER),
Yulya Truskinovsky (Harvard SPH),
Marcos Vera Hernández (UCL & IFS)

**Manoj
Mohan**

(Duke /
Sanford Public
Policy)

NIPFP
Conference
“New Thinking
in Health
Policy”
Oct 26-27
2017

FUNDING

This research was made possible by funding and support from:

- 3ie & DFID-India (Grant #OW2:205 co-PIs: Mohanan & Miller),
- World Bank HRITF (Grant number TF099435: PI Mohanan),
- Government of Karnataka.
 - Usual disclaimers about funders apply.

MOTIVATION

- **Performance incentives to correct principal-agent problems**
(Hall & Liebman 1998, Holmstrom & Milgrom 1991, Jensen & Murphy 1990, Khalil & Lawarree 1995, Lazear 2000, Prendergast 2002, Roland 2004, Rosenthal et al. 2004).
 - Reward inputs or outputs?
 - Output contracts assume that agents know the production function and can find optimal combination of inputs in their own contexts
 - With low skilled worker, if inputs are observable & verifiable, rewarding inputs is first best
(Khalil and Lawarree 1995; Prendergast 2002).
 - Do agents perform better when they are able to innovate or do they do better when they ‘follow orders’?
- **P4P in health globally ...** (Finan, Olken, and Pande, 2015)
 - **Most P4P programs in health reward measures of service delivery**
(Ashraf, Bandiera, and Jack 2014, Basinga et al. 2011, Celhay et al. 2015, Dupas and Miguel 2016, Gertler and Vermeersch 2013, Miller and Babiarz 2014)
 - **Evidence of P4P’s impact on health (outcomes) is mixed**
(Miller and Babiarz 2014; Sherry, Bauhoff, and Mohanan 2017)
- **V low quality / provider effort in LMIC ... large ‘Know-do’ gaps**

OBJECTIVES & CONTRIBUTION

- Wide theoretical literature on input / output contracts, but empirical evidence is relatively scarce
- 2 Key Contributions that we make:
 - **Test effectiveness of input and output incentive contracts**
 - First to empirically compare performance of agents when contracted on inputs or outputs (esp in health)
 - Study **differential effectiveness** of input and output contracts among providers with **varying levels of human capital**
 - Extends literature on optimal contracts and performance incentives
- 2 Concerns:
 - Selecting low risk patients
 - Multitasking (Holmstrom and Milgrom 1991)

BRIEF CONCEPTUAL FRAMEWORK

- Agents produce health outcomes $y = h(\theta_1 e_1, \theta_2 e_2, \varepsilon)$.
- Agents with high and low skills; beliefs about θ
- Input contract,
 - Principals can reward specific inputs directly regardless of agents' beliefs on productivity shifters θ
 - But –because θ are local (principals do not know), this could lead providers in input contracts to pick inefficient levels of effort.
- Output contract,
 - Risk: outcome is not fully under agents' control. Premiums need to compensate agents for this risk
 - But if providers have correct beliefs about productivity shifters and hence can choose inputs optimally.
- **Testable implication:** performance will depend on provider skills in output contracts; but independent under input contracts.

THE EXPERIMENT

- **Field experiment:** randomize three types of performance contracts to maternal care providers:
 - (a) Rewards based on outputs (PPH, Sepsis, Pre-eclampsia and Neonatal Mortality)
 - (b) Rewards based on adherence to inputs (WHO / G.o.I. Guidelines)
 - (c) Control
- All providers are given identical WHO / G.o.I. Guidelines as information material
- All providers sign an agreement - A & B sign performance contracts, C sign an agreement to participate in study on MCH
- All providers receive identical participation payments (~\$45 at each visit) as compensation for time to answer surveys etc.

THE CONTRACTS

- Outputs (↓ in PPH, Sepsis, Pre-eclampsia)

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

- Positive payments for reductions below a pre-intervention level of outcomes.
- INR 15000 for avoiding neonatal deaths
- α is set based on allocating available balance across range of improvements in 3 outputs (Exp. 5% min incidence)
- Example: pre-intervention rate of PPH = 35% (\bar{x}). α = INR 850 (~\$17); if $x_i = 25\%$ then the provider gets \$170.

CONTRACTS... CONTD.

- Similar for inputs, except payments for improvements in adherence to input guidelines above a min. level of performance on 5 domains of care:
 - ANC,
 - Childbirth,
 - Post Natal Maternal Counseling,
 - Newborn care,
 - Post Natal **Newborn** Counseling.
- Providers do not know what inputs are measureable (from validation) and what the survey questions would look like.

PROVIDERS AND RANDOMIZATION

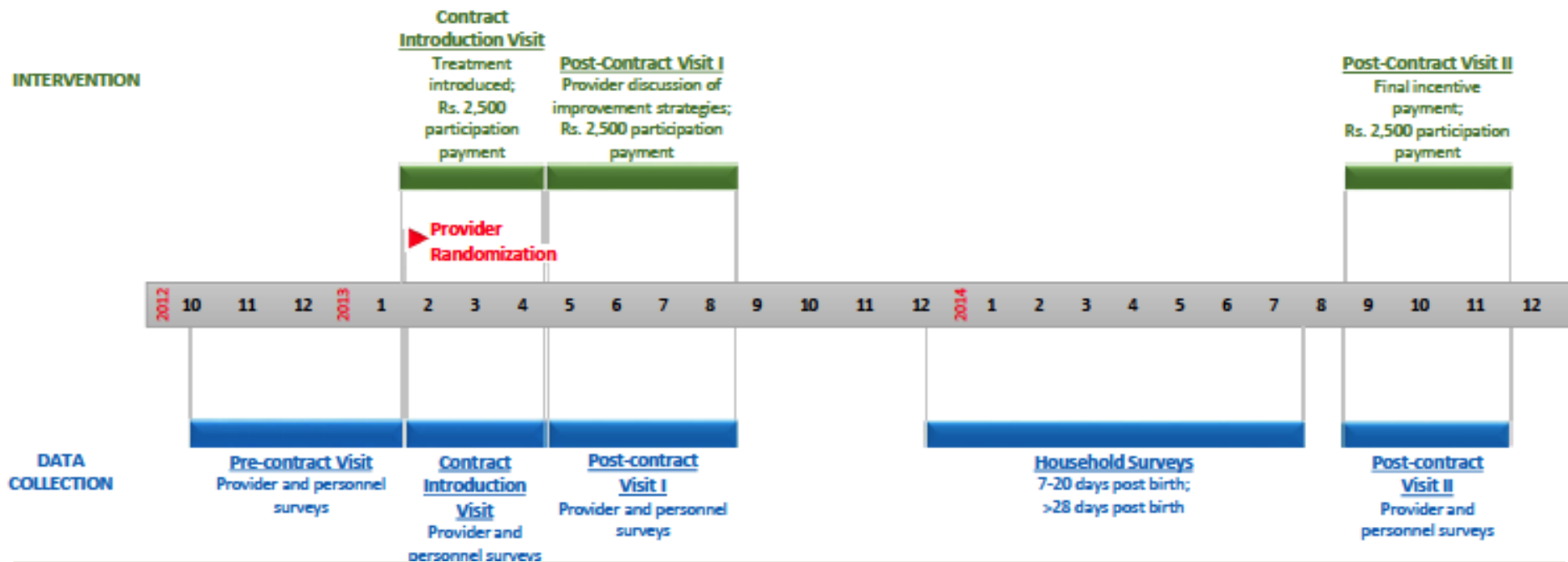
- 135 providers

	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

- Eligibility: Pvt. rural practice in areas not served by large public facilities.
- 120 from govt. survey data, 15 additional found during our field visits
- 56% female, 59% have advanced OBGYN training, 47 yrs, 20 yrs experience, 17 years clinic. (Table 2; for balance see Appendix Table A1)

DATA COLLECTION

■ Timeline:



- Analysis sample: 25 mothers who delivered at study facility
- Additional community sample

DATA - 2

■ Household surveys

- Interview Mothers / attending family member within 2-3 weeks after each delivery.
- Not a cross section, to avoid recall problems (Das et al 2012)
- Questions on health history, symptoms of outputs, and recall of inputs provided (survey included validated and non-validated questions).

■ Provider surveys

- Expectations, demographics, capacity – before contract
- Strategies – 2 months after contract
- Follow up surveys – after contract including qualitative.

ANALYSIS

- Pre-analysis plan on AEA registry

- <https://www.socialscisceregistry.org/trials/179/history/728>

$$y_{ip} = \alpha + \beta T_p + \theta X_p + \gamma Z_i + s_d + \lambda_e + u_{ip},$$

- OLS to estimate effect of treatment, controlling for household and provider characteristics, district and enumerator FE
- Clustered at level of provider
- Multiple hypothesis testing:
 - 4 outputs, 5 input indices, 2 treatment arms
 - Familywise error rate (Westfall & Young 1993)

RESULTS - 1

■ (Table 3)

	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)

- Both groups reduce PPH by about 21% (rel to 36.5% in C)
- PPH most amenable to improvement?

RESULTS - 2

■ Inputs (Table 4)

	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)

- No significant improvements in any of the indices – especially those activities pertaining to PPH
- Improvement in postnatal maternal counseling (6) has unadjusted $p = 0.033$, but after multiple outcomes correction it is not significant.
- PNCC (10) is important – we will revisit in a few slides.

RESULTS - 3

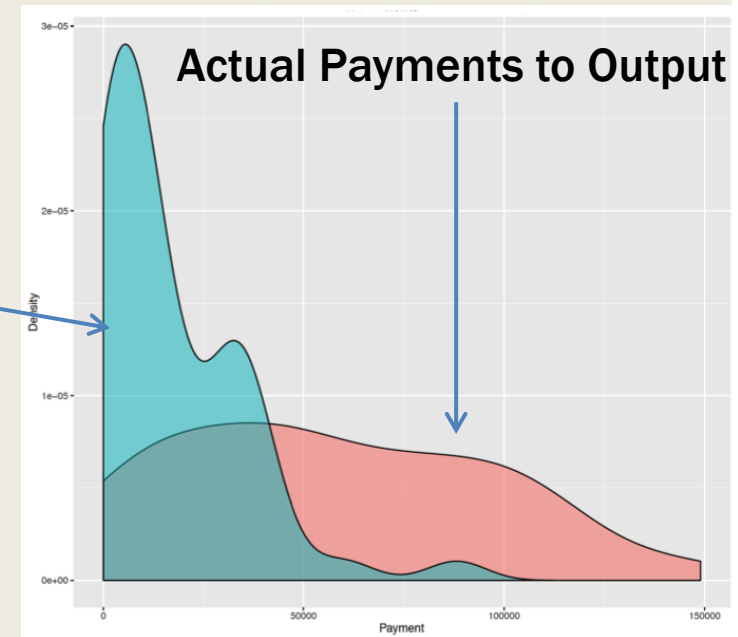
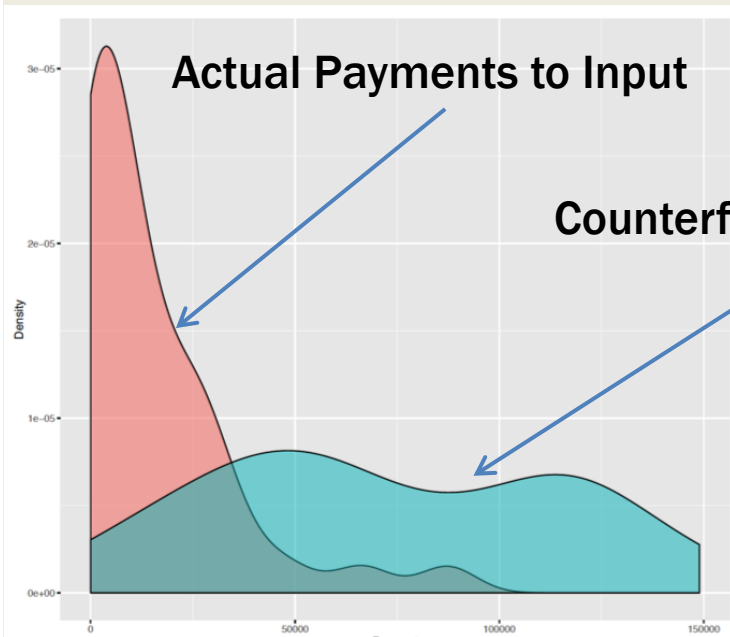
- Potentially difficult to see improvements in inputs due to aggregation of many items into index (Anderson 2008)
 - E.g. Active Management of Third Stage of Labor (AMTSL)
 - Early Cord Clamping, Controlled Cord Traction, Abdominal Massage note
 - 2 specific actions that are most closely related to PPH:
 - Parenteral Oxytocic Drugs and
 - Manual Removal of Placenta (potentially reflects complications)
- Table 5:

	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)

- Manual removal of placenta is conducted less often (7/27 = 26%)

COST OF CONTRACTS

- Ex-post, we see that average payments to outputs was much higher than to input contracts (\$1033 v/s \$252)



- Potentially reflects the risk premium for output contracts
- In our setting, input contract was more efficient
 - Unable to make generalizable inference about efficiency.

RESULTS ON TYPE OF AGENTS

TABLE VI
IMPACT OF INCENTIVES ON POST PARTUM HEMORRHAGE,
BY PROVIDER QUALIFICATIONS

MBBS Plus	-0.002 (0.052)
Input incentives	-0.052 (0.043)
Output incentives	-0.007 (0.044)
Input X MBBS-Plus	-0.054 (0.054)
Output X MBBS-Plus	-0.094* (0.052)
District & Enumerator fixed effects	Yes
Household- and provider-level controls	Yes
Observations	2748
R ²	0.280

- Our conce
- determin
- Innovate
- We look

skills would
cts
e latter
) v/s others

RESULTS ON TYPE OF AGENTS

- Do MBBS+ providers Innovate more?

TABLE VII

PROVIDER QUALIFICATIONS AND ASSOCIATION WITH IMPLEMENTING NEW STRATEGIES

	Implement New Strategies
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

RESULTS ON TYPE OF AGENTS

- Table 6:
 - **OUTPUT contract: MBBS+ providers reduced PPH 9pp more** than other providers
 - INPUT contract: No better or worse
- Exploring whether the MBBS+ folks innovated more (Table 7):
 - **OUTPUT contract: increased Pr(new strategy) for MBBS+ providers**
 - Lincom coeff = 0.36 (se = 0.14)
 - INPUT contract: No increase; lincom coeff = 0.14 (se = 0.17)

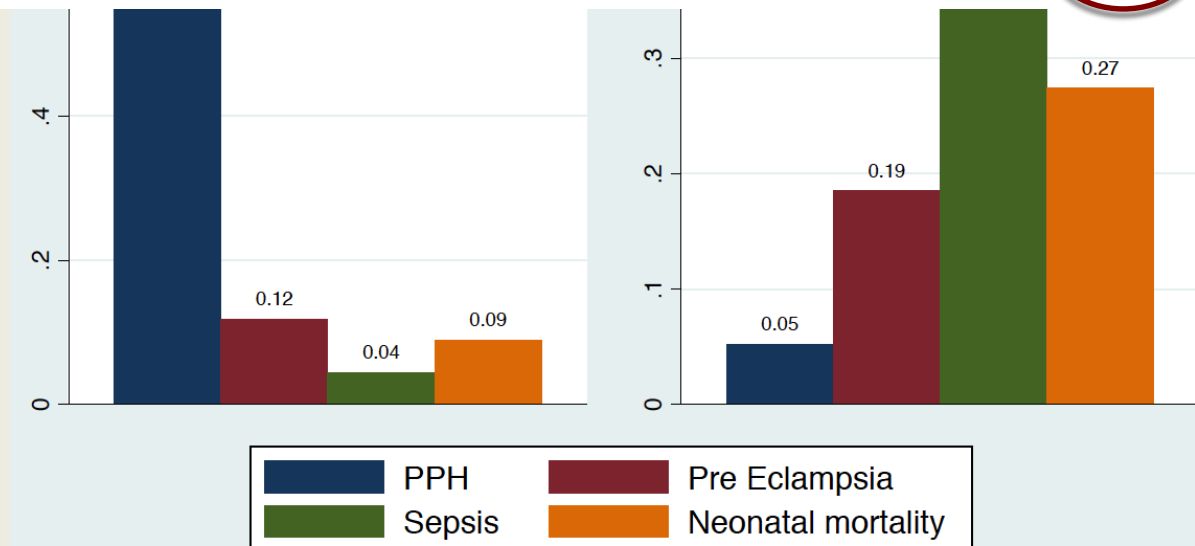
WHY PPH? MULTITASKING?

- Also relates to reduction in PNCC (col 10 in T4) in output arm
- 75% providers thought PPH was most important to improve care of their patients

	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)

- Col 10 – refers to counseling about postnatal care

- No change in input “followed orders” v/s Reduction in output



CONCLUSIONS & POLICY IMPLICATIONS

- Output and Input contracts can achieve comparable gains – and also reduce PPH significantly (major health issue)
- Heterogeneity based on skills:
 - With high skilled workers, output contracts might induce better performance
 - In contrast, output contracts with low skilled workers might not be as effective
- Current focus of ongoing incentive programs globally to reward inputs might in fact be appropriate despite lack of previous empirical evidence on the rationale for this choice.

Acknowledgements

This research was made possible by funding and support from 3ie and DFID-India (Grant number OW2: 205 co-PIs: Mohanan and Miller), World Bank HRITF (Grant number TF099435: PI Mohanan) and Government of Karnataka.

We are grateful, for comments and suggestions, to Alessandra Voena, Alessandro Tarozzi, Amar Hamoudi, Duncan Thomas, Erica Field, Jerry La Forgia, Jishnu Das, Imran Rasul, Meredith Rosenthal, Michael Callen, Nava Ashraf, 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 University, and University of Southern California.

Manveen Kohli provided excellent project management. We are thankful to Kultar Singh, Swapnil Shekhar, and Anil Lobo from Sambodhi, as well as the field team for project implementation and data collection.

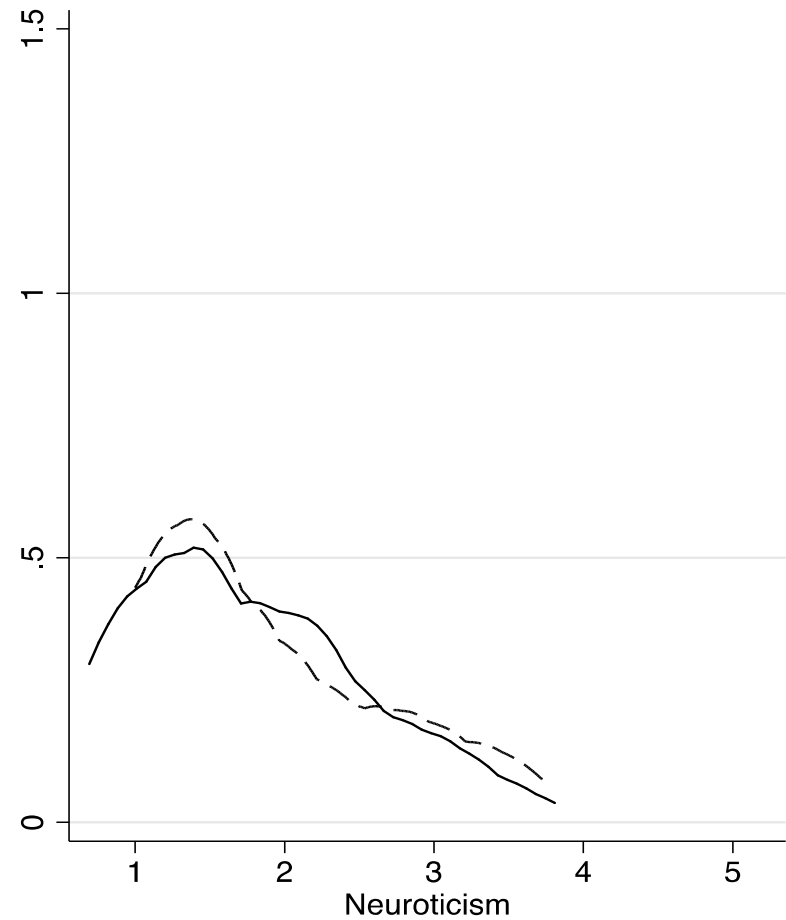
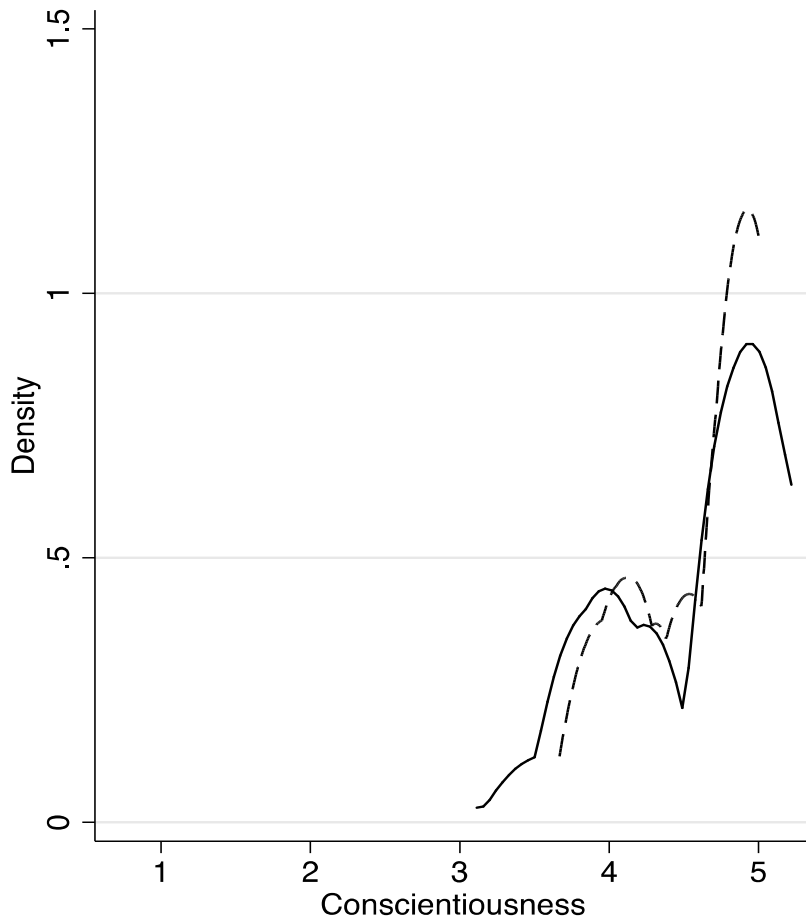
We gratefully acknowledge the support we received from World Bank (Paolo Belli, Patrick Mullen, and Vikram Rajan) and the Government of Karnataka (Vandita Sharma, Selva Kumar, Suresh Mohammed, Raghavendra Jannu, Atul Tiwari, Dr. Nagaraj, Dr. Sridhar, Dr. Prakash Kumar, Dr. Amruteshwari, and several others). We are especially grateful to the many doctors and clinical experts who provided valuable guidance and feedback, including Matthews Mathai, Dinesh Agarwal, Ayaba Worjolah, Vinod Paul, Sharad Iyengar, Kirti Iyengar, Amarjit Singh, Suneeta Mittal, Lalit Baveja, and Sunesh Kumar

PERSONALITY TRAITS & PERFORMANCE CONTRACTS

AMERICAN ECONOMIC REVIEW
VOL. 107, NO. 5, MAY 2017
(pp. 506-10)

- Growing literature on incentives, performance, and personality traits
- Focus on Conscientiousness and Neuroticism – correlated with labor market outcomes (Borghans et al. 2008, Heckman, Stixrud and Urzua 2006, Heckman and Rubinstein 2001)
 - Conscientiousness : dependability, organization skills, perseverance, and achievement oriented thinking
 - Neuroticism: the converse of emotional stability – is associated with anxiety, worry, anger, and insecurity

CONSCIENTIOUSNESS & NEUROTICISM



— Control Arm - - - - Incentive Arm

Note: Providers in incentive arm received output-based contracts that rewarded improvements in maternal health outcomes

PERSONALITY TRAITS (AER P&P)

- Conscientious providers do relatively better absent incentives.
- Beneficial effect of incentive is **weaker among high Cons. providers**
 - At 25th percentile of Conscientiousness (4.3/5), the incentive contracts reduce PPH risk by 13.3 %pts.
 - At 75th percentile of C. (5/5), no statistically significant effect
- No evidence of association b/w neuroticism and performance in absence of incentives (could be due to selection into MD?)
- Performance improvement from incentives is amplified among low Neuroticism (high emotional stability):
 - 13%point reduction at 25th percentile (1.25/5),
 - No significant results at 75th percentile.
 - “Choking under pressure” hypothesis, -- performance deteriorates due to over-arousal and distraction that accompany high stakes (Ariely et al. 2009, Baumeister 1984, Yu 2015)

PERSONALITY TRAITS (AER P&P)

Panel A: Regression Results	PPH	
	(1)	(2)
Incentive	-1.133*** (0.284)	-0.255*** (0.079)
Conscientiousness	-0.193*** (0.057)	
Conscientiousness X Incentive	0.231*** (0.063)	
Neuroticism		-0.0329 (0.039)
Neuroticism X Incentive		0.0997** (0.048)
Panel B: Linear Combination Results	Conscientiousness	Neuroticism
Treatment + interaction at 25th percentile	-0.133*** (0.032)	-0.13*** (0.032)
Treatment + interaction at mean	-0.063** (0.032)	-0.074** (0.031)
Treatment + interaction at 75th percentile	0.021 (0.045)	-0.031 (0.044)
N	1993	1993
R-sq	0.297	0.296
Dep Var Mean	0.364	

* p<0.10 ** p<0.05 *** p<0.01

CONCLUSIONS & POLICY IMPLICATIONS

- Output and Input contracts can achieve comparable gains – and also reduce PPH significantly (major health issue)
- Heterogeneity based on skills:
 - With **high skilled workers**, output contracts might induce better performance
 - Output contracts with low skilled workers might not be as effective
- Personality traits
 - Evidence of significant **heterogeneity by personality traits** – both high conscientiousness and high neuroticism providers don't show improvements with incentives.
 - Among high conscientiousness providers, the dampened effect is suggestive of crowding out.
 - Among high neuroticism (the converse of emotional stability) these results are consistent with “choking”
- Current focus of ongoing incentive programs globally to reward inputs might in fact be appropriate despite lack of previous empirical evidence on the rationale for this choice.