



# Three Field Experiments on Incentives for Health Workers

## Citation

Lee, Scott S. 2015. Three Field Experiments on Incentives for Health Workers. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences.

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:17467500>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

# Three Field Experiments on Incentives for Health Workers

A dissertation presented

by

Scott Sang-Hyun Lee

to

The Committee on Higher Degrees in Health Policy

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Health Policy

Harvard University

Cambridge, Massachusetts

April 2015

© 2015 *Scott Sang-Hyun Lee*

All rights reserved.

# Three Field Experiments on Incentives for Health Workers

## Abstract

The economic study of incentives in firms has traditionally focused on one type of incentive—pecuniary—and one causal mechanism—the direct effect of incentives on effort. This dissertation uses three randomized field experiments to explore non-traditional incentives, and non-traditional incentive effects, in the setting of health care delivery.

The first experiment (jointly authored with Nava Ashraf and Oriana Bandiera) addresses an under-appreciated phenomenon: incentives affect not only the effort of agents on the job, but also the selection of agents into the job. We collaborate with the Government of Zambia to experimentally vary the salience of career incentives in a newly created health worker position when recruiting agents nationally. We find that making career incentives salient at the recruitment stage attracts health workers who are more effective at delivering health services, with administrative data showing an improvement in institutional deliveries, child health visits, and immunization rates in the treatment areas. While career incentives attract agents who differ on observables (e.g., they have higher skills and career ambitions), 91% of the performance gap is due to unobservables. The results highlight the importance of incentive design at

the recruitment stage for attracting high performers who cannot be identified on observables alone.

The second and third experiments examine the use of non-pecuniary incentives in health care. The second experiment (jointly authored with Nava Ashraf and Oriana Bandiera) studies non-monetary awards. Awards may affect behavior through several mechanisms: by conferring employer recognition, by enhancing social visibility, and by facilitating social comparison. In a nationwide health worker training program in Zambia, we design a field experiment to unbundle these mechanisms. We find that employer recognition and social visibility increase performance, while social comparison reduces it, especially for low-ability trainees. These effects appear when treatments are announced and persist through training. The findings are consistent with a model of optimal expectations in which low-ability individuals exert low effort in order to avoid unfavorable information about their relative ability. The results highlight the importance of anticipating the distributional consequences of incentives in settings in which the performance of each worker affects social welfare.

The third experiment turns from extrinsic incentives (such as career opportunities and non-monetary awards) to “intrinsic incentives”—that is, incentives that make work more intrinsically rewarding. In the context of a rural health worker program in India, I develop and test a novel, mobile phone-based self-tracking app designed to increase agents’ intrinsic returns to effort. At nine months of follow-up, the self-tracking app leads to a 27% increase in performance as measured by the main job task (home visits). Moreover, the app is most effective when it leverages pre-existing intrinsic motivation: it produces a 46% increase in performance in the top tercile of intrinsically motivated workers, but no improvement in the bottom tercile. Evidence from survey and performance data indicates that the treatment effect is mediated primarily by making effort more intrinsically rewarding, and not by other mechanisms

such as providing implicit extrinsic incentives. The results suggest the potential for wider use of intrinsic incentives that may increase performance at low cost, when agents are intrinsically motivated.

# Contents

<b>1 Do-Gooders and Go-Getters: Career Incentives, Selection, and Performance in Public Service Delivery</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.2 Context and research design . . . . .	8
1.3 The effect of career incentives on performance via selection . . . . .	22
1.4 Impact on facility utilization and health practices and outcomes . . . . .	37
1.5 Selection on observables vs. unobservables . . . . .	46
1.6 Conclusion . . . . .	53
<b>2 Awards Unbundled: Evidence from a Natural Field Experiment</b>	<b>57</b>
2.1 Introduction . . . . .	57
2.2 Experimental design . . . . .	63
2.3 Analysis . . . . .	69
2.4 Interpretation . . . . .	82
2.5 Conclusion . . . . .	86
<b>3 Intrinsic Incentives: A Field Experiment on Leveraging Intrinsic Motivation in Public Service Delivery</b>	<b>88</b>
3.1 Introduction . . . . .	88

3.2	Framework . . . . .	93
3.3	Context . . . . .	97
3.4	Experimental design . . . . .	101
3.5	Analysis . . . . .	112
3.6	Conclusion . . . . .	130
	<b>Bibliography</b>	<b>132</b>
	<b>A Appendix: Do-Gooders and Go-Getters</b>	<b>142</b>
A.1	Differences in sorting vs. differences in recruitment . . . . .	142
A.2	Data Appendix . . . . .	148
A.3	District Instruction Appendix . . . . .	160
	<b>B Appendix: Awards Unbundled</b>	<b>169</b>
	<b>C Appendix: Intrinsic Incentives</b>	<b>176</b>
C.1	Data sources . . . . .	176
C.2	Psychometric scales: list of statements . . . . .	178



# Chapter 1

## Do-Gooders and Go-Getters: Career Incentives, Selection, and Performance in Public Service Delivery<sup>1</sup>

### 1.1 Introduction

The study of how individuals sort into jobs according to their preferences, skills, and the jobs' own attributes, such as different incentive packages, has a long tradition in economics (Roy, 1951). This sorting gives organizations a tool to attract the

---

<sup>1</sup>Jointly authored with Nava Ashraf and Oriana Bandiera. We thank the Ministry of Health of Zambia and especially Mrs. Mutinta Musonda for partnership on this project. We thank the IGC, JPAL Governance Initiative and HBS DFRD for financial support. We also thank Philippe Aghion, Charles Angelucci, Roland Benabou, Tim Besley, Pedro Bordalo, Gharad Bryan, Robin Burgess, Greg Fischer, Matt Gentzkow, Paul Gertler, Maitreesh Ghatak, Bob Gibbons, Brian Hall, Kelsey Jack, Alan Krueger, Gerard Padro, Imran Rasul, Vandana Sharma, Jesse Shapiro, Bryce Millet Steinberg, Andrei Shleifer, Dmitry Taubinsky and Jean Tirole for comments, as well as seminar participants at several institutions. Adam Grant, Amy Wrzesniewski, and Patricia Satterstrom kindly provided guidance on psychometric scales. We thank Kristin Johnson, Conceptor Chilopa, Mardieh Dennis, Madeleen Husselman, Alister Kandyata, Allan Lalisan, Mashekwa Maboshe, Elena Moroz, Shotaro Nakamura, Sara Lowes, and Sandy Tsai, for the excellent research assistance and the Clinton Health Access Initiative in Zambia for their collaboration.

“right” employees and, indeed, understanding the selection effect of incentives is a core question in the economics of organizations (Lazear & Oyer, 2012, Oyer & Schaefer, 2011).<sup>2</sup> The stakes are large because the dispersion of employee quality is substantial and a key determinant of performance in settings as diverse as firms and schools (see e.g. Abowd & Kramarz, 1999; Staiger & Rockoff, 2010) . Incentives that attract high quality employees can therefore have a significant effect on the success of these diverse organizations.

In this paper, we test whether incentives, in the form of promotion prospects and career advancement, affect who self-selects into a public health job and, through selection, their performance while in service. We collaborate with the Government of Zambia to design and implement a nationwide field experiment when recruiting for new health worker position, the Community Health Assistant (CHA). This is a large recruitment drive that aims to substantially increase health staff numbers in targeted communities: recruiting agents who deliver health services effectively thus has important welfare implications.<sup>3</sup>

The key challenge in identifying the selection effect of incentives is that any incentive scheme that affects selection at the recruitment stage also affects effort once agents are hired. Our identification strategy relies on the fact that, since the CHA position is new, potential applicants are uncertain about the potential for career advancement. This allows us to experimentally vary the salience of career incentives

---

<sup>2</sup>Whether high-powered incentives can attract agents who will perform well is ambiguous. Incentive schemes that reward good performance should attract agents with the skills needed to perform well on incentivized tasks (Lazear, 2000). At the same time, high-powered incentives might crowd out other desirable traits, like pro-sociality, that lead to good performance on tasks that cannot be incentivized (for instance by sending a signal about the nature of the job, as in Bénabou & Tirole, 2003 and Deserranno, 2014)

<sup>3</sup>In the average community in our sample, the arrival of two CHAs represents a 133% increase in health staff.

at the recruitment stage, while providing the same actual incentives to all agents once hired. The difference in performance between agents recruited with salient career incentives and those recruited without identifies the effect of career incentives on performance through selection. There are two reasons why advertising career incentives at the recruitment stage might affect performance even if all agents face the same incentives once hired: (i) agents with traits that lead to better performance are more likely to apply and (ii) agents whose elasticity of effort with respect to career incentives is high are more likely to apply. Both are selection effects that affect performance because they imply that agents in the control group have worse traits and/or respond less to career incentives even if they face the same incentives once hired.

Our analysis proceeds in three stages. First, we measure the effect of recruiting with career incentives on the inputs provided by the CHAs once hired—i.e., the quantity and quality of services they deliver. Second, we test whether recruiting CHAs with career incentives affect facility utilization, health practices and health outcomes in the areas where they operate. Third, we assess the extent to which the performance gap can be explained by selection on observables such as skills and personality traits.

Our experimental design is as follows. In control districts, recruitment materials make salient benefits to the community, thus highlighting the similarities between the new CHA positions and existing informal positions (e.g., village health workers, traditional birth attendants, barefoot doctors) that are common in these areas. In treated districts, recruitment materials make career possibilities salient by highlighting that CHAs are part of the Ministry of Health’s hierarchy and that this gives them access to a career path leading to higher-ranked positions such as nurse, clinical officer, and doctor. Importantly, since the program requires that CHAs must belong the commu-

nity they want to work in, treatment and control communities draw from their own separate pools, thus career incentives cannot draw in talent from control areas. Once hired, all agents move to the same training school where they are trained together for one year and given the same information about the job, including career benefits. Thus all CHAs have the same information and face the same incentives after being hired.

The first stage of the analysis follows the CHAs in the field over the course of 18 months to measure their performance in delivering health services. At this stage, all CHAs are similarly aware of career benefits, and thus performance differences, if any exist, cannot be driven by differences in incentives on the job. Importantly, attrition between recruitment and deployment is trivial, thus allaying the concern that CHAs might drop out after finding out that career benefits exceed those advertised in the control group.

The CHAs' main task is to visit households to conduct environmental inspections, counsel on women's and children's health, and refer them to the health post as needed (e.g. for routine checks for children and pregnant women, or for giving birth). Our core performance measure is the number of household visits completed over the study period. In addition to visits, CHAs are supposed to devote one day per week to work at the health post and to organize community meetings. We measure the numbers of patients seen and meetings organized.

We find that CHAs recruited with career incentives conduct 29% more household visits and organize over twice as many community meetings, while the difference in the number of patients seen at the health post is also positive but not precisely estimated. Supplementary evidence suggests that the difference is not due to measurement error and is not compensated by improvements on other dimensions, such as the duration of visits, targeting of women and children, or visiting hard-to-reach households.

The second stage of the empirical analysis tests whether the selection induced by career incentives affects outcomes that are related to the services delivered by the CHAs, but not directly chosen by them. Given that CHAs are supposed to focus on maternal and child health, we use administrative data on government facilities to test whether our treatment affects women's and children's use of health services (as it should if CHAs are doing their job effectively). Difference-in-difference estimates based on the comparison of treated and control areas before and after CHAs started working reveal that treatment increased the number of women giving birth at the health center by 30%, and the number of children under 5 undergoing health checks by 24%, being weighed by 22% and receiving immunization against polio by 20%. Next, we use survey data from our own survey of 738 households in the 47 districts served by the CHAs to measure treatment effects on health practices and outcomes. We find consistent increases in a number of health practices: breastfeeding and proper stool disposal increase by 5pp and 12pp, deworming treatments by 15% and the share of children on track with their immunization schedule by 5pp (relative to a control mean of 5%). These changes are matched by changes in outcomes as the share of under 5s who are underweight falls by 5pp, or 25% of the mean in control areas.

Finally, we assess the extent to which the observed performance gap can be explained by selection on observables, which informs whether the effect of incentives can be mimicked by a change in the eligibility criteria. We measure standard determinants of performance such as skills, as well as pro-social preferences that might be relevant given the nature of the job, and might be crowded out by our treatment. We find that career incentives attract different types: CHAs in the treatment group have better skills (as measured by test scores during the training program), stronger

career ambitions (as measured by psychometric scales),<sup>4</sup> and are more likely to choose career over community as the main reason to do the job, although only a handful do so. In line with this, CHAs in the two groups score similarly on psychometric scales that measure pro-sociality and donate similar amounts in a contextualized dictator game.<sup>5</sup>

We find that several of these characteristics correlate with performance: most notably, CHAs with higher test scores perform better, while those that put career over community perform worse, which supports the idea that pro-sociality improves public service delivery. Controlling for observables, however, only explains a small part of the performance gap, suggesting that career incentives attract agents whose unobservable traits make them more productive or more responsive to the incentives themselves. The finding that the selection effect acts through unobservable traits echoes the importance of unobservables in other settings where agents self-select such as in applying for welfare programs (Alatas et al., 2015) or purchasing health products (Ashraf et al., 2010). In those settings, like in ours, self-selection cannot be mimicked by targeting on observable traits.

Taken together, the evidence discussed in this paper highlights the importance of incentive design at the recruitment stage to attract strong performers that cannot be identified on observables alone. That differences in performance are matched by dif-

---

<sup>4</sup>To measure preferences, we draw on the literature in organizational behavior that correlates individual psychometric traits with job attributes and performance (Amabile et al., 1994; Wrzesniewski et al., 1997; Barrick et al., 2001; Wageman, 2001; Barrick et al., 2002; Grant, 2008; Gebauer & Lowman, 2008; Duckworth et al., 2007).

<sup>5</sup>As we have data on all applicants who were interviewed, we can further decompose the selection effect into self-selection; namely, the treatment attracts different types, and employer selection; namely, recruitment panels choose candidates with different characteristics. We show that the treatment attracts a different applicant pool while recruitment panels put the same weights on the same traits. Observed differences are thus mostly driven by self-selection.

ferences in outcomes further strengthens the case for focusing on recruitment strategy as a tool to improve performance in organizations, and underscores the impact such differences in performance can make.

Our paper contributes evidence on the selection effects of incentives to the literature that studies the effects of incentives on performance (see Lazear & Oyer, 2012, Oyer & Schaefer, 2011 for recent surveys). In particular, our findings complement the literature that evaluates the effect of introducing material incentives for agents who deliver public services, like teachers (Muralidharan & Sundararaman, 2011; Duflo et al., 2012; Miller et al., 2012), by showing that material incentives affect who sorts into these jobs in the first place, and that this selection affects performance.

Our findings on the effect of career incentives on applicant traits are in line with Dal Bó et al. (2013), who exploit two randomized wage offers for a civil servant job in Mexico and show that higher wages attract more qualified applicants without displacing pro-social preferences.<sup>6</sup> In contrast, Deserranno, 2014 shows that offering higher pay to community health promoters, whose job consists of both commercial sales and public health information delivery, displaces pro-social preferences by signalling that sales are the more important component. This mechanism is muted in our setting as the job is made of similar tasks, and the treatment does not provide information on their relative importance.

The rest of the paper is organized as follows. Section 2 describes the context and research design, Section 3 evaluates the treatment effect on performance in delivering health services. Section 4 evaluates the treatment effect on health behaviors and outcomes using administrative and survey data. Section 5 assess the extent to which

---

<sup>6</sup>That higher wages attract better-quality applicants is also found in a related literature on wages and job queues in the private sector (Holzer et al., 1991; Marinescu & Wolthoff, 2013) and on the effect of wages on the selection of politicians (Ferraz & Finan, 2011; Gagliarducci & Nannicini, 2011).

the observed performance gap can be explained by selection on observables. Section 6 concludes with a discussion of external validity, welfare implications and general equilibrium effects relevant for program scale-up.

## 1.2 Context and research design

### 1.2.1 Context and data

In 2010, the Government of the Republic of Zambia (GRZ) launched a program to create a new civil service cadre called the Community Health Assistant (CHA) to address staff shortages in rural areas.<sup>7</sup> GRZ sought to formalize and professionalize a position similar to community-based lay health workers (e.g., village health workers, traditional birth attendants, barefoot doctors) that are common in rural Zambia; these informal positions had been the primary providers of health services to rural populations.

CHAs are recruited from the communities where they will eventually work, trained together for one year in a central location and posted back to their communities after that. CHAs are expected to devote 80% of their time (4 out of 5 working days per week) to household visits. The visits' main goals are to provide advice on women's health—including family planning, pregnancy, and postpartum care—and child health, including nutrition and immunizations. In addition, CHAs are expected to inspect the household and provide advice on health-related practices such as safe water practices, household waste management, sanitation, hygiene and

<sup>7</sup>The goal of this program was to create an “adequately trained and motivated community-based health workforce, contributing towards improved service delivery [and] the attainment of the Millennium Development Goals (MDGs) and national health priorities” (Government of Zambia, 2010).



ventilation. During visits, CHAs are also tasked with providing basic care to any sick persons and referring them to the health post as needed. In the remaining time, CHAs are expected to assist staff at the health post (the first-level health facility in rural Zambia) by seeing patients, assisting with antenatal care, and maintaining the facility. They are also supposed to organize community meetings such as health education talks at the health post and in schools.

The CHA position confers career benefits because it is an entry point into the civil service from which agents can advance to higher-ranked and better paid cadres. Promotion into higher-ranked cadres within the Ministry of Health from the position of CHA requires additional training (for example, nursing or medical school). Being part of the civil service, CHAs are eligible for “in-service training,” meaning that they attend school as a serving officer and the government pays their tuition for all of their training.

In the program’s first year, GRZ sought to recruit, train, and deploy roughly 330 Community Health Assistants across seven of Zambia’s nine provinces.<sup>8</sup> Within these seven provinces, based on population density, GRZ chose the 48 most rural of the 58 constituent districts. Finally, across these 48 districts, GRZ identified 165 health posts that were deemed to be facing the most severe health worker shortages. From each community that surrounded each health post, the intention was to recruit two CHAs. We collaborated with GRZ at each stage of the recruitment process in all 48 districts as described below.

<sup>8</sup>The two other provinces, Lusaka and Copperbelt, were excluded by GRZ on grounds that they are the most urbanized of Zambia’s provinces.

## Stage 1: Job ads and application requirements

The recruitment and selection process occurred at the community (health post) level, with on-the-ground implementation coordinated by district health officials. In each community, paper advertisements for the job were posted in local public spaces, such as schools, churches, and the health post itself. District health officials were responsible for ensuring that the recruitment posters were posted. To ensure that the recruitment process was carried out in a uniform manner across the 165 communities, GRZ included detailed written instructions in the packets containing the recruitment materials (posters, applications, etc.) that were distributed to district health officials (see Appendix A.3).

The recruitment poster provided information on the position, the application requirements and process. The posters specified that applicants had to be Zambian nationals, aged 18-45 years, with a high school diploma and two “O-levels.”<sup>9,10</sup> All recruitment in the seven provinces occurred between August and October 2010. The

<sup>9</sup>Ordinary levels, or O-levels, are written subject exams administered to Zambian students in their final year of secondary school. They are the primary entry qualification into tertiary education. The Examinations Council of Zambia requires candidates to take a minimum of six O-level exams, including English and mathematics as compulsory subjects that have to be passed. There are currently 33 O-level subjects, such as biology, chemistry, civic education, woodworking, and accounting. Exam performance is rated on a nine-point scale, ranging from “distinction” to “unsatisfactory;” all but the lowest point-score are considered passing. The cost of taking O-level exams comprises a registration fee of roughly USD 16 and an exam fee of USD 10 per subject.

<sup>10</sup>The posters instructed eligible applicants to retrieve application forms from the health center associated with the health post. Applicants were to hand in their application forms, along with photocopies of their national registration cards and high school transcripts, to the health center within two weeks of the posters being posted. In line with the principle that CHAs should be members of the communities that they serve, the application form also required applicants to obtain the signed endorsement of a representative of the applicant’s “neighborhood health committee” (NHC), followed by the signed verification of the application by the health worker in charge of the associated health center. The NHC is a parastatal institution at the community level in rural Zambia. It is comprised of elected volunteer community representatives, whose collective responsibility is to coordinate community health efforts, such as immunization campaigns and village meetings about common health issues.

recruitment drive yielded 2,457 applications, an average of 7.4 applicants for each position. Both the total number of applicants and their distribution across health posts is similar in the two treatment groups: the treatment poster attracts 1,232 applicants in total and an average of 7.2 per position, while the control poster attracts 1,225 applicants in total and an average of 8.0 per position.

## **Stage 2: Interviews and selection by panels**

Once the application window closed, all completed application forms were taken to the district Ministry of Health office. There, district health officials screened applications to ensure that eligibility requirements were met. No discretion was given at this stage; applicants who did not meet the objective criteria were rejected, and those who did were invited for interviews. Overall, 1,804 (73.4%) applicants passed the initial screening and were invited for interviews; of these 1,585 (87.9%) reported on their interview day and were interviewed; of these, 48% came from the career incentives treatment and 52% from the control group. District officials were in charge of organizing interview panels at the health post level.<sup>11</sup> GRZ explicitly stated a preference for women and for those who had previously worked as community health workers, but the ultimate choice was left to the panels.<sup>12</sup>

---

<sup>11</sup>Each selection panel had five members: the district health official, a representative from the health post's associated health center, and three members of the local neighborhood health committee. These committees vary in size, but they typically have more than 10 members.

<sup>12</sup>In addition to submitting panel-wide nominations, individual panel members were instructed to rank their top five preferred candidates independently and, to this end, were given ranking sheets to be completed privately. Specifically, the ranking sheet instructions stated: "This ranking exercise should occur BEFORE panel members formally deliberate and discuss the candidates. Note that the ranking sheets are private and individual. Each panel member should fill out the ranking sheet confidentially so as to encourage the most honest responses. This step must be completed before the panel discussion."

### **Stage 3: Final selection, training, and deployment**

Out of the 1,585 interviewees, for the 165 health posts, the panels nominated 334 applicants as “top 2” candidates and 413 as reserves. The nominations were reviewed centrally by GRZ, and 334 final candidates were invited to join a yearlong CHA training.

Of these, 314 applicants accepted the invitation and, in June 2011, moved to the newly built training school in Ndola, Zambia’s second-largest city. All applicants lived on site and were trained together for one year, during which treatment and control CHAs received the same information on the job, including the same information on career possibilities. Of the applicants who joined the program, 307 graduated and started working as CHAs in August 2012. All CHAs were deployed to their communities of origin.

#### **1.2.2 Experimental design**

The experiment aims to identify the effect of career incentives performance through selection. We use the recruitment posters described above and the information materials distributed to health officers to experimentally vary the salience of career incentives at the recruitment stage. All applicants are then given the same information on career possibilities once recruited.

Since recruitment for the CHA position was organized by district officials, we randomized treatment at the district level in order to maximize compliance with the experimental assignment, evenly splitting the 48 districts into two groups. This implies that each district official is only exposed to one treatment and is unaware of the other. As district officials are the main source of information for aspiring CHAs, randomization at the district level minimizes the risk of contamination. Randomization

at the district level also mitigates the risk of informational spillovers between communities, as the distance between health posts in different districts is considerably larger. Random assignment of the 48 districts is stratified by province and average district-level educational attainment.<sup>13</sup> To ensure compliance with the randomization protocol, we worked closely with GRZ to standardize the information given to the district officials to organize the recruitment process.<sup>14</sup>

The recruitment posters are shown in Figures 1.1 and 1.2. The treatment poster makes career incentives salient. To do so it lists, as the main benefit, the opportunity to ascend the civil-service career ladder to higher and better-paid positions, which are illustrated and enumerated in the poster—e.g., environmental health technician, nurse, clinical officer, and doctor. This incentive is summarized in a bold caption stating, “Become a community health worker to gain skills and boost your career!” In this setting, the pay gradient associated with career advancement is steep, as the starting monthly wage is USD 290 for CHAs, USD 530 for entry-level nurses, USD 615 for environmental health technicians, and USD 1,625 for resident doctors.<sup>15</sup>

---

<sup>13</sup>We stratify by the proportion of adults in the district who have a high school diploma, as reported in the most recent World Bank Living Conditions Measurement Survey, conducted four years prior in 2006. We sort districts by province and, within each province, by high school graduation rate. Within each sorted, province-specific list of districts, we take each successive pair of districts and randomly assign one district in the pair to the career incentives treatment and the other to the control group. For provinces with an odd number of districts, we pool the final unpaired districts across provinces, sort by educational attainment, and randomize these districts in the same pair-wise manner.

<sup>14</sup>District officials are given a packet containing 10 recruitment posters and 40 application forms for each health post and are asked to physically distribute each packet to the respective health center and, from there, to ensure that recruitment posters are posted, application forms are made available, and so forth. The packets are sealed and labeled according to the health post and health center for which it should be used. GRZ provides fuel allowances to the district officials to enable the districts to follow through on the protocol. We conduct a series of follow-up calls over several weeks to the district point-persons to ensure that the recruitment process is conducted as planned.

<sup>15</sup>At the time of the launch of the recruitment process in September 2010, GRZ had not yet determined how much the CHAs would be formally remunerated. Accordingly, the posters did not

Figure 1.1: Recruitment poster: Career incentives treatment

**REPUBLIC OF ZAMBIA  
MINISTRY OF HEALTH**



DESIGNATED HEALTH CENTRE:	FOR POSTING AT:

**TRAINING OPPORTUNITY**

**ONE-YEAR COURSE IN COMMUNITY HEALTH**

The Ministry of Health of the Republic of Zambia is launching a new national Community Health Worker (CHW) strategy and invites applicants to participate in the inaugural training of community health workers.

The training will begin on **30<sup>th</sup> August 2010** and will be held at the Provincial level for selected applicants. All participation costs, including transportation, meals and accommodation will be covered by the Ministry of Health.

**BENEFITS:**

- Become a highly trained member of Zambia's health care system
- Interact with experts in medical fields
- Access future career opportunities including:
  - Clinical Officer
  - Nurse
  - Environmental Health Technologist

**QUALIFICATIONS:**

- Zambian National
- Grade 12 completed with two "O" levels
- Age 18-45 years
- Endorsed by Neighborhood Health Committee within place of residence
- Preference will be given to women and those with previous experience as a CHW

**APPLICATION METHOD:**

Submit to the **DESIGNATED HEALTH CENTRE** indicated above:

- Completed application form with necessary endorsements. If no blank forms are attached to this notice, kindly obtain a blank one at the nearest health centre.
- Photocopy of school certificate documenting completion of Grade 12 and two "O" levels.
- Photocopy of Zambian national registration card.

**For more information:** Contact the designated health centre indicated above.

**CLOSING DATE: 30<sup>th</sup> JULY 2010.**  
Only shortlisted candidates will be contacted for interview.

Figure 1.2: Recruitment poster: control group

**REPUBLIC OF ZAMBIA  
MINISTRY OF HEALTH**



DESIGNATED HEALTH CENTRE:	FOR POSTING AT:

**TRAINING OPPORTUNITY**

**ONE-YEAR COURSE IN COMMUNITY HEALTH**

The Ministry of Health of the Republic of Zambia is launching a new national Community Health Worker (CHW) strategy and invites applicants to participate in the inaugural training of community health workers.

The training will begin on **30<sup>th</sup> August 2010** and will be held at the Provincial level for selected applicants. All participation costs, including transportation, meals and accommodation will be covered by the Ministry of Health.

**BENEFITS:**

- Learn about the most important health issues in your community
- Gain the skills you need to prevent illness and promote health for your family and neighbors
- Work closely with your local health post and health centre
- Be a respected leader in your community

**QUALIFICATIONS:**

- Zambian National
- Grade 12 completed with two "O" levels
- Age 18-45 years
- Endorsed by Neighborhood Health Committee within place of residence
- Preference will be given to women and those with previous experience as a CHW

**APPLICATION METHOD:**

Submit to the **DESIGNATED HEALTH CENTRE** indicated above:

- Completed application form with necessary endorsements. If no blank forms are attached to this notice, kindly obtain a blank one at the nearest health centre.
- Photocopy of school certificate documenting completion of Grade 12 and two "O" levels.
- Photocopy of Zambian national registration card.

**For more information:** Contact the designated health centre indicated above.

**CLOSING DATE: 30<sup>th</sup> JULY 2010.**  
Only shortlisted candidates will be contacted for interview.

Importantly, since there are shortages of health staff at every level, advancing to higher cadres does not require leaving the community.

The control poster, in contrast, lists as the main benefit the opportunity to contribute to one’s community, such as “[gaining] the skills you need to prevent illness and promote health for your family and neighbors” and “[being] a respected leader in your community.” This incentive is summarized in a caption stating, “Want to serve your community? Become a community health worker!” Potential applicants exposed to the control poster are thus presented with a description akin to the informal community health workers that are common in these areas, a position they would be familiar with.<sup>16</sup>

As part of the treatment, we include a basic written script that the district officials are invited to use to orient health centers and neighborhood health committees on the CHA program and recruitment process. In the career incentives treatment, the script describes the new program as follows: “This is an opportunity for qualified Zambians to obtain employment and to advance their health careers. Opportunities for training to advance to positions such as Nurse and Clinical Officer may be available in the future.” In contrast, in the control group, the script states, “This is an opportunity for local community members to become trained and serve the health needs of their community.” (see Section A.3).

display any information about compensation. Although the CHA wage was unknown to applicants at the time of application (indeed, unknown even to GRZ), applicants would likely have been able to infer an approximate wage, or at least an ordinal wage ranking, based on the “community health” job description and the relatively minimal educational qualifications required, both of which would intuitively place the job below facility-based positions in compensation. In Section 1.2.3, we present evidence against the hypothesis that wage perceptions may have differed by treatment.

<sup>16</sup>When the recruitment process was launched, the position was called “Community Health Worker” or “CHW” in both treatment and control areas. It was later renamed “Community Health Assistant” everywhere to avoid confusion with informal community health workers.



Once recruited, all CHAs lived and were trained together for one year during which they received the same information about job characteristics. Most importantly for our identification strategy all of them were told the benefits they were entitled to as civil servants and the career opportunities in the Ministry. As treatment and control CHAs face the same incentives once hired, performance differences, if any, are due to selection.

### **1.2.3 Experimental checks**

To provide evidence on whether the applicants' reported motivation for applying in treatment and control areas matches the treatment, we survey CHAs when they arrive at the training school. This timing is ideal because control CHAs have not been told about career incentives yet, and at the same time both treatment and control CHAs have already been selected, so they have no incentive to answer strategically. To elicit information about their motives to apply for the position, we give each CHA a bag of 50 beans and ask her to allocate them to different cards describing potential benefits of the job in proportion to the weight they gave to each benefit when applying. This method has two desirable features: first, it forces respondents to take into account the trade-off between different motives, namely that giving more weight to one motive necessarily implies that other motives will be given less weight; second, it allows us to test whether the treatment affected other motives besides career advancement and community service.

The answers tabulated in Table A.1 show that the reported motivations match the treatment and control posters well. The weight on career benefits is significantly higher in the career treatment (16.5% vs. 12.0%,  $p=.002$ ) while the weight given to “service to the community” and “earn respect and status in the community” are both

lower in the treatment group (39.6% vs. 43.2%,  $p=.050$  and 3.7% vs. 5.7%,  $p=.048$ , respectively). Two further points are of note. First, “service to the community” is the main reason to apply in both groups, suggesting that pro-social preferences might be equally strong in both groups, an issue to which we return in Section 1.5.1. Second, all other motivations to apply are balanced across groups, suggesting that the poster did not convey different expectations about pay or the nature of the job. To investigate this further, we ask CHAs where they expect to work in 5-10 years’ time. Over 90% of them expect to be with the Ministry, suggesting that the treatment and control posters do not convey different expectations about tenure.

#### **1.2.4 Context descriptives and balance**

Tables 1.1 and 1.2 describe three sets of variables that can affect the supply of CHAs, the demand for their services, and their working conditions. For each variable, the tables report the means and standard deviations in treatment and control, as well as the p-value of the test of means equality, with standard errors clustered at the level of randomization, the district. Tables 1.1 and 1.2 show that the randomization yielded a balanced sample as all p-values of the test of equality are above .05. As treatment and control means are very close throughout, we report values in the treatment group in what follows in this section.

Panel A reports statistics on the eligible population drawn from the 2010 Census, which shows that the eligibles—namely, 18-45 year-old Zambian citizens with at least Grade 12 education—account for 4.4% of the district population, and that among them 37% are female. A large fraction (13%) are unemployed and a further 7.6% are full-time housewives. The employed (63.1% of the total) are equally split between self-employment/unpaid labor in family business and wage employment. Among the self-

Table 1.1: Eligible population by treatment (randomization balance)

	treatment	control	p-value of the difference	
<b>A. Characteristics of the eligible population</b>				
<b>I. Eligible candidates</b>				
Share of eligibles in the district (18-45 year olds with grade 12 or above)	.044 (.205)	.043 (.203)	.917	
Share of women among the eligibles	.371 (.483)	.391 (.488)	.241	
Eligibles' average years of education	12.55 (.827)	12.55 (.829)	.929	
<b>II. Main activity of eligible candidates during the past 12 months</b>				
Unemployed	.133 (.340)	.125 (.331)	.640	
Housework/homemaking	.076 (.266)	.067 (.251)	.273	
Fulltime student	.086 (.280)	.087 (.282)	.860	
Self-employed or unpaid laborer in family business	.284 (.451)	.304 (.460)	.557	
	Farming	.170 (.376)	.173 (.378)	.938
Employees	.347 (.476)	.337 (.472)	.581	
	Teachers	.132 (.339)	.158 (.365)	.134
	Health workers	.023 (.149)	.025 (.156)	.615
	Low skill occupations	.133 (.341)	.099 (.298)	.127

Notes: Columns 1 and 2 show means and standard deviations in parentheses, Column 3 reports the p-value of the test of equality of means based on standard errors clustered at the district level. All variables are drawn from the 2010 Census (10% PUMS sample). Activities codes follow the ILO ISCO88 convention. Low skill occupations include workers engaged in services, sales, agriculture, crafts, manufacturing.

Table 1.2: Area and population characteristics by treatment (randomization balance)

	treatment	control	p-value of the difference
<b>B. Catchment area characteristics</b>			
Number of staff in health post	1.49 (1.09)	1.36 (1.17)	.559
Geographical distribution of households in catchment area:			
<i>Most people live in their farms, none in villages</i>	.082 (.276)	.091 (.289)	.848
<i>Some people live in farms, some in small villages (5-10hh)</i>	.529 (.502)	.532 (.502)	.855
<i>Most people live in medium/large villages (more than 10hh), a few on their farms</i>	.388 (.490)	.364 (.484)	.749
Poor cell network coverage	.082 (.277)	.065 (.248)	.675
<b>C. Target population characteristics</b>			
District population density (persons/km <sup>2</sup> )*	13.58 (8.88)	14.08 (9.92)	.854
Share of district population under 5*	.187 (.390)	.187 (.390)	.915
Average years of education of district residents*	4.20 (3.83)	4.20 (3.82)	.993
Number of assets owned by average HH in district*	12.67 (4.58)	12.76 (4.46)	.741
Main type of toilet: Pit latrine or better	.718 (.449)	.667 (.471)	.494
Household water supply: Protected borehole or better	.361 (.480)	.416 (.492)	.248

**Notes:** Columns 1 and 2 show means and standard deviations in parentheses; Column 3 reports the p-value of the test of equality of means based on standard errors clustered at the district level. Number of staff in health post is the total number of nurses, environmental health technicians, and clinical officers assigned to the health post as reported by district officials surveyed by phone. Information on the geographical distribution of HHs was obtained from a survey of the deployed CHAs before deployment. CHAs were shown stylized maps accompanied by the description above and asked to choose the one that most closely resembled the catchment area of their health post. Questions were asked to each CHA individually so that two CHAs from the same health post could give different answers. For the 5 out of 161 cases in which the two CHAs gave different answers, we use the information provided by supervisors to break the tie. To measure cell network coverage we attempt to call all CHAs after deployment. We make daily calls for 118 consecutive days. The health post is classified as having poor coverage if we do not manage to reach either of its two CHAs during this period. Variables with \* are drawn from the 2010 Census (10% PUMS sample). Variables with \*\* are drawn from the 2010 Living Conditions Monitoring Survey (LCMS), which covers 20,000 HHs and is representative at the district level. Main type of toilet: Pit latrine or better equals 1 if the surveyed household uses a pit latrine, ventilated improved pit (VIP), or flush toilet, and 0 if bucket, other, or no toilet. Household water supply: Protected borehole or better equals 1 if the water supply comes from a protected borehole or well, communal tap, or other piped water systems, and 0 if it comes from an unprotected well or borehole, river/dam/stream, rain water tank, other tap, water kiosk, water vendor, mineral/bottled water, or other. Number of assets owned is the number of durable goods and livestock owned by the household.

employed/unpaid laborers the most common occupation is farming, which accounts for 17% of the eligibles. Among those who work for a wage, the most common occupations are teachers (13.2%) and low-skilled occupations (13.3%), which include services, sales, agriculture, crafts, and manufacturing. Only a small minority (2.3%) are already employed in the health sector. Taken together, the evidence suggests that, despite their educational achievements, the majority (65.3%) of the eligibles are not in stable wage employment. This indicates that the CHA program can draw talent from these areas without crowding out other skilled occupations.

Panel B illustrates the characteristics of the catchment areas. These variables are drawn from surveys administered to district officials and the CHAs themselves. Three points are of note. First, health posts are poorly staffed in both the treatment and control groups; the average number of staff (not including the CHA) is 1.5. Given that the aim is to assign two CHAs to each health post, the program more than doubles the number of health staff in these communities. Second, the areas vary in the extent to which households live on their farms or in villages, but the frequency of either type is similar in the treatment and control groups. This is relevant as travel times between households depend on population density and are higher when households are scattered over a large area, as opposed to being concentrated in a village. Third, over 90% of the catchment areas in both groups have at least some cell network coverage, which is relevant for our analysis, as some performance measures are collected via SMS messages.

Panel C illustrates the characteristics of the target population that are relevant for the demand for CHA services. First, population density is fairly low in both groups, which implies that CHAs have to travel long distances between households. This also implies that the ability to plan and efficiently implement visits is likely to play a key role in determining the number of households reached. Second, children under 5,

who (together with pregnant women) are the main targets of CHAs, account for 19% of the population. Third, the educational achievement of the average resident is 4.2 years, well below the average for those eligible for the CHA position (12.6 years, panel A). Fourth, Panel C shows that access to latrines and—most noticeably—protected water supply is limited in these areas. Lack of latrines and protected water supply favor the spread of waterborne infections, to which pregnant women and children are particularly vulnerable and, through this, the demand for CHAs’ services.

## **1.3 The effect of career incentives on performance via selection**

### **1.3.1 Measuring performance in service delivery**

The CHAs’ main task, to which they are required to devote 80% of their time, or 4 out of 5 days per week, is to visit households. Our performance analysis focuses on the number of visits completed over the course of 18 months, from August 2012 (when CHAs started work) until January 2014. The number of household visits is akin to an attendance measure for teachers or nurses: CHAs are supposed to work in people’s houses, and we measure how often they are there. Naturally, differences in the number of visits can be compensated by behavior on other dimensions; we discuss this possibility after establishing the main results in Section 1.3.3.

Our primary measure of household visits is built by aggregating information on each visit from individual receipts. All CHAs are required to carry receipt books and issue each household a receipt for each visit, which the households are asked to sign. CHAs are required to keep the book with the copies of the receipts to send to GRZ when completed. They are also required to send all information on these

receipts—consisting of the date, time, and duration of the visit, as well as the client’s phone number—via text message to the Ministry of Health. These text messages are collected in a central data-processing facility, which we manage. CHAs know that 5% of these visits are audited.

Since visits are measured by aggregating text messages sent by the CHAs themselves, identification can be compromised by the presence of measurement error that is correlated with treatment. For instance, CHAs in the career treatment might put more effort in reporting visits via text messages or might report visits that never took place, leading to a positive bias in the estimated treatment effect. Outright cheating is made difficult by the fact that CHAs would need to falsify the household signature on the official receipt to report a visit that did not happen. While the SMS submissions carry no signature, CHAs are required to send their household visit receipt books containing carbon copies of the receipts to the Ministry of Health for cross-checking. Fabricating receipts thus entails a potentially high cost for no direct benefit. Nevertheless, the estimated treatment effect might be upward biased because of differential effort in reporting.

We validate our visits measure by comparing it to administrative data and households’ own reports of CHA activity. The administrative data is drawn from the Health Management and Information System (HMIS), which is the Ministry of Health’s system for reporting, collecting, and aggregating routine health services data at government facilities. These are reported at the end of each month and sent electronically to the Ministry via a mobile platform, jointly by the two CHAs and the other staff working in each health post. While HMIS visit data are also collected by the CHAs themselves, the effort required is considerably lower since HMIS reports are compiled monthly rather than on every visit, and cheating is more difficult as the reports are compiled jointly by the two CHAs and the health post staff. As HMIS data are

only available aggregated at the health post level—i.e., summed over the two CHAs in each health post—we regress these on our visit measure, also aggregated at the health post level. Columns 1 and 2 in Table A.2 shows that the two measures are strongly correlated ( $r=.766$ ) and that the correlation is the same in treatment and control, which contradicts the differential reporting hypothesis.

The households’ reports are collected via a survey that we administered to 16 randomly chosen households in each of 47 randomly selected communities chosen from the set of 161 communities where CHAs operate, stratified by district.<sup>17</sup> For each CHA, we ask respondents whether they know the CHA (97% do), whether they have ever been visited (43% of them have), and their level of satisfaction with each CHA. Columns 3-6 show a precisely estimated correlation between our visit measure and the probability that a household reports a visit, as well as their level of satisfaction with the CHA’s performance. Again, there is no significant difference between the treatment and control groups, casting doubt on the relevance of differential reporting.

Taken together, the findings in Table A.2 validate our visits measure. Ultimately, however, we will not be able to detect a treatment effect on households’ health outputs in Section 1.4 if measured differences in visits capture differences in reporting rather than in actual visits .

### 1.3.2 Treatment effect on household visits

Table 1.3 reports the estimates of

$$v_{ihdp} = \alpha + \beta C_{id} + Z_h \gamma + \delta E_d + \rho_p + \epsilon_{ihdp} \quad (1.3.1)$$

---

<sup>17</sup>As CHAs are supposed to focus on mother and child’s health we interview the wife of the head (if this is male) or the head herself (if female).



where  $v_{ihdp}$  is the number of visits completed by CHA  $i$  in catchment area  $h$  district  $d$  and province  $p$ ,  $C_{id}$  equals 1 if agent  $i$  is recruited and operates in a district assigned to the career incentives treatment.  $Z_h$  is a vector of area characteristics, which includes the number of staff at the health post, cell network coverage, and the distribution of households between farms and villages described in Table 1.2. We control for the stratification variables, district-level high school graduation rate  $E_d$  and provinces indicators  $\rho_p$  throughout. Standard errors are clustered at the level of randomization—the district.

The coefficient of interest is  $\beta$ , which measures the effect of making career incentives salient at the recruitment stage on the number of visits completed over 18 months. Given that all CHAs are given the same information on career incentives during the yearlong training,  $\beta$  captures the effect of career incentives on performance through selection. Note that selection can affect performance by increasing productivity for a given level of effort or by increasing the marginal return to effort. An example of the former is talent for logistics: for the same amount of effort, a more talented CHA plans better and reaches more households in the same amount of time. An example of the latter is the utility weight put on career advancement: CHAs who value career more draw a higher marginal benefit from a given unit of effort and therefore exert more effort.

The causal effect of career incentives on performance can be identified under the assumptions that (i)  $C_{id}$  is orthogonal to  $\epsilon_{ihdp}$ , (ii) there are no spillovers between the two groups, and (iii) the salience policy itself does not affect behavior directly. Orthogonality is obtained via random assignment, but measurement error in visits correlated to  $C_{id}$  can bias the estimates. We return to this in Section 1.3.3 below.

Spillovers via movements of CHAs between treatment and control areas are ruled out by the program requirement that CHAs must have been residing in the com-

Table 1.3: The effect of career incentives on the number of visits

dependent variable	Household visits			
	source	SMS receipts	SMS receipts	SMS receipts
	time horizon	months 1-18	months 1-6	months 7-12
	level	CHA	CHA	CHA
	(1)	(2)	(3)	(4)
Career incentives	93.95**	33.93**	29.56**	30.46**
	(37.19)	(15.97)	(13.49)	(12.92)
Area characteristics	Yes	Yes	Yes	Yes
Mean of dependent variable in control	318.6	167.1	92.1	59.8
Adjusted R-squared	0.112	0.115	0.064	0.105
N	307	307	307	307

Notes: OLS Estimates, standard errors clustered at the district level. The dependent variable is total number of household visited over the relevant time horizon. SMS receipts are sent by individual CHAs to MOH for each visit. All regressions include the stratification variables (province dummies and share of high school graduates in the district). Area characteristics include: number of staff in the health post, geographical distribution of households in the catchment area, and an indicator variable that equals 1 if the CHA reports to have good cell network coverage most of the time or all the time.

munity they want to work in prior to applying. This implies that career incentives cannot draw in talent from control areas as treatment and control communities draw from their own separate pools. Spillovers of information, caused, e.g., by potential applicants in control seeing the treatment poster, would introduce a downward bias because they would reduce the information differences between treatment and control. Information spillovers are minimized by design, as recruitment messages were randomized at the district level, which, given the travel distance between rural communities in different districts, makes it very unlikely that applicants in one group might have seen the poster assigned to the other group. Importantly, information cannot spillover through the district officials that implement the program or through the recruitment panels, as these are only exposed to one treatment only.

Finally, in Section 1.3.3 we present evidence to allay the concern that  $\beta$  captures the effect of the salience policy rather than career incentives themselves.

Column 1 reveals a large and precisely estimated effect of career incentives on household visits: CHAs recruited by making career incentives salient do 29% more visits over the course of 18 months. The magnitude of the difference is economically meaningful: if each of the 147 CHAs in the social treatment had done as many visits as their counterparts in the career treatment, 13,818 more households would have been visited over the 18-month period. Given that for most of these households CHAs are the only providers of health services, the difference between treatments is likely to have implications for health outputs in these communities. We return to this issue in Section 1.4.

Figure 1.3 provides evidence of treatment effects on the distribution of household visits. Both the comparison of kernel density estimates and quantile treatment effect estimates reveal that the difference between the two treatments is driven by a group of strong performers in the treatment group. The effect of career incentives is positive

throughout but flat until the 40th percentile and increasing thereafter. The quantile estimates indicate that career incentives lead to better performance by attracting a group of individuals who perform much better than the average CHA.

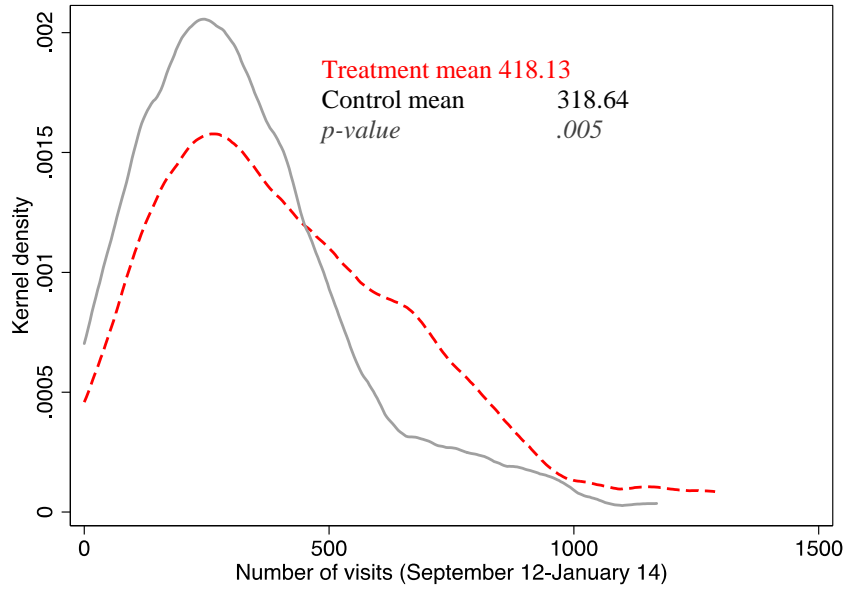
### **1.3.3 Identification: the effect of salience**

The experimental design allows us to identify the effect of career incentives on performance through selection if the salience policy itself does not directly affect the applicants' utility once the real career benefits are known by both treatment and control CHAs. Since career benefits are greater than or equal to the values agents knew at the application stage, we need to rule out behavioral biases that make agents value a given benefit differently if its value exceeds their expectation. This assumption might fail for two reasons. First, if agents are made worse off by discovering that the actual value of a given benefit is larger than the value advertised by the salience policy, agents for whom the participation constraint is met ex-ante but not ex-post would drop out once hired, and differences in performance among stayers would not be interpretable as the effect that career incentives have on performance through their effect on the applicant pool. Reassuringly, the drop-out rate at the relevant stage is minimal. Namely, 314 agents join training informed by the salience policy. They are then told about the actual benefits of the job at the start of the one-year training program. Contrary to the implication that some are made worse off by discovering that the actual value of a given benefit is larger than the value advertised by the salience policy, 98% of selected candidates stay on after discovering the actual benefits and complete the training program.

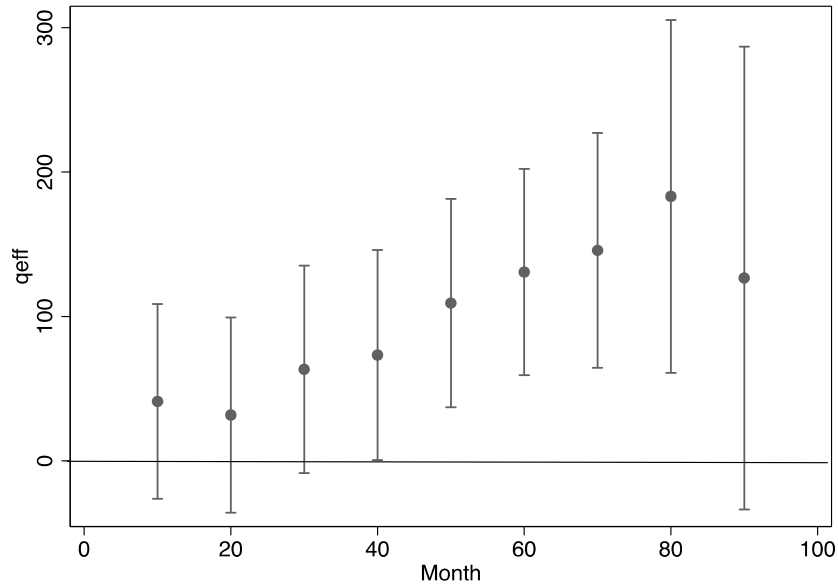
Second, if agents are made better off by discovering that the actual value of a given benefit is larger than the value advertised by the salience policy, they may react to

Figure 1.3: The effect of career incentives on performance

A. Kernel density estimates of visits



B. Quantile treatment effects



Notes: Total number of household visited, aggregated from individual SMS receipts sent by individual CHAs to MOH. Panel A plots kernel density estimates. Panel B reports quantile treatment effects using the same covariates as in Column 2, Table 6. Each point represents the treatment effect at the decile on the x-axis, each bar represents the 90% confidence interval. Confidence intervals are based on bootstrapped standard errors with 500 replication clustered at the district level.

the positive surprise by working harder. This would imply, for instance, that the effect of career incentives on effort would be stronger in the control group, to whom career benefits are revealed after being hired, than in the treatment group, who knew about career benefits all along. Given that treatment CHAs do more visits, the only way in which our estimates overstate the effect of career incentives is if the “surprise” effect is actually negative for agents in the control group (i.e., their effort response to finding out about career benefits is negative and larger -in absolute value- than what it would have been had they known the career benefits at the outset).

While we cannot measure the surprise effect directly, we can exploit the long time series of performance data to test whether the treatment effect changes with time in a manner that is consistent with there being a “surprise” effect. Specifically, if estimated differences between treatment and control are overstated due to the “surprise” effect, we expect treatment effects to shrink with time as the surprise wanes.

To test this implication, in columns 2-4, we divide the 18-month period into three semesters. We find that the estimated treatment effect is identical in the three sub-periods: in each semester, the average CHA recruited under the career salience policy does between 30 and 34 more visits. Since the number of visits falls over time, the percentage effect *increases* with time from 20% to 51%. This casts doubt on the interpretation that CHAs’ behavioral responses to differences between salience policy and actual incentives lead us to overstate the effect of career incentives on performance through selection.

### **1.3.4 Compensation mechanisms and work styles**

Table 1.4 investigates the hypothesis that CHAs in the control group take other actions that compensate for the lower number of visits. Column 1 tests whether

career incentives improve performance at the expense of retention—e.g., whether they attract individuals who leave with their newly acquired skills as soon as it is feasible to do so. In our context, the CHAs are bonded to their position for one year.<sup>18</sup> Thus, we measure retention by the number of CHAs who make at least one visit after the one-year commitment has elapsed. We find that, by this measure, 18% of CHAs drop out, though some of this may be due to a combination of malfunctioning phones and the rainy season (falling between months 15-18 in our analysis window) making travel to cell network-accessible areas difficult. This attrition rate is balanced across treatments. It is important to note that according to the Ministry’s rule, CHAs have to wait two years before applying for higher-ranked positions, such that none of those who left their positions did so for career progression. It is possible that career incentives will affect retention rates after the two-year mark. As we discuss in the Conclusion, the welfare implications of this effect (were it to materialize) are ambiguous.

Columns 2 and 3 investigate whether CHAs in the control group compensate by spending more time with each household or are better at reaching those they are supposed to target. The results show that CHAs in both groups devote the same time to a single visit, on average, and are equally likely to target their primary clients—women and children.

Columns 4 and 5 decompose the number of total visits into the number of unique households visited and the average number of visits per household to test whether CHAs in the career treatment do more visits because they cover a smaller number of easy-to-reach households. Contrary to this, columns 4 and 5 show that CHAs

<sup>18</sup>The CHAs were told that, if they quit before one year of service, they would be required to pay monthly wages for any months not worked (rather than simply relinquishing pay) to compensate the government for the free one-year training that they received.

Table 1.4: Compensation mechanisms

dependent variable	retention		visit duration		no of women and children visited per HH		no of unique HHs visited		no of visits per HH		community mobilization meetings		patients seen at health post		emergency calls		
	SMS receipts	CHA	SMS receipts	CHA	HMS records	HMS records	SMS receipts	CHA	SMS receipts	CHA	HMS records	HMS records	HMS records	health post	HMS records	CHA	
level	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	
Career incentives	0.0469 (0.0582)	0.265 (1.850)	0.0437 (0.0947)	36.35** (15.49)	0.488* (0.246)	17.06*** (5.220)	31.79 (260.4)	0.0469 (0.0582)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Area characteristics	0.796	33.9	2.07	179.4	1.817	20.32	1126.6	0.457	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of dependent variable in control	0.041	0.011	0.006	0.121	0.125	0.072	0.027	0.002	307	307	142	146	146	146	146	146	298
Adjusted R-squared																	
N																	

Notes: OLS Estimates, standard errors clustered at the district level. The dependent variable is total number of household visited over the relevant time horizon. SMS receipts are sent by individual CHAs to MOH for each visit. The Health Management and Information System (HMIS) is the Zambian Ministry of Health's system for reporting health services data at government facilities. The two CHAs are required to submit monthly reports that summarize their activities at the health post/community level. The number of observations varies because some health posts do not submit the reports; these are equally distributed between treatments. The time use survey was administered in May 2013 during a refresher training program. Emergency calls=1 if the CHA takes at least 1 out of hours call in a typical week. Retention=1 if CHA is active after 1 year. All regressions include the stratification variables (province dummies and share of high school graduates in the district). Area characteristics include: number of staff in the health post, geographical distribution of households in the catchment area, and an indicator variable that equals 1 if the CHA reports to have good cell network coverage most of the time or all the time.



in the career incentive treatment reach more households and make more follow-up visits. The point estimates indicate that just over one-third (36/94) of the total treatment effect is due to career CHAs visiting more households and two-thirds to them visiting the same household more than once. This is consistent with the two groups of CHAs having a similar number of households in their catchment area and visiting them at least once, but treatment CHAs doing more follow-up visits. Note that longitudinal follow-up with households is considered an integral part of the CHA job, in view of which Ministry of Health guidelines state CHAs should attempt to visit each household on a quarterly basis. Column 5 indicates that CHAs in both groups fall short of this target, suggesting that differences in performance are relevant to welfare.

The results in columns 4 and 5 also cast doubt on the hypothesis that observed differences are driven by measurement error, because it is equally costly to send SMSs for first or repeated visits, but differences are larger for the latter.

Besides household visits, CHAs are expected to assist staff at the health post by seeing patients, assisting with antenatal care, and maintaining the facility. They are also supposed to organize community meetings such as health education talks at the health post and in schools. Columns 6-7 investigate whether differences in household visits are compensated by differences in secondary tasks using HMIS data on the number of community meetings CHAs organize and the number of patients they attend to at the health post. The latter should be seen as a proxy of the quantity of services delivered by CHAs at the health post, as seeing patients is mostly a nurse's job. We find that CHAs recruited by making career incentives salient organize twice as many meetings over 18 months (43 vs. 22), and the difference is precisely estimated. The effect of career incentives on the number of patients CHAs see at the health post is also positive but small and not precisely estimated.

To provide further evidence on possible compensation mechanisms, we administer a time use survey that is meant to capture differences in work style. We surveyed CHAs in May 2013, nine months after they started working.<sup>19</sup> The survey asked CHAs to report the frequency of emergency visits typically done outside of working hours. The median CHA does one emergency call per week, and column 8 shows that this holds true for CHAs in both groups.

The time use survey is designed to collect information on hours worked and the time allocated to different activities. This allows us to assess whether the differences in performance documented above are due to differences in time allocation across tasks; namely, whether treatment CHAs do more visits because they devote more time to that task. To collect information on the latter, CHAs were given 50 beans and asked to allocate the beans in proportion to the time devoted to each activity within each task. Besides household visits, community meetings and time at the health post, we allow for two further activities: traveling and meeting with supervisors. For each activity, we calculate the share of time devoted to each activity by dividing the number of beans allocated to that activity by the total number of beans allocated to all activities. The share of time allocated to these five activities is .32, .22, .16, .22 and .09, respectively. We then estimate a system of equations for hours worked and share of time devoted to each task, omitting traveling. Table 1.5 reports our findings.

Column 1 shows that the average CHA reports working 43 hours per week in the typical week and there is no difference in reported working hours by treatment. This suggests that CHAs in the control group do not compensate for visiting fewer households by devoting more hours to other, possibly informal, tasks. It also provides

---

<sup>19</sup>To implement this survey we took advantage of a refresher course organized by GRZ in the CHA School in Ndola. Of the 307 CHAs, 298 (97%, equally split by treatment groups) came to training and took part in the survey.

Table 1.5: The effect of career incentives on time allocation

dependent variable	Hours worked		Share of time spent in:		
		HH visits	Health Post	Community meetings	Meeting with supervisor
	(1)	(2)	(3)	(4)	(5)
Career incentives	-.588 (1.19)	.007 (.014)	-.021* (.012)	.011 (.011)	-.001 (.008)
Area characteristics	yes	yes	yes	yes	yes
Mean of dependent variable in control	42.8	.312	.171	.213	.085
Adjusted R-squared	.071	.055	.081	.031	.063
N	298	298	298	298	298

Notes: SURE Estimates, standard errors clustered at the district level bootstrapped with 1500 replications. Data source is the Time Use Survey that was administered in May 2013 during a refresher training program. Hours worked is defined as the number of hours worked in a typical week as reported by the CHAs. To measure the "Share of time spent in," CHAs were given 50 beans and asked to allocate them on cards listing the different activities listed above plus travel. The cards were scattered on a table in no particular order. All regressions include the stratification variables (province dummies and share of high school graduates in the district). Area characteristics include: number of staff in the health post, geographical distribution of households in the catchment area, and an indicator variable that equals 1 if the CHA reports to have good cell network coverage most of the time or all the time.

further assurance that CHAs in the career treatment do not have differential incentives to overstate their contribution, as self-reported hours are unverifiable and hence easy to “game.”

Columns 2-5 show that CHAs in the two groups allocate their time in a similar manner; thus, observed performance differences are not driven by differences in time allocation. Two, possibly complementary, explanations are possible. First, treatment CHAs might work more effective hours—e.g., by taking shorter breaks over the 43 weekly hours. Second, treatment CHAs might be more efficient at their jobs. Household visits take place in remote, low-density areas: the median 78 square km area has 200 households, with an interquartile range of 130 to 360. It is thus rather time consuming to go from house to house, and this is compounded by the fact that roads are bad. In this setting, the ability to plan—e.g., by making appointments with specific households or collecting information as to whether members are likely to be home before setting out to visit them—is an important determinant of completing visits successfully. These effects might be strengthened by peer externalities because each CHA works alongside another CHA hired through the same treatment, thus CHAs in the treatment group are more likely to have a highly productive peer than CHAs in the control group. Peer effects might be driven by imitation, social comparison or a perception that the other CHA competes for the same promotion.

Finally, Appendix Table A.3 tests whether CHAs in the two groups allocate their time differently within each activity, namely whether they have different work “styles.” Panel A shows that CHAs in control devote more time to counseling, inspections, and visiting sick members, but, taken one-by-one, these differences are small and not precisely estimated. CHAs in the career incentives treatment devote 1.6% less time to filling in forms and receipts and submitting SMSs, but the difference is not precisely estimated at conventional levels. Because the quality of reports is the same,

this implies that career CHAs are more productive at this task. Panel B shows a similar pattern for time allocation during work at the health post: collecting data and filling in reports is an important component of the job, which takes 23% of the CHAs' time in the control group, but only 18% in the career treatment. As with household visits, there is no evidence that CHAs in the career treatment collect fewer data at the health post level or that these data are of worse quality. CHAs in the two groups are equally likely to submit HMIS reports in a given month, and these are equally accurate. Thus, the evidence suggests that CHAs in the career treatment are more productive, and this frees time for other tasks.

## **1.4 Impact on facility utilization and health practices and outcomes**

The CHA program leads to a substantial increase in the number of health staff operating in the communities where CHAs are deployed: the number of staff associated with the community health post increases on average from 1.5 to 3.5. Given the size of the increase and the magnitude of the treatment effect on household visits and community mobilization meetings, it is reasonable to expect treatment to affect health outcomes in these communities. CHAs can directly affect facility utilization and health practices by increasing both demand, e.g. by providing information and promoting behavioral changes, and supply, e.g. by helping cover staff shortages at the health post or delivering medical treatments to the households. In turn, improved facility utilization and practices should lead to better outcomes. To provide evidence on whether treatment affected facility utilization we use data from the Ministry's HMIS administrative records, to measure effects on health practices and outcomes

we survey households residing in the communities where CHAs operate. As the main remit of the CHA job is mother and child health, we focus on this throughout.

### 1.4.1 Impact on facility utilization

The Ministry’s HMIS administrative records are compiled by facilities’ senior staff and transmitted to MoH via an electronic platform. Two level of facilities serve these communities: health centers and health posts.<sup>20</sup> CHAs are supposed to encourage women to give birth at the closest health center and to bring in children for regular visits and immunizations at the closest facility (health center or health post). The importance of institutional deliveries in this context cannot be understated: Zambia’s maternal mortality rates are very high and health centers have the equipment and medical supplies that can prevent these deaths. Regular children’s visits ensure that conditions such as diarrhea are treated before they become dangerous. Immunizations protect children from potentially fatal illnesses.

To test whether the treatment affected facility utilization, we obtain information on institutional deliveries, children’s visits, and immunizations for the period January 2011-June 2014 and estimate the following difference-in-difference specification:

$$y_{hdpt} = \alpha + \beta C_{hd} + \gamma A_t + \delta C_{hd} * A_t + Z_h \theta + E_d \phi + \rho_p + \xi_{hdpt}$$

<sup>20</sup>Health facilities in Zambia are structured according to a population-based hierarchy. Health posts are the first-level health facility for most rural communities and provide basic medical care (no inpatient or surgical services). Health centers, which typically serve a population encompassing four to five health posts, provide both outpatient and inpatient services, including labor and delivery and minor surgical procedures. District hospitals in turn encompass several health center catchment areas and are primarily focused on inpatient care.

where  $y_{hdpt}$  is the outcome in health facility  $h$  in district  $d$  and province  $p$  at quarter  $t$ .<sup>21</sup>  $h$  represents the lowest level of government facility to which the CHAs can refer their patients. This is the health post if it is operational; if not, the closest health center. The only exception is childbirths that are always measured at the health center level, as that is where they are supposed to take place.  $C_{hd}=1$  if facility  $h$  is located in a district where CHAs were recruited via career incentives. We have data for 14 quarters, equally divided before and after the CHAs' arrival, and  $A_t=1$  after the CHAs' arrival (4th quarter of 2012). To minimize composition bias and to test for robustness to facility fixed effect models we restrict the sample to the facilities for which we have at least three observations before and after the CHAs' arrival.<sup>22</sup>  $Z_h$  is a vector of area characteristics, which includes the number of staff at the health post, cell network coverage, and the distribution of households between farms and villages described in Table 1.2. We control for the stratification variables, district-level high school graduation rate  $E_d$ , and provinces indicators  $\rho_p$  throughout. Standard errors are clustered at the level of randomization—the district.

The parameter of interest is  $\delta$ , the difference in differences between facilities in treatment and control districts before and after the CHA's arrival. Under the parallel trend assumption  $\delta$  captures the effect of career incentives for CHAs on these outputs.

Table 1.6 shows that indeed, career incentives improved clinic utilization outputs. In particular, the number of women giving birth at the health center increases by 30% relative to the mean in control areas at baseline. Regarding child health, the number of children under age five visited increases by 24%, the number of children

<sup>21</sup>HMIS data should be transmitted to MoH monthly, but in practice (due to poor connectivity), reports are missing for some months and the information added to the following month. We aggregate the data at the quarterly level to smooth out monthly fluctuations due to this.

<sup>22</sup>This restriction keeps 77% of the health posts and 70% of the health centers in the sample.

under five weighed increases by 22%, and the number of children under 12 months of age receiving polio vaccination increases by 20%. The effects on postnatal visits for women, BCG, and measles vaccinations are also positive and in the 8-22% magnitude range, but are not precisely estimated. The average standardized treatment effect over all outcomes is .277, significantly different from zero at the 1% level. Reassuringly, there are no significant differences between treatment and control areas in any of these outcomes before the CHAs' arrival: all the estimated  $\beta$  coefficients are small and not significantly different from zero.

To provide support to our identifying assumption, in Table A.7 (Panel A) we run a placebo test where we split the pre-CHA period in two halves and test whether outcomes improve in treatment areas over time even in the absence of CHAs. Reassuringly they do not. Finally, Table A.7 (Panel B) estimates (2) with facility fixed effects; the fact that all estimated  $\delta$  coefficients remain stable provides evidence that they are not biased by time-invariant facility unobservables correlated with treatment.

### 1.4.2 Impact on health practices and outcomes

To provide evidence on the effect of treatment on health practices and outcomes we survey households in 47 randomly chosen communities located in each of the 47 districts where the CHAs operate. We randomly choose 16 households in each community, surveying 738 in total.<sup>23</sup> As the main focus of the CHA job is mother and child health, we only survey households that contain at least one child under five. The

<sup>23</sup>The sample frame had 752 households. The 14 households difference is due to several factors. In some communities, safety concerns related to local political tensions forced the survey team to leave the community before completing surveying. In other communities, especially low-density communities where travel times between households could exceed one hour, the survey team was unable to find 16 eligible households within the allotted survey time. One household interview was lost due to malfunction of the mobile device on which the interview was recorded. The minimum number of households surveyed in a community was 13.



Table 1.6: The effect of career incentives on facility utilization

Dependent variable: total over each quarter 2011:1-2014:2	institutional deliveries	postnatal (0-6 weeks) visits	children under 5 visited	children under 5 weighed	children under 1 receiving BCG vaccinations	children under 1 receiving polio vaccinations	children under 1 receiving measles vaccinations
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Career incentives	-1.796 (10.29)	-13.81 (9.534)	-80.19 (142.0)	-86.81 (133.2)	10.48 (12.07)	-1.335 (9.075)	0.747 (10.05)
After	3.479 (4.425)	15.08** (5.191)	55.40 (63.22)	102.7 (63.91)	-1.611 (4.566)	-1.643 (3.717)	-1.517 (3.591)
Career incentives*After	14.68** (6.322)	8.253 (9.562)	318.1** (98.05)	284.3** (110.2)	7.158 (8.906)	14.98** (4.803)	11.47 (7.255)
Area characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of dependent variable in control in year 1	46.7	49.9	1312.8	1261.5	89.8	73.9	73.6
Adjusted R-squared	0.331	0.203	0.246	0.246	0.146	0.143	0.108
Number of facilities	89	118	123	123	121	120	121
Number of observations	1269	1528	1618	1610	1518	1530	1535

Notes: OLS Estimates, standard errors clustered at the district level. Data source is the Health Management and Information System (HMIS) available monthly from January 2011 until June 2014. Health center and health post staff are required to submit monthly reports that summarize their activities at the health post/community level. These are aggregated at the quarter level in the regressions. The variable in Column (1) is defined at the health center level because health centers are equipped for child births and health posts are not. The variables in Columns (2)-(7) are defined at the health post level if this reports data, at the health center otherwise. After=1 after September 2012 (from 2012:4 onwards), when CHAs started working. All regressions include the stratification variables (province dummies and share of high school graduates in the district). Area characteristics include: number of staff in the health post, geographical distribution of households in the catchment area, and an indicator variable that equals 1 if the CHA reports to have good cell network coverage most of the time or all the time.

survey contains modules on health and sanitation knowledge, health practices, incidence of illnesses and anthropometrics for the youngest child. Knowledge, practices, and illnesses are self-reported; deworming and immunization data are drawn from the child health card, and anthropometrics are measured by trained enumerators. We interview the main carer of the child, which is their mother in 90% of the cases and either a grandparent or a sibling in the remaining 10%. All questions are drawn from the DHS Zambia questionnaire, with the exception of the health knowledge module which we designed based on the CHA curriculum, and mid-upper arm circumference, which the DHS does not measure.

Table 1.7 reports the estimates of:

$$y_{idp} = \alpha + \beta C_{id} + D_i \gamma + \delta E_d + \rho_p + \epsilon_{idp} \quad (1.4.1)$$

where  $y_{idp}$  is the outcome of child (or respondent)  $i$  in district  $d$  and province  $p$ ,  $C_{id}$  equals 1 if child (or respondent)  $i$  lives in a district that is assigned to the career incentives treatment.  $D_i$  is a vector of child, respondent and household characteristics that include child age and gender, household size and number of assets, and the education level of the respondent. As above, we control for the stratification variables, district-level high school graduation rate  $E_d$  and provinces indicators  $\rho_p$  throughout and cluster standard errors at the district level.

Column 1 shows that the average respondent answers 74% of the knowledge questions correctly and this is does not differ by treatment status. In contrast, treatment affects all the health practices we collect information on. In particular, Columns 2 and 3 show that children under 2 living in treatment areas are 5 percentage points more likely to be breastfed,<sup>24</sup> and their stools are 12 percentage points more likely

<sup>24</sup>WHO recommends breastfeeding until the age of two years.

Table 1.7: The effect of career incentives on health practices and outcomes

Dependent variable	Information		Health practices				Incidence of illness		Anthropometrics			
	% of correct answers in medical knowledge test		= 1 if child under 2 yr old is breastfed	= 1 if child's stool are safely disposed	number of deworming treatments	= 1 if child exposed to CHA is on track with immunization schedule	= if child has experienced fever in the last two weeks	= if child has experienced cough in the last two weeks	= 1 if age z score <2 (moderately or severely undernourished)	= 1 if MUAC < 12.5 (moderately or severely wasted)	= 1 if MUAC < 11.5 (severely wasted)	
Career incentives	0.002 (0.010)	yes	0.051** (0.023)	0.121*** (0.039)	0.225* (0.129)	0.047** (0.020)	-0.003 (0.037)	0.037 (0.027)	-0.053* (0.030)	-0.028* (0.015)	-0.023 (0.015)	-0.014 (0.014)
household controls		no	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
child controls			yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Mean of dep var in control	.740		.641	.595	1.44	.058	.469	.255	.210	.051	.036	.014
Adjusted R-squared	0.057		0.561	0.161	0.263	0.024	0.077	0.017	-0.006	0.003	0.018	0.017
N	738		613	736	659	462	731	731	582	582	581	581

Notes: OLS estimates, standard errors, clustered at the district level. The medical knowledge test contains 14 questions on topics that CHAs are supposed to cover; these questions were drafted by the researchers in consultation with CHA program officials and the CHA curriculum. Breastfeeding and stool disposal are self-reported. In line with UNICEF (2014), we define stools to be safely disposed if flushed in toilet/latrine. Deworming, immunisation data and schedule as reported in the child health card. A child is defined to be on track if they have completed all immunisations required for their age in months. The immunisation sample is restricted to children who were 3 months or younger (including unborn) when the CHAs started working. Thresholds for weight-for-age and MUAC are taken from WHO guidelines; following these, data are restricted to children between 6-59 months. Household controls include size, education level of the respondent, number of assets. Child controls include age and gender. All regressions include the stratification variables.

to be safely disposed; these effects represent a 8% and 20% increase from the control group mean, respectively. Columns 4 and 5 show that treatment also increases the incidence of deworming treatments by 16% and the likelihood that the child is on track with the immunization schedule by 4.7 percentage points, which is 81% of the control group mean (5.8%).<sup>25</sup> Importantly, the treatment affects the incidence of immunizations for children who are young enough to have been exposed to CHAs when their immunization period started (as shown in Column 5) but not for those that were too old to start the cycle when the CHAs started working. This echoes the findings in Table 1.6 that show no difference in immunization rates between treatment and control areas before the CHAs started working.

Columns 6-8 measure treatment effects on the incidence of three main illness symptoms: fever, diarrhea and cough. These are fairly common as 47%, 26% and 45% of children in control areas had experienced them in the past two weeks. As it is widely acknowledged, self-reported symptoms can actually worsen as knowledge improves and individuals learn how to recognize them, so these effects are lower bounds. We find that treatment reduces the incidence of cough symptoms by 7 percentage points while leaving the others unchanged. Finally, Columns 9-12 show treatment effects on anthropometric measurements. We report weight-for-age z-scores and mid-upper arm circumference (MUAC). The combination of these two allows us to measure both chronic and acute malnutrition.<sup>26</sup> Following WHO's guidelines we

<sup>25</sup>A child is defined to be on track if she has completed all immunizations required for her age. At age 3 months, this includes BCG, OPV 0-2, PCV 1-2, DPT-HepB-Hib 1-2, and rotavirus 1-2. At 4 months, this includes, additionally, OPV 3, PCV 3, and DPT-HepB-Hib 3. At 9 months, this includes OPV 4 if OPV 0 was not given, and measles 1. The immunization series is complete at age 18 months with measles 2. Finally, we consider a child to be on track for vitamin A supplementation if she has ever been supplemented.

<sup>26</sup>We did not measure weight-for-height, an alternative to MUAC for assessing acute malnutrition, for three reasons. First, compared to weight and MUAC, height measurement is more invasive,

use the -2SD and -3SD thresholds for weight-for-age z-scores to measure moderate and severe underweight, respectively, and 12.5cm and 11.5cm for MUAC to measure moderate and severe wasting, respectively (Food and Nutrition Technical Assistance Project, 2011). According to these measures, 21% of the children in control areas are underweight, and 5% severely so. The incidence of wasting is much lower, with 3.6% of the children exhibiting some wasting and 1.4% severe wasting. These data, which match the corresponding DHS figures for rural Zambia (Government of Zambia, 2014), suggest that these areas are characterized by high rates of chronic malnutrition but low rates of acute malnutrition.

The findings in columns 9-10 show that children in treatment areas are 5 percentage points less likely to be underweight (25% of the control group mean) and 3 percentage points less likely to be severely underweight (55% of the control group mean). In line with this, columns 11 and 12 show a large percentage reduction in wasting, but given the limited occurrence of this in our sample the effects are not precisely estimated.

The average standardized treatment effect across all variables (coded so that higher values correspond to better outcomes) is .108, significantly different from zero at the 1% level.

Taken together, the findings in this and the previous section show that differences in the inputs provided by treatment and control CHAs are matched by differences in facility utilization and household health practices. The selection effect of career requiring, for children under two, laying the child down on a height board and having two enumerators hold the child while collecting the measurement. During survey piloting, many respondents (and the children themselves) balked at this procedure. Second, accurate height measurement is made difficult by high measurement error relative to standard effect sizes (Mwangome et al., 2012). For example, 1 millimeter is 12 percent of the increase in height-for-age typically observed in dedicated child nutrition programs (Dewey & Adu-Afarwuah, 2008). Finally, MUAC is a more accurate predictor of mortality (Myatt et al., 2006).

incentives is strong enough to generate discernible differences in household behaviors and child health outcomes.

## **1.5 Selection on observables vs. unobservables**

There are two reasons why advertising career incentives at the recruitment stage might affect performance even if all agents face the same incentives once hired: (i) agents with traits that lead to better performance are more likely to apply and (ii) agents whose elasticity of effort with respect to career incentives is high are more likely to apply. Both are selection effects that affect performance because they imply that agents in the control group have worse traits and/or respond less to career incentives even if they face the same incentives once hired. We now analyze whether career incentives attract agents who differ on observable traits (sub-section 1.5.1) and the extent to which this selection on observables can explain the performance gap identified above (sub-section 1.5.2). The answer informs the choice between career incentives and eligibility criteria at the recruitment stage. If the entire gap is due to observables, there exists a set of eligibility criteria that can mimic the effect of career incentives under the assumption that the participation constraint of those who meet the criteria is met in the absence of career incentives. In contrast, if the gap is due to unobservables, no set of eligibility criteria can mimic the effect of incentives.

### **1.5.1 The effect of career incentives on observable traits**

Table 1.8 measures the effect of career incentives on CHAs' traits that can affect performance. We group these in four categories: skills, preferences, outside option, and demographics. For each variable, the table reports the means and standard deviations in treatment and control, as well as the p-value of the test of means equality,

Table 1.8: The effect of career incentives on CHA's traits

	treatment	control	p-values
<b>Panel A: Skills</b>			
Average test score at training [0-100] *	69.2 (7.23)	68.0 (6.75)	.067
O-levels total exam score *	25.3 (9.92)	24.5 (8.70)	.559
O-levels passed in biology and other natural sciences *	1.47 (.868)	1.39 (.824)	.801
<b>Panel B: Motivation and preferences</b>			
Psychometric scale: Career orientation [1-5]	3.30 (1.050)	3.08 (.939)	.025
Psychometric scale: Pro-social motivation	3.64 (.541)	3.63 (.541)	.623
Psychometric scale: Desire for positive pro-social impact [1-5]	4.43 (.444)	4.43 (.509)	.824
Psychometric scale: Affective commitment to beneficiaries [1-5]	3.81 (1.153)	3.83 (1.170)	.873
Donation to local hospital (dictator game)	4063 (4018)	3922 (3937)	.739
Main goal is "career advancement" vs. "service to community"	.138 (.346)	.055 (.228)	.015
<b>Panel C: Outside opportunity</b>			
Farmer (=1 if yes)	.717 (.452)	.659 (.476)	.441
Houseworker (=1 if yes)	.103 (.025)	.141 (.030)	.586
Trader (=1 if yes)	.090 (.287)	.081 (.275)	.928
Teacher (=1 if yes)	.041 (.200)	.015 (.121)	.108
<b>Panel D: Demographics and socio-economic status</b>			
Gender (=1 if female)	.450 (.499)	.585 (.494)	.083
Age (years)	28.66 (6.42)	26.93 (5.49)	.005
Married (=1 if yes)	.462 (.500)	.510 (.502)	.156
Number of dependents	3.50 (2.54)	3.26 (2.56)	.369
Aims to remain in the same community in 5-10 years (=1 if yes)	.575 (.496)	.612 (.489)	.392
Poor (self reported) (=1 if yes)	.219 (.419)	.204 (.404)	.507
Number of household assets	5.07 (2.58)	5.22 (3.11)	.477
Owens transport (=1 if yes)	.781 (.439)	.741 (.415)	.651

Notes: Columns 1 and 2 show means and standard deviations in parentheses. Column 3 reports the p-values of the null hypothesis that the career treatment effect equals zero conditional on stratification variables and with standard errors clustered at the district level. Variables denoted by \* are drawn from MOH administrative data, all other variables are drawn from surveys administered to CHAs at the interview or during the training program. The sample is the 307 CHAs deployed. Average test score at training equals the average score in 11 tests on basic medical practices taken during the training program. Ordinary levels or O-levels are administered by the Examinations Council of Zambia (ECZ) to 12th-grade students, the highest grade in the Zambian secondary education system. O-levels total exam score is constructed as the sum of inverted O-levels scores (1=9, 2=8, and so on) from all subjects in which the applicant wrote the exam, so that larger values correspond to better performance. O-levels passed in biology and other natural sciences equals the number of O-levels passed in biology, chemistry, physics, science and agricultural science. Career orientation: from Wrzensniewski et al.'s (1997) Career-Calling Orientation scale, which consists of three items: "I expect to be in a higher-level job in five years," "I view my job as a stepping stone to other jobs," and "I expect to be doing the same work as a CHA in five years," each scored on a five-point scale from "strongly disagree" to "strongly agree." The psychometric measures of pro-sociality are adopted from Grant (2008). Each measure takes on a value between 1 and 5 and represents, among the statements listed below, the extent to which the applicant agreed, on average. Levels of agreement are 1 (strongly disagree), 2 (disagree), 3 (neither agree nor disagree), 4 (agree), and 5 (strongly agree). Statements for the other variables are as follows: Desire for positive pro-social impact includes "It is important to me to do good for others through my work," "I care about benefiting others through my work," "I want to help others through my work," "I want to have positive impact on others through my work," "I get motivated by working on tasks that have the potential to benefit others," "I like to work on tasks that have the potential to benefit others," "I prefer to work on tasks that allow me to have a positive impact on others," "I do my best when I'm working on a task that contributes to the well-being of others," "It is important to me to have the opportunity to use my abilities to benefit others," "It is important to me to make a positive difference in people's lives through my work," "At work, I care about improving the lives of other people," and "One of my objectives at work is to make a positive difference in other people's lives." Sees self as pro-social: "I see myself as caring," "I see myself as generous," and "I regularly go out of my way to help others." Affective commitment to beneficiaries includes "The people who benefit from my work are very important to me" and "The people who benefit from my work matter a great deal to me." Donation to local hospital: trainees are given 25,000 Kwacha (approximately \$5) and invited to donate any portion (including nothing) to the local hospital to support needy patients. This donation decision occurs privately and confidentially in concealed donation booths.

controlling for the stratification variables and with standard errors clustered at the level of randomization—the district.

To measure skills we use the CHAs’ test scores in the examinations they took during the one-year training program. These examinations test the material taught in the program that will directly inform the work of the CHAs in the field. As all trainees are informed about career incentives at the beginning of the training program, differences in test scores solely reflect the selection effect of career incentives. We complement these test scores with MoH’s records of the CHAs’ high school results.<sup>27</sup> Panel A shows that career incentives attract higher-skilled candidates: treatment CHAs’ test scores are 18% of a standard deviation higher than control CHAs’. Differences in test scores date back to high school as treatment CHAs’ O-level scores are 9% of a standard deviation higher, and the number of O-level exams passed in the natural sciences is 10% of a standard deviation higher, although these differences are not precisely estimated.

Panel B measures two sources of motivation that are relevant in this context: career ambition and pro-sociality. Differences in career ambitions and pro-sociality can drive differences in performance if more ambitious CHAs work harder to reach their goals and more pro-social CHAs work harder because they put a larger weight on the welfare of the individuals they serve. To measure these preferences we give trainees a battery of psychometric tests using validated scales commonly used in employment surveys. Full descriptions of these variables can be found in Appendix A.2.4. We also implement a contextualized dictator game to measure the strength of

---

<sup>27</sup> As noted above, applicants were required to have finished grade 12 with two passed O-levels. The Examinations Council of Zambia requires that candidates take a minimum of six O-level exams, with English and mathematics being compulsory. In addition, students choose among subjects in the natural sciences, arts and humanities, and business studies.



pro-social preferences.<sup>28</sup> Finally, we measure the relative strength of career vs. pro-social preferences by asking trainees to choose whether they see “career advancement” or “service to community” as the main goal of the CHA job. While both career ambitions and pro-sociality can lead to higher performance, there might be cases in which a tradeoff arises between the two goals, and the effect on performance is ambiguous *a priori*.<sup>29</sup>

The data in Panel B show that treatment CHAs have stronger career ambitions but the same level of pro-social motivation as control CHAs. In line with this, when asked to choose between “career advancement” or “service to community,” only a minority chooses “career advancement,” but this is larger in the treatment group (14% vs. 6%,  $p=.015$ ).

Panel C reports CHAs’ occupation at the time of application. This is relevant both because it allows us to assess whether the CHA program crowds out talent from other sectors, and because CHAs with worse outside options might work harder to keep their CHA job (although, given the low frequency of dismissals of government employees, this effect is unlikely to be strong). Four categories account for over 90% of occupations and all four are similar in treatment and control. Over two-thirds of applicants in both treatment and control groups are farmers. This is more than

---

<sup>28</sup>In the dictator game, we gave trainees 25,000 Kwacha (approximately USD 5; half of a CHA’s daily earnings) and invited each to donate any portion (including nothing) to the local hospital to support needy patients. This donation decision occurred privately and confidentially in concealed donation booths. Previous work has found dictator games adapted for specific beneficiary groups to be predictive of performance on pro-social tasks Ashraf et al. (2014) and choices of public sector nurses to locate to rural areas Lagarde & Blaauw (2013).

<sup>29</sup>To interpret the results in Panel B we need to keep in mind that these measures are self-reported and CHAs might give answers that are consistent with the recruitment poster rather than express their true preferences. Two considerations allay this concern: (i) the measures are collected after CHAs have been selected, so they have no incentive to modify their answers to affect the probability of selection, and (ii) psychometric tests are not straightforward to game.

double the share of farmers in the general population of eligibles (Table 1.1). The two other occupations listed by respondents are “trader” and “teacher,” both of which are likely to have a higher return to skills than farming. These are slightly, but not significantly, more common in the treatment group and substantially lower than in the general population of eligibles. Housework is slightly, but not significantly, more common in the control group and higher than in the general population of eligibles. Noticeably, only 13% of the sample reports being unemployed, but in the absence of information on hours worked we cannot rule out that the data in Panel C hides underemployment. Regardless of the true share of unemployed, Panel C makes clear that a large majority of CHAs were not in jobs fit to their skill levels. The program might crowd out some agricultural production, but it is not drawing talent from other professions.

Finally, Panel D shows that treatment CHAs are older and more likely to be male, but have similar socio-economic status as the control CHAs.

Taken together, the data in Table 1.8 reveal that individuals in the two groups differ on some relevant traits. In the Appendix we show that this is driven by differential sorting, namely by the fact that career incentives attracted different types, rather than by differential selection by recruitment panels. In short, panels in the treatment and control groups put the same weight on the same traits, but they face different applicant pools.

## **1.5.2 Explaining the performance gap**

We now establish the extent to which differences in performance identified in Section 1.3 are due to selection on observables vs. unobservables. To do so, we augment specification (1.3.1) by adding the individual traits that differ significantly between treat-

ment and control groups. If differences in performance disappear, we can attribute the selection effect entirely to the fact that career incentives attract applicants with different observable traits. If differences in performance remain, we conclude that the selection effect is partly due to the fact that career incentives attract applicants with different unobservable traits.

Table 1.9, column 1 replicates the baseline estimates in Table 1.3. Columns 2 to 5 add skills, preferences, and demographics, individually and then jointly. Column 2 shows that, as expected, skills are positively correlated with performance, but their effect is of a magnitude smaller than the effect of career incentives. A one standard deviation increase in test scores increases visits by 28, which is 32% of the effect of career incentives. Differences in skills only explain a small share of the performance gap: after controlling for skills, the difference in visits done by treatment and control CHAs drops from 94.0 to 89.1.

Column 3 shows that the intensity of career preferences is positively correlated with performance, as we would expect, but the effect is small (a one standard deviation increase leads to 6.5 more visits) and not precisely estimated. In contrast, CHAs who put career advancement over service to the community do 58 *fewer* visits. Because these types are more common in the treatment group the estimated effect of career incentives slightly increases from 94.0 to 97.1. This is in line with the hypothesis that strong incentives can crowd out pro-social types, and this can harm performance, but the crowding out is rather weak—only 14% of CHAs in the treatment group (and 5% in control) put career advancement over service to the community; the remaining 86% who do not perform better than their counterparts in the control group.

Finally, column 4 shows that there are no gender differences in performance (the coefficient is small and not significantly different from zero) but older CHAs perform better: one standard deviation increase in age (5.5 years) increases visits by 34.1.

Table 1.9: The effect of career incentives on performance: observables vs. unobservables.

	Household visits				
	<i>(SMS receipt data: Total August 12-January 14)</i>				
	(1)	(2)	(3)	(4)	(5)
Career incentives	93.95** (37.19)	89.08** (37.46)	97.10** (37.98)	83.05** (38.57)	85.69** (38.74)
Average test score at training		4.185** (2.001)			3.013 (1.997)
Main goal is "career advancement" vs. "service to community"			-57.79* (32.12)		-63.75* (32.54)
Psychometric scale: Career orientation [1-5]			6.458 (15.33)		8.576 (15.12)
Female				7.842 (35.75)	17.26 (36.41)
Age				6.240** (2.251)	5.382** (2.238)
Area characteristics	Yes	Yes	Yes	Yes	Yes
p-value of the test that individual controls are jointly=0		0.04	0.17	0.03	0.00
Mean of dependent variable in control	318.6	318.6	318.6	318.6	318.6
Adjusted R-squared	0.112	0.122	0.106	0.128	0.126
N	307	307	307	307	307

Notes: OLS Estimates, standard errors clustered at the district level. The dependent variable is total number of household visited between August 12 and January 14. SMS receipts are sent by individual CHAs to MOH for each visit. Average test score at training equals the average score in 11 tests on basic medical practices taken during the training program. Career orientation: from Wrzenski et al.'s (1997) Career-Calling Orientation scale, which consists of three items: "I expect to be in a higher-level job in five years," "I view my job as a stepping stone to other jobs," and "I expect to be doing the same work as a CHA in five years," each scored on a five-point scale from "strongly disagree" to "strongly agree." All regressions include the stratification variables (province dummies and share of high school graduates in the district). Area characteristics include: number of staff in the health post, geographical distribution of households in the catchment area, and an indicator variable that equals 1 if the CHA reports to have good cell network coverage most of the time or all the time.

Since CHAs in the treatment group are on average older, the difference in visits done by treatment and control CHAs drops from 94.0 to 83.1.<sup>30</sup>

Taken together, the evidence in Table 1.8 and 1.9 indicates that career incentives attract agents with different observable traits, but while these are correlated with performance, their effect is small relative to the effect of career incentives and they explain a small share of the observed performance gap. Comparing columns 1 and 5, Table 1.9, shows that indeed the gap falls by 9% when all the traits that differ significantly between the two groups are accounted for.

## 1.6 Conclusion

Attracting effective employees is a core objective for all organizations. Our analysis shows that advertising career incentives at the recruitment stage draws in individuals who perform well in the health sector. Importantly, the selection effect deriving from incentives cannot be mimicked by a modification of the eligibility criteria, which highlights the importance of incentive design at the recruitment stage.

The findings suggest that estimates of the effects of incentives on performance obtained by strengthening incentives for a given set of agents might understate their true impact, both because they do not take into account the selection effect and because they measure the response of agents who have self-selected into jobs with low-powered incentives, and hence might be less responsive to high-powered incentives in the first place.

---

<sup>30</sup>Further analysis, not reported, shows that the effect of observable traits on performance is the same in both groups, suggesting that these traits affect performance directly rather than by determining the response to career incentives—the sole exception is age, which is associated with performance in the control, but not in the treatment group.

The findings also allay the concern that offering material rewards for public service delivery jobs displaces applicants with desirable social preferences and ultimately worsens the quality of services provided. Naturally, the type of material benefit offered—a career in the Ministry of Health—was unlikely to attract purely selfish types, since government service implies some pro-social benefit. The findings do not rule out the possibility that there exists a level of financial compensation that attracts callous types, but rather they suggest that the material benefits that can be reasonably associated with these jobs have no drawbacks in terms of pro-social motivation and performance. The findings have implications for policy strategies based on this concern, such as maintaining the volunteer status of community-based work, or low salaries and lack of career incentives in teaching and health professions (World Health Organization, 2006; Lehmann & Sanders, 2007).

Our research provides evidence on factors that inform the welfare analysis of providing career incentives, but is not designed to conduct a full welfare analysis for three reasons. First, due to political constraints, all agents had to be paid the same amount. This implies that we cannot judge whether agents attracted by career incentives have a higher reservation wage, such that their higher performance comes at a price; in other words, the government could get the agents in the control group to work for a lower wage. A priori, the difference in reservation wages between applicants in the two treatments is difficult to sign: that applicants to the career incentives treatment are more skilled suggests that it might be positive, whereas the fact that they expect to move on to better-paid positions suggests that it might be negative (like interns are typically willing to forego compensation for the sake of career opportunities). Regardless, our results suggest that higher wages and career incentives can be substitutes for drawing candidates with better outside options and consequently higher skills. However, career incentives may be cheaper for the organization if the

organization also requires higher-level positions to be filled, and has trouble filling them.

Second, while retention rates after 18 months are the same in the two groups, agents in the career incentives treatment might leave their posts for higher-ranked positions sooner than those in the control group. Whether this entails a welfare cost depends on whether they can be easily replaced and whether the government can use their skills in other jobs. In our context, replacement is straightforward; the number of applicants per post was above seven, and the government faces scarcity of health staff at all levels, such that promoting high-performing CHAs to nursing and other higher-level cadres is likely to be welfare-improving. In contexts where retention in the original post is more important, the welfare cost of attracting agents who expect to move on will be higher.

Third, since over 80% of CHAs were engaged in subsistence farming or housework we cannot quantify the opportunity cost of the CHAs' time, namely the value of the activities they give up to become full time health workers and the size of this difference between treatment and control. If productivity in these alternative occupations is increasing in the same qualities that make a CHA productive the findings imply that the opportunity cost is higher in the career treatment, namely the career treatment draws in more productive farmer or houseworkers. By revealed preferences we know that the private value of the CHA jobs must be at least equal to the private value of these activities (otherwise these individuals would have not switched occupations) but we cannot quantify the extent to which the social value produced by career CHAs in their new jobs exceeds the loss in social value from agriculture and housework.

Finally, the fact that CHAs are recruited locally from the communities where they are meant to serve implies that there is no competition for talent across communities. This has implications for the scale-up of the program as career incentives can be

offered in each community without losing effectiveness as each community can only hire from their own pool, and most communities in these areas have access to a pool of skilled individuals who are either unemployed or in low skills jobs. More generally, in a context of nearly full employment when different organizations compete for the same pool of talent, focussing on incentive design at the recruitment stage might still be valuable to the extent that it improves the quality of the match between employees and organizations.



# Chapter 2

## Awards Unbundled: Evidence from a Natural Field Experiment<sup>1</sup>

### 2.1 Introduction

What are the advantages which we propose by that great purpose of human life which we call bettering our condition? To be observed, to be attended to, to be taken notice of with sympathy, complacency, and approbation, are all the advantages which we can propose to derive from it.

—ADAM SMITH, “OF THE ORIGIN OF AMBITION, AND OF THE  
DISTINCTION OF RANKS,” THE THEORY OF MORAL SENTIMENTS  
(1759)

---

<sup>1</sup>Jointly authored with Nava Ashraf and Oriana Bandiera. This chapter is the fruit of a collaboration with the Ministry of Health, Zambia, to whose staff we owe a tremendous debt, especially Mutinta Musonda and Miriam Libetwa. We also thank Alice Nyirenda and the teachers of the CHA Training School for their patience and valuable feedback. Madeleen Husselman, Mardieh Dennis, Manpreet Singh, Katherine Otto, Johann Blauth, and, especially, Kristin Johnson provided skilled research assistance. We thank Charles Angelucci, Ed Glaeser, Francesca Gino, Daisuke Hirata, Kelsey Jack, Bryce Millett Steinberg, Emily Oster and Michele Rigolizzo, and participants at the Stanford Institute for Theoretical Economics, Psychology and Economics session, for very helpful input. Financial support by the International Growth Centre is gratefully acknowledged.

The innate human desire for approbation might make status awards a cost-effective tool to incentivize good performance (Besley & Ghatak, 2008; Moldovanu et al., 2007). Awards can motivate employees to exert effort in order to gain recognition and visibility, both of which are free for the employer to bestow but valuable to the employee. However, given that awards derive their value from their scarcity, they inevitably facilitate social comparisons, which might be *demotivating* to employees.<sup>2</sup>

Our goal in this paper is to “unbundle” awards—that is, to provide evidence on the mechanisms that underlie their effectiveness and potential harm. We conduct a natural field experiment to separately identify channels through which awards can affect behavior, unbundling the effect of social comparison through the disclosure of rank information, from the effect of employer recognition and social visibility.

We study the effect of awards in the context of a nationwide training program for health workers in Zambia. Our agents are 314 health workers recruited from 162 rural communities and brought to professional school for a one-year training program aimed at teaching community-based health care. After training, trainees will be employed by the Ministry of Health and deployed to their communities of origin, where they will become the first point of contact for health services. Incentivizing learning is key in this context because trainees have no previous medical training; thus, the skills they learn will determine their effectiveness in the field.<sup>3</sup>

---

<sup>2</sup>Lazear (1989) describes the tradeoff in relative performance evaluation: it could motivate employees to work harder, but could also create an excessively competitive work environment and decrease employee morale. Major et al. (1991)’s review of the literature in social psychology provides evidence on the demotivating effect of social comparisons. A related literature in management emphasizes the importance of concealing relative performance information to improve employee motivation (Milkovich & Newman, 1996).

<sup>3</sup>A number of field experiments have evaluated the effect of financial incentives on student learning; the evidence of their efficacy is mixed (Fryer, 2011; Angrist & Lavy, 2009; Kremer et al., 2009; Leuven et al., 2010).

During the training program, trainees take courses on several topics, on which they are tested at baseline (at the beginning of the year) and at the end of each course. The field experiment randomly allocates trainees to two broad classes of treatments (in addition to control): those that only provide information on trainees' relative performance, and those that also offer awards. After each exam, trainees in the control group receive a letter from the school reporting their absolute score and their value added over their baseline score for the given course. Trainees in the "private social comparison" treatment (T1) receive the same letter with added information on their rank in the class distribution of both absolute score and value added measured as improvement from baseline. Trainees in the "public social comparison" treatment (T2) receive the same letter as in the previous treatment as well as the names of the top four performers in the class (top two by absolute score and top two by value added).

The third and fourth treatments add awards to rank information. Awards are given to the trainees with the top two scores and those with the top two most improved scores (from baseline). The latter ensures that weaker trainees have a chance to win and are therefore motivated by the award. In the "employer recognition award" treatment (T3), the top four performers receive a congratulatory letter from the Ministry of Health. In the "social visibility award" treatment (T4), one of the top four performers is randomly selected to be featured in an interview, which is printed along with the candidate's photo in a newsletter distributed back to their community of origin. Under a linearity assumption the difference between each of the award treatments and the "public social comparison" treatment isolates the effect of awards from the effect of the social comparisons they inevitably create.

Our setting has three key features that make it ideal for the purpose of this experiment. First, since trainees take four courses during the study period and treatments

are announced at the beginning of the first course, we can assess whether they change their behavior in anticipation of receiving rank information and awards or only after these have been provided. Second, during training, the performance of the health workers is measured by an institution (the school) that is different from their employer (the Ministry), and the health workers are physically removed from their home communities. This allows us to separate the effect of information on relative performance (provided by the school) from that of the employer’s recognition and from visibility to one’s social circle (the home community). In most settings, the employer measures and provides information on performance, such that the provision of information necessarily entails some recognition. The fact that trainees are distant from their communities is similarly useful, as no treatment other than the social visibility award can be used to enhance visibility within their social circles. In most settings in which agents are co-located with their social network, any treatment that reveals their rank in the distribution could potentially be used to enhance visibility.

Third, performance in this setting is uni-dimensional (trainees are solely meant to attend classes and study the topics on the syllabus), and thus not subject to a multitasking problem in the face of additional incentives. Moreover, performance can be measured objectively and precisely by test scores. Value added in test scores is a good measure of learning, as is often the case when knowledge at baseline is very limited (Muralidharan & Sundararaman, 2011). Critically, we can show that in this context exam performance is correlated with future performance in the field Ashraf et al. (2015).

The analysis reveals that social comparison and awards have opposite effects on performance. Compared to trainees in the control group, the “private social comparison” treatment significantly reduces test scores by 0.31 standard deviations, and the “public social comparison” treatment reduces it by 0.38 standard deviations. Im-

portantly, the two social comparison treatments reduce performance as soon as they are announced—i.e. before trainees get the first letter with their rank information. A likely explanation is that individuals value the *belief* that they have high relative ability, and the anticipatory utility this provides. They may thus prefer to exert low effort in order to decrease the informativeness of the ranking signal. This is akin to refusing to take a medical test for a disease, so as to justify holding an optimistic belief about one’s health status (Oster et al., 2013), and is consistent with the literature on belief utility and information avoidance Bénabou & Tirole (2002); Köszegi (2002). In our context, the negative effect due to information avoidance seems to dominate the potential positive effect of competition among trainees (Charness & Grosskopf, 2001; Freeman & Gelber, 2010).<sup>4</sup>

Adding awards to rank information significantly improves performance. Compared to trainees in the “public social comparison” treatment, mean scores of trainees in the “employer recognition award” and “social visibility” treatments are 0.38 and 0.44 standard deviations higher, respectively. Recognition from one’s employer can increase performance if agents have career concerns (Dewatripont et al., 1999) or preferences for reciprocity (Fehr & Schmidt, 1999). The net effect of either type of award is nil: because the positive effects of employer recognition and social visibility are nullified by the negative effect of providing information on relative ranks, trainees in the two award treatments perform as well as trainees in control.<sup>5</sup>

---

<sup>4</sup>The empirical evidence on the ex-post effect of providing rank information is markedly mixed. Tran & Zeckhauser (2012), Azmat & Iriberry (2010), and Dur et al. (2013) show positive effects of rank information on performance, while Bandiera et al. (2013) and Barankay (2012) find this information reduces productivity.

<sup>5</sup>This echoes the findings of Bandiera et al. (2013) who show the impact of receiving information on relative rankings can offset the positive impact of monetary prizes. In their setting, however, the effect is driven by changes in team composition.

Quantile treatment estimates show that both the negative effect of social comparison and the positive effects of recognition and visibility are stronger on the left tail of the conditional productivity distribution, and both are zero at the top two deciles. In line with this, we also find that these negative and positive effects are stronger for low-ability trainees, and zero for high-ability trainees, where ability is measured by baseline test score. The fact that the negative effect of ranking is stronger for the weakest trainees is intuitive, as these are more likely to receive a negative signal about their skills.<sup>6</sup> That the positive effects of recognition and visibility are also stronger for the weakest trainees may be due to the fact that, since awards are given to trainees with the highest value added (the “most improved”), those who start at the bottom have a better chance to win. That the effects are zero for the top two deciles is consistent with the fact that scores are capped, such that the highest performing trainees have little room for improvement. Evidence from settings with no cap on performance suggests that, in contrast, awards are most effective at the top of the distribution (Nalbantian & Schotter, 1997; Müller & Schotter, 2010; Bandiera et al., 2013; Leuven et al., 2010).

An important implication of these findings is that, due to the negative effect of social comparison, awards might increase the dispersion of performance by weakening the weakest. In our setting, this may matter just as much as—and perhaps more than—mean effects. Health workers wield considerable power to influence the utility of their patients, especially for the worse (e.g., missed diagnoses, incorrectly dosed medications, wrong-site amputations). Both because of the potential for harm and the government’s mandate to ensure equity of services across populations, the distri-

---

<sup>6</sup>Tran & Zeckhauser (ibid.) find that private rank disclosure motivates high-ability more so than low-ability trainees, whereas Azmat & Iriberry (ibid.) find uniformly positive effects across the distribution.

bution of performance during training in this field experiment is crucial. The findings thus suggest caution in using mechanisms that facilitate interpersonal comparisons in contexts in which worsening performance at the bottom of the distribution is costly. This is particularly important in the policy domains of public service delivery, such as in health and education, where the use of awards is increasingly common (Mathauer & Imhoff, 2006; Ashraf et al., 2014), but where distributional effects on agents' performance could have severe welfare consequences for those they are serving.

The paper is organized as follows. Section 2.2 describes the context and the experiment. Section 2.3 presents the empirical analysis. Section 2.4 interprets the findings in light of optimal expectations theory (Brunnermeier & Parker, 2005; Oster et al., 2013). Section 3.6 concludes.

## 2.2 Experimental design

### 2.2.1 Context

In 2010, the Government of Zambia (GOZ) launched a national effort to create a new civil service cadre called the Community Health Assistant (CHA). In the program's first year, GOZ sought to recruit, train, and deploy approximately 300 Community Health Assistants across seven of Zambia's nine provinces.<sup>7</sup> Within these seven provinces, based on population density, GOZ chose the 48 most rural of the 58 constituent districts, and across these, GOZ identified 165 underserved communities, each with an average population of 3,500 individuals (Government of Zambia,

---

<sup>7</sup>This is the first generation of community health workers trained by the Government of Zambia. Although this paper does not evaluate the efficacy of community health workers, they have been shown to improve health outcomes in randomized controlled trials in other countries (Baqui et al., 2008; Bhandari et al., 2011; Spencer et al., 2011).

2011). From each community, the intention was to recruit two Community Health Assistants. The recruitment and selection process occurred at the community level, with on-the-ground implementation coordinated by district health officials.<sup>8</sup>

In total, 314 individuals accepted GOZ's training offer and moved to Ndola, Zambia's second-largest city, to join a newly established training school. The training program lasted one year and was structured in a modular format (see Figure B.1).<sup>9</sup> At the beginning of the training, all trainees took a baseline exam which covered all the material that would be subsequently taught during the training year. After each course, trainees took an exam covering the material taught in that course.<sup>10</sup> During this time, the trainees engaged only in attending classes and studying for their exams. The training school was divided into five classrooms, each accommodating roughly 60 trainees. The school was led by a principal and staffed by ten full-time teachers. The trainees were not formally paid during the training year, but their tuition and room and board were covered by the Ministry. In addition, the participants were aware that wages upon completing training would be the same for all CHAs, and that opportunities for promotion would be available. The program is thus effectively training "on the job" and career concerns were likely at play.

Once deployed to the field, the CHAs were to routinely visit households and provide a variety of home-based services: basic medical care to any sick persons, health education and counseling, and referrals to nearby health facilities as needed. Two key

---

<sup>8</sup>See Ashraf et al. (2015) for details on the selection process of the Community Health Assistants.

<sup>9</sup>The training curriculum was designed in the months leading up to the training launch through a consultative process led by the Zambian Ministry of Health, with input from health educators, clinicians, and public health and development practitioners.

<sup>10</sup>Each course also had a "practical" component for which trainees visited field sites. Performance in this component was not tested.



features of the job illuminate how critical the training period was to subsequent performance. First, CHAs were expected to provide a very broad scope of health services to all age groups. Second, they were to do so with a great deal of autonomy. In contrast to nurses, whose job it classically is to implement a physician's orders, and who typically are not trained to diagnose and treat illnesses, the CHA is more like a physician, making decisions autonomously without direct supervision. The human capital required to perform such varied activities is substantial, and the one-year training was consequently critical.

### **2.2.2 Performance measurement**

Since the trainees' only task during training was to attend classes and study the material taught therein, we measure performance by exam test scores. All exams were based on multiple choice questions created by external medical advisors based on the content of the official training textbook.<sup>11</sup> Grading was done electronically by the research team. After each course exam, each student's completed exam was returned to him or her, along with an answer key.

Several measures were taken to prevent cheating or gaming. First, all exams were administered under timed, proctored conditions. Second, each exam had four versions, in which the order of the answer choices for each question was randomly varied. The exam versions were distributed within a classroom in sequential fashion, such that no two neighboring trainees had the same exam version. Even if they were alike, the exam version was indicated discreetly in the lower corner of the exam,

---

<sup>11</sup>Multiple choice questions are a standard question type on exams in Zambian secondary and tertiary education.

such that a student attempting to cheat by copying a neighbor’s answers would have difficulty determining whether the neighbor’s exam version was the same as hers.

### 2.2.3 Experimental design

We worked with the school administrators to randomly allocate trainees to five groups (one control and four treatments) of approximately 60 individuals each, stratified by average baseline exam score and other potential determinants of performance.<sup>12</sup> To minimize contamination across treatments, all trainees in a given treatment group were assigned to the same classroom, and classrooms were kept together for the entire duration of the experiment that lasted nine consecutive months. For each course, each classroom was co-taught by two teachers. The teachers rotated and were assigned to different classrooms after each course, using a schedule that was determined by the principal and by the researchers with the aim of ensuring even coverage of teachers across classrooms. Teachers and trainees in all groups used the same textbook that was developed by GOZ for the CHA training.

The experimental treatments are as follows (see Figure 2.1 for a schematic diagram). After each exam, trainees in the control group receive a letter from the school reporting their absolute score and their value added over their baseline sub-score for the relevant course content after each exam (see Figure B.2). Trainees in the “private social comparison” treatment (T1) receive the same letter from the school with added information on their rank in the distribution of both absolute score and value added (see Figure B.3). Trainees in the “public social comparison” treatment (T2) receive

---

<sup>12</sup>We used the “T-min-max” method to balance the classrooms on gender, baseline exam score, any previous health experience, employment status and district-level recruitment strategy (different districts advertised the CHA position with different emphases on social vs. private benefits as described in Ashraf et al. (2015)). For a discussion of this randomization method, see Bruhn & McKenzie (2009).

the same letter from the school with added information on their rank and the names of the top four performers in the class—that is, the top two by absolute score and the top two by value added (see Figure B.4).<sup>13</sup>We include the “most improved” category based on theoretical and empirical evidence that multiple prizes at different points in the distribution are more motivating across the distribution than a single prize (Moldovanu & Sela, 2001; Freeman & Gelber, 2010; Dur et al., 2013). Importantly, trainees did not know that they could have won “value added” prizes when they took the baseline test, so there is no scope for gaming by obtaining a low score at baseline.

Trainees in the “employer recognition award” treatment (T3) receive the same letter as those in the “public social comparison” treatment, and, in addition, the top four performers receive a congratulatory letter from the Ministry of Health, handwritten and signed by the director of the department that runs the CHA program (see Figure B.5 for an example). Trainees in the “social visibility award” treatment (T4) receive the same letter as those in the “public social comparison” treatment, and, in addition, one of the top four performers is randomly selected to be featured in an interview that is printed together with the candidate’s photo in a newsletter that is distributed back to their community of origin (see Figure B.6 for an example).

The timeline of the experiment is as follows. Trainees took four sequential courses during the experimental period (covering 9.5 out of the 12 months of training) and sat exams at the end of each course. Courses varied in duration from two weeks (course 2) to four months (course 1). At the beginning of the first course, trainees were told about their treatments, and a reminder was delivered towards the middle of the same course.

---

<sup>13</sup>The design decision to make comparisons only at the very top public was done in consultation with senior teachers who were concerned about severe demotivating effects of advertising the lowest performers. Although ranking throughout the entire distribution is publicly displayed in many professional schools in Zambia, there is no consensus on whether this is helpful or harmful.

Figure 2.1: Schematic of experimental treatment conditions

	Components				
	After each course exam, each trainee is given a private letter that indicates his or her <b>absolute score</b> on the exam as well as his or her <b>value added</b> with respect to his or her course sub-score in the baseline exam.	The letter reports the trainee's <b>rank</b> within the classroom on absolute score and value added, respectively.	The two trainees with the highest exam score and the two trainees with the highest value added are <b>publicly named in the letters</b> given to all trainees within the classroom.	The four top-performing trainees are given a <b>personalized letter from the program director at the Ministry of Health</b> congratulating them on their achievement.	One of the four top-performing trainees is randomly chosen to be <b>profiled in a newsletter</b> that is sent to all of the trainees' communities of origin.
<b>C:</b> Control	✓				
<b>T1:</b> Private Social Comparison	✓	✓			
<b>T2:</b> T1 + Public Social Comparison	✓	✓	✓		
<b>T3:</b> T2 + Employer Recognition Award	✓	✓	✓	✓	
<b>T4:</b> T2 + Social Visibility Award	✓	✓	✓		✓

During the course of the training year, 6 of 314 (1.9%) trainees dropped out of the program, in most cases because it was discovered that the trainee was pregnant, and the school could not accommodate infants in the dormitories. These trainees were distributed across four of the five treatment conditions (one in the control group, one in T2, three in T3, and one in T4) and have been excluded from our analysis. Due to the very low rate of attrition, any imbalance between treatment conditions arising from this attrition is indistinguishable from random error.

Trainees were told the content of the letters that they were to receive after each exam and, in addition, trainees in the two awards treatments were shown sample employer recognition letters and community newsletters. In an assessment given to all trainees after the first letter was sent out, 79% of CHAs responded to the question, “How clear do you find the information presented in the letters?,” with “very clear” and an additional 14% with “somewhat clear.”

Table 2.1 shows balance across the five groups on a number of variables including baseline exam score, English exam score, gender, age, health experience and employment status. Only 3 out of 24 (12%) pairwise differences are different from zero at the 10% level, as expected by chance.

## **2.3 Analysis**

### **2.3.1 Methodology**

During the experimental study period trainees attend four courses and take an exam at the end of each. Trainees in all treatments take the same courses and complete the same exams at the same time.

Table 2.1: Randomization balance

	C: Control	T1: Private Social Comparison	Test of equality: C=T1	T2: T1 + Public Social Comparison	Test of equality: C=T2	T3: T2 + Employer Recognition Award	Test of equality: C=T3	T4: T2 + Social Visibility Award	Test of equality: C=T4	Joint test of equality: C=T1=T2= T3=T4
Baseline test score (0-100)	0.610 (0.088)	0.600 (0.090)	$p = 0.541$	0.598 (0.094)	$p = 0.463$	0.609 (0.087)	$p = 0.936$	0.613 (0.084)	$p = 0.864$	$p = 0.865$
English test score (0-100)	39.97 (11.89)	44.56 (11.07)	$p = 0.048$	42.81 (15.70)	$p = 0.219$	40.1 (10.70)	$p = 0.944$	43.26 (13.62)	$p = 0.155$	$p = 0.193$
Number of O-Levels	5.97 (1.36)	5.93 (1.56)	$p = 0.898$	5.89 (1.57)	$p = 0.807$	5.91 (1.68)	$p = 0.847$	5.86 (1.67)	$p = 0.720$	$p = 0.993$
Age (years)	26.24 (5.73)	26.02 (5.60)	$p = 0.843$	27.01 (5.95)	$p = 0.466$	27.33 (6.05)	$p = 0.313$	28.47 (6.60)	$p = 0.040$	$p = 0.169$
Gender (=1 if female)	0.459 (0.502)	0.583 (0.497)	$p = 0.174$	0.484 (0.504)	$p = 0.784$	0.484 (0.504)	$p = 0.784$	0.516 (0.504)	$p = 0.529$	$p = 0.695$
Unemployed at time of application (=1 if yes)	0.705 (0.460)	0.639 (0.484)	$p = 0.460$	0.565 (0.500)	$p = 0.112$	0.565 (0.500)	$p = 0.112$	0.565 (0.500)	$p = 0.112$	$p = 0.393$
Work experience in the health sector (=1 if yes)	0.361 (0.484)	0.344 (0.479)	$p = 0.855$	0.484 (0.504)	$p = 0.168$	0.452 (0.502)	$p = 0.309$	0.516 (0.504)	$p = 0.082$	$p = 0.221$
District used community recruitment (=1 if yes)	0.475 (0.503)	0.525 (0.504)	$p = 0.589$	0.500 (0.504)	$p = 0.786$	0.419 (0.497)	$p = 0.537$	0.484 (0.504)	$p = 0.926$	$p = 0.828$
Number of trainees	61	61		62		62		62		62

Notes: Table reports means and standard deviations (in parentheses) for the five experimental conditions, as well as tests of equality for each control-treatment pair. Baseline test score is the score (0-100) on a baseline test administered at the beginning of the training program to measure trainees pre-existing knowledge of the materials taught during the training program. English test score is the score (0-100) on a test administered at the beginning of the training program to measure trainees' knowledge of English. Unemployed at time of application is a self-reported variable that equals 1 if the trainee had no full- or part-time employment when they applied for the CHA position. Work experience is a self-reported variable that equals 1 if the trainee has held a paid or unpaid position in the health sector in the past. District used community recruitment equals 1 if trainees come from a district where the main mission of the CHA job was advertised as "serving the community," it equals 0 if trainees come from a district where the main mission of the CHA job was advertised as "fostering your career."

To evaluate treatment effects we estimate the following model using panel data at the trainee-course level:

$$score_{ic} = \alpha + \sum_{t=1}^4 \beta_t T_i^t + \gamma B_{ic} + X_i \delta + l_{ic} \eta + \varepsilon_{ic} \quad (2.3.1)$$

where  $score_{ic}$  is trainee  $i$ 's test score in course  $c$  normalized by the mean and standard deviation of test scores for the same course in the control group; treatment effects are thus measured in standard deviation units.  $B_{ic}$  is trainee  $i$ 's score in the baseline exam content for course  $c$  that was administered at the start of the training program. The difference between the “post” exam score and the baseline score is used to measure the value added that is reported in the treatment letters and to rank trainees for the “most improved” awards.  $X_i$  are individual characteristics that include all stratification variables as well as age and trainee  $i$ 's test score for an English language test.  $l_{ic}$  are teacher-specific traits such as teacher ratings (as reported anonymously by the trainees) and expertise in the subject matter of course  $c$ . Since all trainees take the same courses at the same time, different trainees have different teachers for the same course. The teacher rotation schedule was determined by the principal and by the researchers with the aim of ensuring even coverage of teachers across classes.

Standard errors are clustered at the trainee level, as trainee-specific unobservables in the error term create correlation within trainee. Since courses are of different durations and trainees have more time to exert learning effort the longer the course is, we weight observations by course duration.

The parameters  $\beta_t$  measure the causal effect of treatment  $t$  vis-à-vis the control group under the identifying assumption of no contamination across treatments. Contamination can occur if the response to treatment  $j$  is affected by the knowledge that treatment  $k$  exists. For instance, trainees might respond differently to being given

information on relative rankings if they know that other trainees are also getting employer recognition awards while they are not. To minimize the risk of contamination, we allocate trainees in different treatments to different classes and keep classes together for the duration of the experiment. Trainees were told that other classrooms may receive different types of letters with their their exam scores, as this was the pilot year of the government program and different classrooms were trying different things. Trainees were not told that this was a research experiment, thus mitigating potential experimenter demand effects. Reassuringly, no student ever complained or raised the issue of different treatments for the entire duration of the experiment. Despite these precautions, trainees in the non-award treatments could have come across the award recognition letter and community newsletter given to trainees in other treatments. This, however, could have occurred only after the awards were distributed—that is, after the first test. To provide evidence on the practical relevance of contamination, we estimate equation (2.3.1) using scores from the first exam only, which was taken before trainees could have seen letters given to their colleagues in other treatments.

To separate the different mechanisms through which awards can affect performance, treatments are designed to be cumulative so that the “public social comparisons” treatment (T2) also contains information about relative ranks (T1), and the two awards treatments (T3 and T4) contain the same information as T2. Under the assumption that the effect of each component does not interact with the effect of the others, the net effect of each additional component can be identified by the appropriate linear combination of  $\beta_t$  estimators. For instance, the net effect of employer recognition is given by  $\beta_3 - \beta_2$ . We report all relevant linear combinations at the foot of each table.



### 2.3.2 Average treatment effects

Table 2.2 reports the estimates of equation (2.3.1). Test scores are normalized by the mean and standard deviation of the control group for each test so that treatment effects are measured in standard deviation units. Columns (1) and (2) estimate average treatment effects with and without teacher characteristics (average teacher rating and whether at least one of the teachers had specific expertise in the subject matter).

Three findings are of note. First, trainees in treatments 1 (private social comparison) and 2 (public social comparison) perform significantly worse than trainees in the control group. Estimates from the baseline specification in Column (1) show that giving private information about relative rankings lowers performance by 0.31 standard deviations, while rank information combined with a public list of the top four performers lowers performance by 0.38 standard deviations. By contrast, trainees in treatments 3 and 4, where top performers receive awards either in the form of a letter from the Ministry of Health or a profile in the organization’s newsletter, performed the same as trainees in the control group.

Importantly, since trainees can do very little other than studying during the program, we can rule out that the drop in exam performance is associated with an increase of effort devoted to other tasks. In addition, in Ashraf et al. (2015) we find that low performance on test scores during training is highly predictive of low performance in the field in CHA work. This strongly suggests that, even if there were other tasks, there is no compensating effort, either on other tasks during training or in later job performance that can make up for negative performance during the training period.

Table 2.2: Average treatment effects

	(1)	(2)	(3)	(4)	(5)
	Heterogeneous effects by exam period				
	Baseline	Teacher controls	First exam	Following exams	Test of equality
T1: Private Social Comparison	-0.308*	-0.351**	-0.287*	-0.322*	$p = 0.844$
	(0.161)	(0.171)	(0.168)	(0.194)	
T2: T1 + Public Social Comparison	-0.379***	-0.409***	-0.456***	-0.324**	$p = 0.378$
	(0.145)	(0.155)	(0.172)	(0.156)	
T3: T2 + Employer Recognition Award	0.005	-0.044	0.128	-0.080	$p = 0.123$
	(0.133)	(0.174)	(0.165)	(0.136)	
T4: T2 + Social Visibility Award	0.064	0.122	-0.140	0.112	$p = 0.087$
	(0.135)	(0.148)	(0.165)	(0.143)	
Net effect of Public Social Comparison (T2-T1)	-0.071	-0.058	-0.170	-0.003	$p = 0.381$
	(0.166)	(0.164)	(0.174)	(0.201)	
Net effect of Employer Recognition Award (T3-T2)	0.384***	0.365**	0.585***	0.244*	$p = 0.026$
	(0.141)	(0.161)	(0.175)	(0.148)	
Net effect of Social Visibility Award (T4-T2)	0.443***	0.531***	0.316*	0.436***	$p = 0.458$
	(0.141)	(0.151)	(0.170)	(0.153)	
Trainee controls	yes	yes		yes	
Teacher controls	no	yes		no	
Number of clusters (trainees)	307	307		307	
Number of observations (trainee-courses)	1149	850		1149	
Adjusted R-squared	0.213	0.212		0.224	

*Notes:* Table reports OLS estimates, weighted by course duration, with standard errors clustered at the trainee level in parentheses. Dependent variable is normalized exam score, normalized by the mean and standard deviation of the control group for each exam. Trainee controls include: score in the baseline test for each of the four courses, English test score, gender, age, previous experience in the health sector, employment status at the time of application, district recruitment strategy. Teacher controls include: average teacher rating and whether at least one of the teachers had specific expertise in the subject matter. Both variables are defined at the course-treatment level. Columns (2) and (3) report coefficients estimated in the same regression where we include all treatments interacted with an indicator variable that takes value 0 in the first period and 1 thereafter.

\*\*\* Significant at the 1 percent level.

\*\* Significant at the 5 percent level.

\* Significant at the 10 percent level.

Second, we can identify the net effect of each additional treatment component by differencing out common elements under the independence assumption discussed above. This exercise reveals that the difference between T2 and T1 is small and not statistically significant, suggesting that making the list of top performers public does not motivate individuals to work harder. This might be driven by the possibility that other trainees are not the natural peer group to whom these individuals compare themselves. Alternatively, given that in T1, individual ranks were printed on a letter, top performers in T1 could have easily made themselves known even if the list was not public.

Similarly, we can identify the effect of employer recognition and social visibility by differencing out the effect of providing rank information. Column (1) shows that both mechanisms have a strong positive effect on performance. The difference between T3 and T2 is 0.38 standard deviations, whereas the difference between T4 and T2 is 0.44 standard deviations. Both effects are precisely estimated.

Third, the estimated treatment effects are not sensitive to the inclusion of teacher controls (i.e., average teacher rating in students' evaluations and whether at least one of the teachers had specific expertise in the subject matter). This allays the concern that differences among treatments could be due to correlated unobservables at the class-course level. Since including teacher controls reduces the sample as these were not collected for course 2, we use the specification without teacher controls in what follows.

Taken together, the findings indicate that the three mechanisms described at the outset of this paper are relevant in this setting, but their signs and magnitudes differ: social comparisons (whether private or public) weaken motivation whereas employer

recognition and social visibility strengthen it.<sup>14</sup> Thus, depending on the size of these opposing effects in a given context, the net effect of relative performance-based awards is uncertain.

### 2.3.3 Timing

Since treatments are announced at the beginning of course 1, we are able to identify whether their effect differs before and after the first round of letters containing rank information and awards are handed out. This test can help us shed light on why individuals change their behavior in response to rank information. For instance, if individuals are uncertain about the production function of test scores and use rank information to update on the marginal return of their effort, we would expect them to respond differentially before and after the information is provided. In contrast, if individuals know the production function and intrinsically care about their performance rank, we would expect their responses to be the same before and after.

To test whether responses vary before and after rank information is first disclosed, we estimate equation (2.3.1) augmented by the interactions of all treatment indicators with an indicator variable that takes value 0 for the first test (i.e. before treatment letters were distributed) and 1 for the following tests. Columns (3) and (4) of Table 2.2 report the estimated treatment effects on performance in the first test and performance in the following tests, respectively. Column (5) reports the p-value of the null hypothesis that these are equal.

---

<sup>14</sup>It could be argued that social visibility (in our case, visibility in one's home community) is a more powerful motivator in this setting; indeed, while the estimated effects are similar, the underlying treatment strength is different, as all top four performers receive the employer recognition award, whereas only one of them is randomly selected for the social visibility award.

The findings show that social comparisons, whether private or public, lead to a similar reduction in performance in the first and subsequent tests. Thus, this effect cannot be driven by “demotivation” in the traditional sense, or updating on the marginal return to effort, since trainees’ performance dropped *in anticipation* of receiving rank information. Likewise, the findings are at odds with the assumption that trainees have competitive preferences (Charness & Grosskopf, 2001; Freeman & Gelber, 2010); otherwise, we should have observed an *increase* in effort in anticipation of receiving rank information. In Section III, we describe a model of optimal expectations that is consistent with this pattern, in which trainees choose low effort (i.e., self-sabotage) to avoid rank information that delivers a signal about their relative ability.

In addition, columns (3) and (4) cast doubt on the relevance of the concern that responses to treatments 1 and 2 might be contaminated by the awareness of the other two treatments. Indeed, we show that providing rank information with or without a public list of top performers reduces performance by the same amount even before trainees in these treatments were likely to have become aware of the other treatments.<sup>15</sup>

The findings also show that both the employer recognition award and the social visibility award are effective at increasing performance conditional on rank information both before and after the first exam. The effect of the recognition award becomes weaker after the first exam, possibly because individuals revise their chance of winning downwards once the first round of rank information is revealed. We do not,

---

<sup>15</sup>It was of course possible for trainees in the award treatments to tell their colleagues in other treatments before the first exam, and before the awards were distributed. However, these claims should have been more credible and hence contamination stronger after the awards letters became potentially visible to all. The fact that treatment effects are stable throughout courses casts doubt on this.

however, find a similar pattern with respect to the social visibility award. We will return to this below when we allow responses to differ by baseline ability levels.

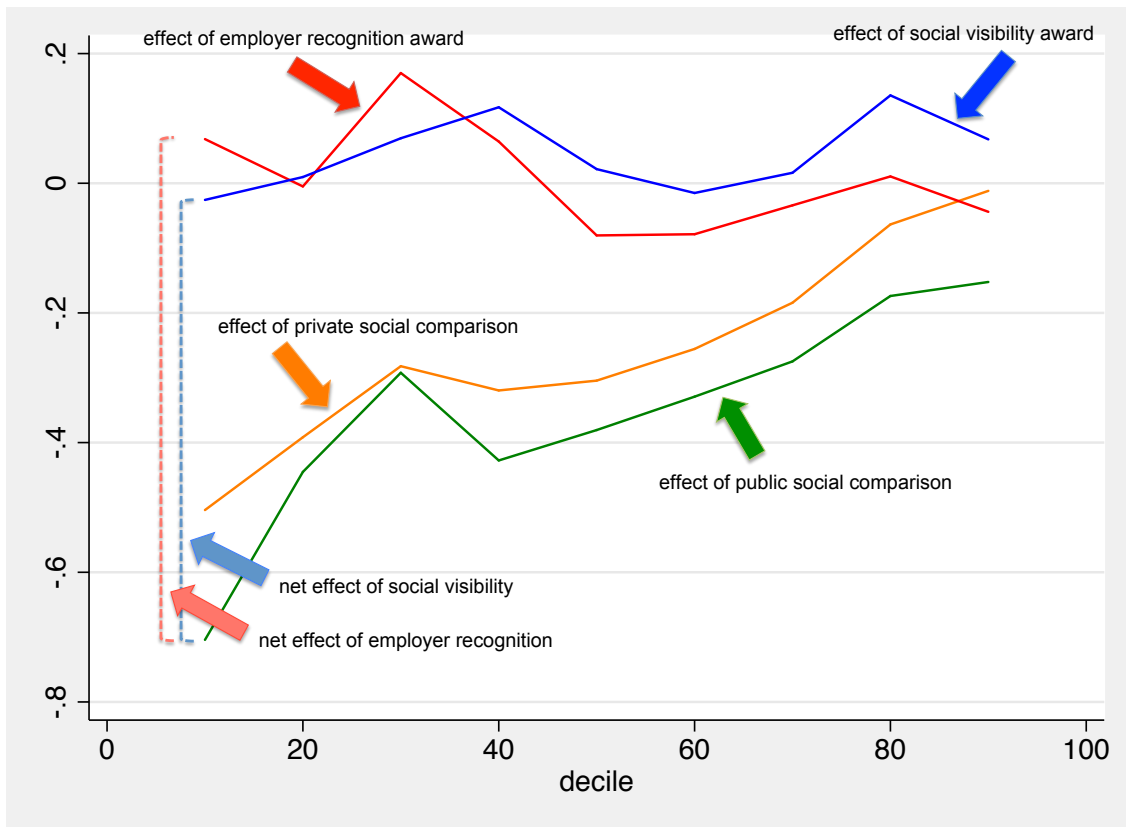
### 2.3.4 Distributional effects

Awards are likely to affect individuals at different points in the performance distribution differently, as the incentive power of awards should be stronger for those who have a greater chance of winning the award—in our case, more able trainees and those who have more potential to improve. Similarly, the effect of social comparison is likely to depend on whether, given their knowledge of their own ability and expectation about others', individuals expects to be ranked high or low.

To provide evidence on these distributional issues we estimate quantile treatment effects at each decile. Figure 2.2 shows this graphically, and Table B.1 reports the regression coefficients by decile. The estimated treatment effect at each decile is the difference in conditional test score between two statistical trainees—one in the treatment group and one in the control group—both positioned at the same decile of the distribution of test scores within her group. Figure 2.2 shows that both the negative effect of relative rankings and the positive effects of recognition and visibility are stronger on the left tail of the conditional test score distribution, and they gradually diminish to zero at the top two deciles. Standard errors reported in Table B.1 show that all effects are statistically different from zero until the seventh decile. Taken together, these findings indicate that information on relative ranks, with or without a public component, increases the dispersion of performance by reducing performance on the left tail.

To provide further evidence on this issue, Table 2.3 allows treatments to have heterogeneous effects by trainees' ability, measured by their score in the baseline

Figure 2.2: Quantile treatment effects



Notes: Each line connects treatment effects of each treatment estimated at each decile. Point estimates and standard errors are reported in Table A1. Individual controls include: score in the baseline test for each of the four courses, English test score, gender, age, previous experience in the health sector, employment status at the time of application, district recruitment strategy.

exam. The estimates show that both the negative effect of relative rankings and the positive effects of recognition and visibility are stronger for low-ability trainees and zero for high-ability trainees. The fact that the negative effect of relative rankings is stronger for the weakest trainees is intuitive, as these are more likely to receive a negative signal about their skills. That the positive effects of recognition and visibility are also stronger for the weakest trainees is presumably due to the fact that, since awards are given to trainees with the highest value added (the “most improved”), those who start at the bottom have a better chance to win. That the effects are zero for the top two deciles is consistent with the fact that scores are capped, such that the very best trainees have little room for improvement.<sup>16</sup>

Finally, Table B.2 allows treatments to have heterogeneous effects by trainees’ ability and exam timing. This confirms that most effects are driven by trainees at the bottom of the ability distribution and that, as discussed above, rank information leads to a similar reduction in performance in the first and subsequent tests. The results also show some evidence that the positive effect of awards becomes weaker after the first exam for trainees in the bottom tercile while it becomes stronger for those in the middle tercile. These findings should, however, be taken with caution as samples are small and tests have low power.

---

<sup>16</sup>Across the entire sample, mean absolute post-test score in the highest decile is 90.0% and in the second-highest decile is 86.5%. The single highest absolute score across all exams was 96.1%. As our team, in consultation with medical training experts, wrote the exams to ensure quality and precision, we purposely included extremely difficult questions making scores of 100% very difficult. We thus take the evidence on top decile scores as support of capping, although we cannot rule out that other effects (such as a lack of desire to be singled out among peers, due to being taxed by expectations of assistance) may have also been occurring at the top decile.



Table 2.3: Heterogeneous treatment effects by baseline test score

	(1)	(2)	(3)
	First tercile of baseline test score	Second tercile of baseline test score	Third tercile of baseline test score
T1: Private Social Comparison	-0.672** (0.298)	-0.233 (0.271)	0.010 (0.213)
T2: T1 + Public Social Comparison	-0.812*** (0.260)	-0.179 (0.199)	-0.113 (0.203)
T3: T2+ Employer Recognition Award	-0.280 (0.216)	0.157 (0.250)	-0.005 (0.159)
T4: T2+Social Visibility Award	-0.102 (0.309)	0.204 (0.223)	-0.031 (0.177)
Net effect of Public Social Comparison (T2-T1)	-0.140 (0.334)	0.054 (0.243)	-0.153 (0.214)
Net effect of Employer Recognition Award (T3-T2)	0.532** (0.255)	0.336 (0.214)	0.107 (0.166)
Net effect of Social Visibility Award (T4-T2)	0.710** (0.342)	0.383** (0.183)	0.082 (0.173)
Trainee controls	yes	yes	yes
Number of clusters (trainees)	92	107	107
Number of observations (trainee-courses)	350	401	398
Adjusted R-squared	0.186	0.079	0.157

*Notes:* Table reports OLS estimates, weighted by course duration, with standard errors clustered at the trainee level in parentheses. Dependent variable is normalized exam score, normalized by the mean and standard deviation of the control group for each exam. Individual controls include: score in the baseline test for each of the four courses, English test score, gender, age, previous experience in the health sector, employment status at the time of application, district recruitment strategy.

\*\*\* Significant at the 1 percent level.

\*\* Significant at the 5 percent level.

\* Significant at the 10 percent level.

## 2.4 Interpretation

Our results suggest that the prospect of receiving information about one’s rank in the distribution makes trainees exert lower effort, whereas the possibility of receiving employer recognition or improving one’s visibility in the community makes them exert higher effort. In this section, we attempt to interpret these findings in a unified utility maximization framework.

The fact that individuals put in more effort when it is announced that doing so might earn them an award is intuitive. The main challenge is to explain why individuals, especially those with low ability, reduce effort in the rank information treatments even before being told their rank. A growing literature models the possibility that information lowers utility leading to information avoidance, at the cost of taking worse decisions (Oster et al., 2013; Brunnermeier & Parker, 2005; Köszegi, 2006; Stone, 2004). In these models, agents choose their beliefs optimally to maximize their lifetime utility, including an interim period of anticipatory utility arising from the belief. This implies that individuals may take actions to avoid a precise signal that has the potential to threaten their belief, for instance by choosing not to take a medical test.

In our setting, all agents know their individual baseline test scores and are told their individual exam scores in all treatment groups (including control). This, combined with the fact that individuals have a relatively good sense of their own individual ability as they all enter as adults, allows us to assume all that all individuals know their own ability. What individuals do not know is the ability of others—a reasonable assumption in our setting given that trainees entered the training school from an extremely diverse set of geographic and skill backgrounds. Realistically, individuals have a noisy expectation over their peers’ ability.

In this setting, rank information (as provided in all treatment groups but not in control) increases the precision of the estimate of each individual's expectation over his peers' average ability. This would not affect behavior in standard models where agents only care about the expected value, but it might matter if agents prefer a fuzzy signal which can support optimistic beliefs about their relative ability.

We assume that agents are risk neutral and individual  $i$  with ability level  $a_i$  chooses effort  $e_i$  to maximize his expected utility:

$$\beta_i^t t(a_i, e_i) + \beta_i^{sc} (a_i - a_{-i}^*) (1 + T_i^{sc} \sigma(e_i)) + \beta_i^a T_i^A p(a_i, e_i, b_i) A - d(e_i) \quad (2.4.1)$$

The first term captures the effect of effort on learning proxied by test scores, which might provide utility either directly as individuals care about learning or through future wages.  $\beta_i^t > 0$  is the weight individual  $i$  puts on learning, and learning is a function of effort and ability with  $t_e > 0$ ,  $t_a > 0$ ,  $t_{ee} < 0$ . We assume that individuals know their own ability, as, again, individuals know their baseline scores and their absolute test scores in control and treatment alike.

The second term captures the utility deriving from social comparisons to which individual  $i$  gives weight  $\beta_i^{sc} > 0$ . Social comparisons enter additively as in Kandel & Lazear (1992) and we assume that individuals care about being of higher ability than their peers rather than having higher test scores *per se*. We assume that individuals enjoy social comparisons when their own ability is higher than the average of their peers  $a_{-i}^*$ , and the effect is larger the larger is the ability gap. Conversely, individuals suffer from social comparisons when their own ability is lower than the average of their peers, and the effect is larger the larger is the ability gap.

Finally, as discussed above, we assume that when individuals are given information on test score ranks, they can obtain more precise information about their relative ability. Thus,  $T_i^{sc} = 1$  if individual  $i$  is in one of the four treatment groups that provide rank information on test scores and 0 in control. In particular, we assume that whether test score ranks give information on ability ranks depends positively on effort through the “signal” function  $\sigma(e_i)$  with  $\sigma_e > 0$ . Intuitively, since test scores are a function of ability and effort, receiving a low rank when exerting low effort could still allow an individual to retain the option of believing that he is of truly high relative ability, but receiving a low rank with high effort could not.<sup>17</sup> The “signal” function  $\sigma(e_i)$  is a reduced form representation of “choosing beliefs” in the spirit of Yariv (2002), Köszegi (2006), Oster et al. (2013) and Brunnermeier & Parker (2005).<sup>18</sup>

The third term captures the utility deriving from award  $A$ , which in our setting is either employer’s recognition or social visibility. Thus,  $T_i^A = 1$  if individual  $i$  is in one of the two treatment groups that provide awards, and 0 otherwise. Note that  $T_i^A = 1 \Rightarrow T_i^{sc} = 1$  but not vice versa. The probability of winning the award  $p(a_i, e_i, b_i)$  depends on effort, ability, and the baseline score  $b_i$ , with  $p_e > 0$ ,  $p_{ee} < 0$ ,  $p_a > 0$ , and  $p_b < 0$ . The latter captures that “most improved” awards are more easily obtainable by those who did poorly in the baseline test.

Finally,  $d(e)$  is the disutility of effort, with  $d_e > 0$ ,  $d_{ee} > 0$  as is standard.

Maximizing equation (2.4.1) with respect to  $e_i$  yields:

---

<sup>17</sup>This holds regardless of whether  $i$  expects others to exert high or low effort. If he chooses high effort and expects others to do so as well, a low rank in test scores implies a low rank in ability. If he expects others to choose low effort, a low rank in test scores implies a low rank in ability *a fortiori*.

<sup>18</sup>Similarly, Benabou and Tirole (2002) consider a model where agents can manipulate their interim belief through the choice of information structure. They show that less information can be preferable as it can weaken the time-inconsistency problem and induce more effort in the future. We do not assume time-inconsistency in our setting, which would exacerbate our effect as it would further discount the future impacts of taking distorted actions for interim belief utility.

$$\left\{ \begin{array}{ll} e_i^C \quad s.t. & \beta_i^t t_e(a_i, e_i^C) - d_e(e_i^C) = 0 \quad \text{if } T_i^{SC} = T_i^A = 0 \\ e_i^{SC} \quad s.t. & \beta_i^t t_e(a_i, e_i^{SC}) + \beta_i^{sc} \sigma_e(e_i^{SC})(a_i - a_{-i}^*) - d_e(e_i^{SC}) = 0 \quad \text{if } T_i^{SC} = 1, T_i^A = 0 \\ e_i^A \quad s.t. & \beta_i^t t_e(a_i, e_i^A) + \beta_i^{sc} \sigma_e(e_i^A)(a_i - a_{-i}^*) + \beta_i^a p_e(a_i, e_i^A b_i) A - d_e(e_i^A) = 0 \quad \text{if } T_i^{SC} = T_i^A = 1 \end{array} \right.$$

The first order conditions for effort in the three cases inform the comparison of performance in each treatment vs. the control group and across treatments. These comparisons map into the empirical findings as follows.

First,  $e_i^C > e_i^{SC}$  iff  $a_i < a_{-i}^*$ ; namely, providing rank information reduces effort for individuals whose ability is lower than their expectation of their peers' ability. Intuitively, all other things equal, effort provides them with something undesirable (a precise private signal of their ranking in the class). This is related to the psychology literature on “self-handicapping,” which can take the form of withdrawing effort in performance settings where there is potential for self-image-damaging feedback (Jones & Berglas, 1978; Berglas, 1985).<sup>19</sup> In line with this, a qualitative survey we administered before implementing the treatments reveals that 43% of the trainees in the bottom quartile of baseline scores did not want to know their relative rank in the class, while only 24% of those in the top quartile said the same.

These findings are similar to Oster et al. (2013), in which people at risk for Huntington Disease prefer not to undergo a test to learn whether they have the genetic mutation that causes the disease because the anticipatory utility of believing they might not get sick outweighs the costs of potentially distorted actions.

<sup>19</sup>This allows the trainee to manipulate the attribution of failure to himself (Kelley, 1971). If failure occurs, the extent to which the outcome is attributed to his lack of ability is discounted because of the equally likely reason for failure: decreased effort. However, if the trainee obtains a high grade on the exam, the attribution to ability is strengthened because the decreased learning effort made it even more difficult to do well.

Second  $e_i^A > e_i^{SC}$  for all  $i$ ; namely, providing awards in addition to rank information can only increase effort as long as awards are valuable, since  $\beta_i^a p_e(a_i, e_i^A, b_i)A > 0$ . The strength of the effect depends on ability and baseline scores, both of which determine the marginal return to effort  $p_e(a_i, e_i^A, b_i)$ .

Third,  $e_i^A > e_i^C$  for all  $i$  such that  $a_i \geq a_{-i}^*$ , as both the second (rank information) and third (award) term in the first order condition are weakly positive, thus increasing the marginal return to effort. In contrast,  $e_i^A \begin{matrix} \leq \\ > \end{matrix} e_i^C$  for all  $i$  such that  $a_i < a_{-i}^*$  since the second (rank information) term is negative and the third (award) term is positive. Thus, providing awards might reduce effort compared to the control group that receives no awards or rank information if the response to rank information is stronger than the response to the award itself.

## 2.5 Conclusion

Our results suggest that awards can have a negative effect on performance as they facilitate social comparison, even though they have a positive effect through employer recognition and enhanced social visibility. In our context, the negative effect of rank information on learning was large enough to undo the positive effects of awards. Since learning is directly related to future labor productivity, and test scores are significantly predictive of future performance, the distortion in effort is practically significant, and detrimental to the objective of increasing overall learning and later on-the-job productivity.

While the relative magnitudes of these effects are likely to be context-specific, the possible negative effects of rank information and social comparisons are important to consider in the overall productivity effects and design of performance awards. In particular, we show that individuals may even reduce effort in anticipation of

learning rank information. This is consistent with a model in which individuals have preferences over their self-perceived ability ranking, and thus engage in information avoidance to be able to retain a positive view of themselves. In effect, they self-sabotage in order to avoid the signal contained in ranking information so as not to have to update their beliefs about their own relative ability. As the risk that the signal of low rank is greater in lower ability ranges, this self-handicapping is worse at the bottom of the distribution.

From an employer or policy maker perspective, the cost of an incentive structure that differentially affects the lower tail depends on the nature of the production function. It is particularly costly when there are complementarities in production or when the performance of the lower tail is critical to the principal's goal, as in our setting where the government wants to ensure equitable provision of health services to remote rural areas. In domains such as innovation in science and finance, the effect on productivity of the upper tail of the distribution might be most important. But in domains such as health services delivery where the potential for harm is high, it is critical to employ incentives that are not detrimental to the lower tail of the distribution. This depends in part, as well, on the ease of exit and entry; if the goal is to induce the lower tail to withdraw effort, and potentially exit, then providing rank information could be a highly effective means to do so. It is left to future research to unbundle these effects across cultures and professional sectors.

# Chapter 3

## Intrinsic Incentives: A Field Experiment on Leveraging Intrinsic Motivation in Public Service Delivery<sup>1</sup>

### 3.1 Introduction

Public services—governance, education, health care, national defense—rely on agents for their provision. In developing countries, the effort of these agents is often a binding constraint, prevailing over other factors such as the agents’ ability and market demand (Das & Hammer, 2007, 2014; Leonard et al., 2013; Maestad et al., 2010). The standard

<sup>1</sup>I thank the Chief Medical Office of Kaushambi District, Dr. Nandini Sharma/Maulana Azad Medical College, and Dimagi, Inc. for hosting this research; Brian DeRenzi, Andrew Ellner, and Neal Lesh for ongoing collaboration as co-Investigators; and Sapana Gandhi, Sangya Kaphle, Sugandha Nagpal, Robert Racadio, and Jeremy Wacksman for excellent research assistance. I am grateful to Nava Ashraf, Oriana Bandiera, Iqbal Dhaliwal, Paul Farmer, Rema Hanna, Michael Kremer, Matthew Rabin, Andrew Weiss, and seminar participants at Harvard Business School, Harvard Department of Economics, and Harvard Medical School for helpful comments. Generous financial support has been provided by the Massachusetts General Hospital Consortium for Affordable Medical Technologies, Child Relief International, and the Harvard Business School Doctoral Program.



agency model offers both an explanation and a solution: agents dislike effort, and they can be persuaded with incentives to exert it. In line with this view, recent field experiments have shown that monetary and non-monetary incentives can improve the performance of agents engaged in public service delivery (Ashraf et al., 2014; Basinga et al., 2011; Duflo et al., 2012; Miller et al., 2012; Muralidharan & Sundararaman, 2011).<sup>2</sup>

What this perspective neglects, however, is the converse scenario: agents who exert effort despite having little extrinsic incentive to do so. In settings in which forty percent of health workers are absent from their posts on any given day (Chaudhury et al., 2006), what explains the presence of the remaining sixty percent? While other extrinsic factors (e.g., monitoring, social pressure, status-seeking) likely contribute, the persistence of effort in the face of weak incentives—and the decision to select into pro-social jobs in the first place—suggests a role for intrinsic motivation.<sup>3</sup>

That agents can be intrinsically motivated is not new in economics (Bénabou et al., 2003; Fehr et al., 1999). But this motivation is typically taken as given—a fixed trait that organizations may wish to select for but cannot influence (Besley & Ghatak, 2005). If it can be influenced, it is only for the worse, as proposed by theories of motivational crowd-out (Deci et al., 1999; Bénabou & Tirole, 2006).<sup>4</sup> In contrast, the business literature has long posited that managers can leverage intrinsic motivation by

---

<sup>2</sup>For evidence challenging the effectiveness of financial incentives in public service delivery, see Banerjee et al. (2008), Glewwe et al. (2010), and, in the US context, Fryer (2013). For theoretical contributions on why financial incentives in public service delivery may fail, see Bénabou et al. (2003); Bénabou & Tirole (2006).

<sup>3</sup>For empirical evidence that supports the hypothesis that those with pro-social preferences select into pro-social jobs, see Kolstad & Lindkvist (2013) and Lagarde & Blaauw (2013).

<sup>4</sup>A small but notable exception is theoretical and empirical work on the role of delegation and empowerment in enhancing intrinsic motivation. See, e.g., Aghion & Tirole (1997); Rasul & Rogger (2015).

manipulating job attributes such as autonomy (Deci & Ryan, 1985), purpose (Weick, 1995), and organizational culture (Schein, 1985).<sup>5</sup> But these attributes are generally conceptualized as static elements of job design, as opposed to incentives that interact with heterogeneous preferences.

To provide a richer view into the relationship between intrinsic incentives and intrinsic motivation, this paper explores the potential role of an intrinsic incentive technology in enhancing the effort of agents engaged in public service delivery. I define an intrinsic incentive as any variable in the principal’s choice set that modifies the agent’s marginal intrinsic utility of effort, analogous to how an extrinsic incentive modifies the marginal extrinsic utility of effort. In the setting of a rural health worker program in India, I develop a novel mobile phone technology—a “self-tracking” app—designed to act as an intrinsic incentive by delivering information that makes effort more intrinsically rewarding.<sup>6</sup> The app comprises a set of graphs that a health worker can access on her phone to view her performance with respect to the job’s primary task: visiting pregnant women in their homes. As a counterfactual, I develop an analogous app—a “generic encouragement” app—designed to be lower-powered in that it provides generic messages of encouragement that are independent of the agent’s effort.<sup>7</sup>

---

<sup>5</sup>For more recent evidence from laboratory experiments, see Grant (2007) and Ariely et al. (2008).

<sup>6</sup>The technology is intended to leverage intrinsic motivation both in the sense used in the psychological literature (motivation to do a task for its own sake—effort as its own reward) and in the sense of prosocial preferences (motivation to do a task due to its positive social externalities—impact as its own reward), both of which are distinguishable from extrinsic motivation in that no benefit external to the task and its output is needed to justify effort. Throughout this paper, unless otherwise stated, I use “intrinsic” motivation to encompass both intrinsic and prosocial preferences.

<sup>7</sup>The conceptual distinction between the two lies in the slope with which they are expected to modify agents’ marginal intrinsic utility of effort, analogous to how a piece rate of  $x$  dollars per piece is higher-powered than one that pays  $\frac{x}{2}$  dollars per piece.

I test these incentives head-to-head by randomly assigning each of 145 health workers to receive one of the two apps on their work phone, and then tracking both app usage and performance on a daily basis for ten months. The experiment yields four main findings. First, both intrinsic incentive technologies are demanded. Across the two treatments, despite receiving minimal encouragement and no directive to do so, the average health worker accesses the software application approximately once every three days. Second, turning to effects on effort, compared to the generic encouragement app, the self-tracking app leads to a 27% increase in performance as measured by home visits. Third, the self-tracking app is most effective when it leverages pre-existing intrinsic motivation; it produces a 46% increase in performance in the top tercile of intrinsically motivated workers, but no improvement in the bottom tercile, indicating that, in this setting, intrinsic incentives and intrinsic motivation are complements. Fourth, despite these gains in client visits, no aggregate impact on the health of the pregnant clients and their children is observed. Finally, in terms of mechanisms, supplementary evidence suggests that the treatment effects of the self-tracking app are not mediated by effects on the production function or on extrinsic preferences. Rather, they appear to increase effort by making effort more intrinsically rewarding.

This paper contributes to two nascent literatures. First, recent empirical work has evaluated information incentives as a tool for motivating prosocial behavior. In particular, various forms of relative performance feedback have been found to be effective in improving home energy conservation (Allcott, 2011), student learning (Tran & Zeckhauser, 2012; Azmat & Iriberry, 2010), and physician quality (Kolstad, 2013). Information incentives, however, can function as extrinsic or intrinsic rewards, and when the targeted task (such as energy conservation or school performance) confers financial benefits, an information incentive that increases effort in the task is

likely to be operating at least in part via an extrinsic channel.<sup>8</sup> In contrast, the self-tracking app studied in this experiment is designed to increase only intrinsic returns to effort. In this regard, this *intrinsic* information incentive conceptually shares more in common with other intrinsic rewards, such as task meaning and organizational mission, than with other information incentives such as performance feedback.

The second literature examines the interaction between incentives and psychological traits such as intelligence and personality in prosocial behavior, in the context, for example, of civil servants (Dal Bó et al., 2013), health agents (Ashraf et al., 2014, 2015), and taxpayers (Dwenger et al., 2014). In the psychologically richest of the analyses, Callen et al. (2015) find that personality traits predict job performance among health officials in Pakistan, and that the experimental response to a novel monitoring technology varies with these personality traits. I extend this approach by (a) testing an intrinsic rather than extrinsic incentive, (b) testing a specific, theoretically guided interaction—that between intrinsic incentives and intrinsic motivation—and (c) elucidating psychological mechanisms.

The paper is organized as follows. Section 3.2 presents a simple principal-agent framework in which the principal can offer extrinsic and intrinsic incentives, and agents have extrinsic and intrinsic preferences. Section 3.3 describes the policy and program context. Sections 3.4 and 3.5 present the experimental design and results, respectively. Section 3.6 concludes.

---

<sup>8</sup>Kolstad (2013) finds that cardiac surgeons respond to physician report cards in ways that cannot be explained solely by profit maximization. He defines this reduced-form residual as an “intrinsic incentive” effect, but is silent about what in the physicians’ utility function drives this effect. As such, the results are consistent with physicians behind motivated not only intrinsically, but also by non-financial extrinsic preferences such as those for prestige, recognition, and career promotion.

## 3.2 Framework

In this section, I develop a simple framework for analyzing the potential role of intrinsic incentives in public service delivery. I take extrinsic incentives as given and instead focus on the principal's ability to stimulate effort by increasing intrinsic—i.e., direct hedonic—returns to effort. The framework assumes that an agent may find an effort task tasteful or distasteful for intrinsic reasons. Modifying or augmenting some aspect of the task may make it more tasteful, thereby increasing its marginal intrinsic returns and, by extension, equilibrium effort. The possibility that the principal may be able to control this modification paves the way for intrinsic incentives. I show that, like an extrinsic incentive, an intrinsic incentive can be high- or low-powered, and the incentive interacts with the agent's intrinsic motivation to determine the overall impact on marginal intrinsic utility. This interaction generates simple testable implications that I then take to the experimental data.

### 3.2.1 Principal's choice

Consider a single-principal, single-agent framework. The principal (e.g., a government, nonprofit organization, hospital) is invested in the production of a social good  $Y$  (e.g., population health). The agent can contribute to the production of  $Y$  with effort  $e_i \geq 0$ . Assume that the production function  $Y(e_i)$  is monotonically increasing in  $e_i$ .

To motivate the agent to exert effort, the principal may use extrinsic and intrinsic incentives. The former provides a reward contingent on output, whereas the latter enhances the hedonic returns to effort itself. Denote by  $\alpha \geq 0$  and  $\beta \geq 0$  the extrinsic incentive contract, where  $\alpha$  is a fixed reward and  $\beta$  is paid linearly for each unit of  $Y$

produced.<sup>9</sup> Because  $\alpha$  does not affect equilibrium effort, for notational convenience, assume that  $\alpha = 0$ .

Denote by  $\psi(\cdot)$  an intrinsic incentive function.  $\psi(\cdot)$  will enter into the agent's utility function as described in Section 3.2.2 below. Here, I comment on its conceptual underpinnings. The  $\psi(\cdot)$  function is left purposefully vague, because it is likely to have many parameters. Definitionally,  $\psi(\cdot)$  is a function that alters the marginal intrinsic utility of effort. It is analogous to  $\beta$ , but whereas  $\beta$  is non-negative, the value of  $\psi(\cdot)$  is unbounded—i.e., it can increase or decrease marginal intrinsic returns. Theories of motivational crowd-out (Deci et al., 1999), for example, imply  $\frac{\partial \psi}{\partial \beta} < 0$ —i.e., an increase in extrinsic returns to effort reduces its intrinsic returns.<sup>10</sup>

Potentially any job attribute could affect marginal intrinsic utility of effort: the nature of the effort task, the technology of production, the degree of monitoring vs. autonomy, organizational norms and culture, and so forth. For the current purposes, let the principal's choice variable in the  $\psi(\cdot)$  function be the psychological salience and observability to the agent of the effort task and its social impact. The principal alters the agent's information environment to achieve this effect. Much more is described about this information technology in Section 3.4.1. The basic concept is that the principal can provide a technology by which the agent is better able to self-observe effort and output and, in so doing, experience greater marginal utility (or disutility) of effort.

---

<sup>9</sup>The conventional interpretation for  $\beta$  would be a variable wage, but the conceptual intuition can be extended to other extrinsic goods such as social status and job security.

<sup>10</sup>I abstract away from this relationship in this framework because it is not empirically relevant in this experiment; as I describe in Section 3.4, there is no variation in  $\beta$  in the sample and hence no way to identify  $\psi(\beta)$ .

The principal chooses intrinsic incentive regime  $j \in \{h, l\}$ , corresponding to “high-powered” and “low-powered,” respectively. The high-powered (low-powered) regime enables the agent to access information that makes the effort task and its social impact highly (minimally) salient. Assume that  $j$  maps onto  $\psi(j)$  such that  $\psi_h = 1$  and  $\psi_l = 0$ . Assume also that providing  $j$  carries zero marginal cost, in terms of both the direct supply cost and the shadow cost of reputation.<sup>11</sup> Thus, the only cost incurred by the principal is  $\beta Y$ , but, as is made explicit below,  $Y$  is affected by  $\psi(j)$  via the latter term’s effect on optimal effort.

Let the principal’s utility be  $V(Y - \beta Y)$ , where, given risk neutrality,  $V' > 0$  and  $V'' = 0$ . The principal chooses  $\{e, \beta, j\}$  to maximize her expected utility, subject to the agents’ individual rationality and incentive compatibility constraints.

### 3.2.2 Agent’s choice

I now turn to the agent’s choice of  $e$ . A risk-neutral agent  $i$  with preferences  $(\theta_i^E, \theta_i^I)$  chooses effort  $e_i$  to maximize his expected utility:

$$U_{ij} = \theta_i^E[\beta Y(e_i)] + \theta_i^I[\gamma_i + (1 + \psi_j)e_i] - \frac{e_i^2}{2} \quad (3.2.1)$$

Equation 3.2.1 has three terms: an extrinsic payoff term, an intrinsic payoff term, and an effort cost function. The two payoff functions are weighted by extrinsic  $(\theta_i^E)$  and intrinsic  $(\theta_i^I)$  preference parameters, respectively. Assume  $\theta_i^E$  and  $\theta_i^I$  are independent

<sup>11</sup>Bénabou et al. (2003) describe this shadow cost as follows: “A teacher or a manager who makes very complimentary comments to every pupil or employee may lose her credibility....[W]hen disclosing soft information to several agents the principal must realize that they will see through her ulterior motivation, and believe her only if she builds a reputation for not exaggerating claims.” Why there is no shadow cost of  $j$  in the experimental intervention is explained in more detail in Section 3.4.1, but in brief, because  $j$  only contains objective information, it does not rely on scarcity for its value, whereas recognition, praise, or positive feedback does.

and that  $\theta_i^E$  is distributed over  $[0, 1]$ . In contrast, assume  $\theta_i^I$  is distributed over  $[-1, 1]$ . Similar to the discussion of  $\psi(\cdot)$  above, whereas preferences for extrinsic rewards can logically only be weakly positive, intrinsic preferences can be positive or negative since effort can be intrinsically utility-enhancing or -undermining. The convex effort cost function,  $\frac{e_i^2}{2}$ , captures disutility of effort in a baseline state absent intrinsic preferences, around which the intrinsic payoff term allows for heterogeneity.<sup>12</sup> That is,  $\theta_i^I$  is meant to account for heterogeneity in intrinsic preferences regarding task-specific effort, which can increase or decrease the cost of effort relative to the baseline of no intrinsic preferences.

To return to the utility function, the first term—the extrinsic payoff function—is simply the principal’s linear payment contract weighted by the agent’s preference for extrinsic rewards,  $\theta_i^E$ .

As with the payment contract, the intrinsic payoff term has two components—one that is effort-independent and another that is effort-dependent. The effort-independent parameter,  $\gamma_i$ , can be thought of as an endowment of intrinsic (dis)utility—e.g., the warm glow experienced from having a prosocial job, or the effort-independent disutility of having a job that runs counter to one’s tastes. Since this endowment affects equilibrium utility but not effort choice, for simplicity, assume  $\gamma_i = 0$ . The second component of the intrinsic payoff term,  $[(1 + \psi_j)e_i]$ , captures the practical intuition that, in some circumstances (e.g., volunteering), even when there is no extrinsic payoff ( $\alpha = \beta = 0$ ), agents are still willing to supply  $e_i^* > 0$ , implying that, over the interval  $[0, e_i^*)$ , marginal intrinsic benefit of effort exceeds marginal effort cost.

In the absence of any intrinsic incentive (i.e.,  $\psi = 0$ ), the agent derives marginal

---

<sup>12</sup>For example, an agent with a fear of blood would find being a surgeon particularly distasteful, and, all else equal, surgical effort would be more costly for this person than for someone without this phobia. Likewise, an individual with anti-abortion political views may have an intrinsic aversion to working in an abortion clinic, and so forth.



intrinsic utility  $\theta_i^j$  per unit effort. With the high-powered intrinsic incentive ( $j = h$ ), the marginal intrinsic utility increases in absolute magnitude to  $2\theta_i^j$ . In contrast, marginal intrinsic utility is negative if  $\theta_i^j < 0$  and 0 if  $\theta_i^j = 0$ .

The agent maximizes Equation 3.2.1 with respect to  $e_i$ , such that:

$$e_{ij}^* = \theta_i^F \beta + \theta_i^j (1 + \psi_j) \quad (3.2.2)$$

This simple first-order condition illustrates the complementarity between intrinsic incentives and intrinsic motivation that this experiment tests. It implies that intrinsic incentives will lead to higher effort for intrinsically motivated agents, but may have no effect or even reduce effort for those who are intrinsically unmotivated. Thus, the aggregate effect of an intrinsic incentive depends on the distribution of intrinsic motivation in the agent population.

### 3.3 Context

#### 3.3.1 Program context

In 2005, the Government of India launched the Accredited Social Health Activist (ASHA) program, a nationwide effort to improve health services at the community level, especially in rural regions.<sup>13</sup> ASHAs are female community health workers who are selected by local village councils to provide for the health needs of the villages in which they reside.<sup>14</sup> Each ASHA is assigned a discrete and non-overlapping catchment

<sup>13</sup>As of 2014, 828,000 ASHAs had been recruited across India Government of India (2015).

<sup>14</sup>Job qualifications include: female gender; married, widowed, or divorced status (due to patrilocality); grade 8 education or higher; age between 25 to 45 years; and preferably, literacy.

area, which typically corresponds to a village or group of villages with a population of roughly 1,000 people.

ASHAs' primary responsibilities revolve around supporting and counseling women during pregnancy, childbirth, and the postpartum period.<sup>15</sup> The typical sequence of serving a pregnant client proceeds as follows. When the ASHA learns of a new pregnancy in her village, she visits the woman and offers to support the woman. If the woman accepts, the ASHA registers the client using a phone-based record-keeping tool.<sup>16</sup>

Over the course of the pregnancy, ASHAs are expected to visit the client at home at least monthly. During these visits, ASHAs carry out a variety of tasks: counseling on nutrition, physical activity, and other day-to-day aspects of pregnancy; counseling on identifying pregnancy-related danger signs requiring urgent medical attention; encouraging the client to obtain facility-based antenatal care; providing iron and folic acid supplements; working with the client to develop a birth plan, which includes calculating the estimated delivery date, advising the client on local health facilities for delivery, identifying means of transport, and engaging family support; and updating the client's maternal health card. The ASHA records and submits these follow-up visits also using her phone-based tool.

At the time of labor, the ASHA accompanies the client to a health center or hospital and remains with her throughout labor and delivery; the ASHA's payment, as discussed below, is contingent upon this presence. After the mother is discharged,

---

<sup>15</sup>In 2006, an expert group convened by *The Lancet* identified one "overwhelming priority strategy" for reducing maternal deaths: "promoting delivery in primary-level institutions (health centers), backed up by access to referral-level facilities," as opposed to home delivery (Campbell & Graham, 2006).

<sup>16</sup>The mobile phone tool is not a feature of the national ASHA program but, rather, the program site in which this experiment takes place. See Section 3.3.2.

the ASHA visits the mother and the child several times over the ensuing six weeks to monitor their health and counsel on newborn care, breastfeeding, family planning, and immunizations. At six weeks postpartum, the ASHA “discharges” the client from care.

The ASHA job is typically not a full-time position, and ASHAs do not receive salaries.<sup>17</sup> Instead, they are paid piece rates for discrete activities such as facilitating institutional delivery, assisting with polio immunization campaigns, and mobilizing men and women to undergo sterilization.<sup>18</sup> ASHAs’ chief source of income arises from a federally sponsored conditional cash transfer scheme designed to encourage institutional delivery, called Janani Suraksha Yojana (JSY). In this scheme, pregnant women are paid INR 1,400 (USD 23 at 2015 exchange rates) for delivering in an accredited public or private health facility. In addition, an ASHA who accompanies the woman to the hospital for delivery is paid INR 600 (USD 10). ASHAs are not paid for visits to the client during the antenatal period. Thus, in the absence of intrinsic motivation, antenatal visits are rational only inasmuch as they increase the probability of institutional delivery.

### **3.3.2 Program site**

With a population of 1.6 million, Kaushambi District is one of 19 (out of 70 total) districts in Uttar Pradesh designated by the state government as “high-focus” in view

<sup>17</sup>In the study population in the pre-experimental period, the average ASHA earned USD 372 in total annual ASHA payments. For context, auxiliary nurse-midwives (ANMs), the supervisory cadre directly above ASHAs and the actual providers of facility-based antenatal care services, earn approximately USD 2,000 per year. Anganwadi workers (AWWs)—child health and nutrition workers who are positioned laterally to ASHAs and have similar job qualifications but work full-time—earn approximately USD 880 per year.

<sup>18</sup>For example, if an infant registered by the ASHA completes his or her complete course of routine immunizations, the ASHA receives a payment of INR 150 (USD 2.50).

of its poor development indicators. Its maternal mortality ratio of 442 deaths per 100,000 live births is nearly twice the national average and 30% higher than the state average (Government of India, 2011a); the district's neonatal mortality rate is twice the national average; and the district has the second-highest scheduled caste population share in the state (Government of India, 2011a).<sup>19</sup> This experiment takes place in Muratganj, one of eight sub-districts in Kaushambi, with a population of 193,355 (Government of India, 2011b). At the time of the experiment launch in 2014, Muratganj had 145 ASHAs, all of whom had been recruited and trained in 2006-2007 when the ASHA program was rolled out in the District.

In 2012, a nongovernmental organization established a partnership with the local government to equip the ASHAs with mobile phones to facilitate their work. The phones contain a software application called CommCare, through which the ASHAs register clients and document home visits as described above.<sup>20</sup> These records, which are synchronized with a cloud-based server, provide the data on which the self-tracking app functions. By the time of the experiment launch, the ASHAs had been using CommCare for 15 months. As all of the ASHAs in the experiment were trained to use CommCare, this experiment is not designed to evaluate the underlying work tool.

Table 3.1, Panel A shows descriptive statistics for the performance of the 72 ASHAs in the control (generic encouragement) treatment over the pre- and post-intervention periods. In contrast to the expectation that ASHAs should visit all clients on a monthly basis, the average control ASHA visits 46% of her 13.1 pregnant clients in the average month. Panel B shows descriptive statistics for a limited set

<sup>19</sup>Scheduled castes are castes designated in the Constitution of India as historically disadvantaged. Nationwide, 17% of the Indian population belongs to a scheduled caste; in Kaushambi, the proportion is 36%.

<sup>20</sup>The application is developed by a US-based company called Dimagi, Inc.

of client characteristics and health outcomes that ASHAs report. Of note, the 77% institutional delivery rate is substantially higher than the 22% rate for rural Uttar Pradesh reported in the most recent (2005-2006) wave of the National Family Health Survey (International Institute for Population Sciences & Macro International, 2007); though possibly overstated due to selection (both in registration of pregnant women and in reporting of outcomes), this statistic is consistent with a broad increase in institutional delivery that has been observed across India since the introduction of the JSY conditional cash transfer scheme in 2005 (Lim et al., 2010).

## 3.4 Experimental design

### 3.4.1 Experimental interventions

I create a novel mobile phone-based “self-tracking app” designed to enhance the intrinsic utility that ASHAs derive from providing care to pregnant women.<sup>21</sup> In this paper, I mean “intrinsic” both in the sense used in the psychological literature (motivation to do a task for its own sake—effort as its own reward) and in the sense of prosocial preferences (motivation to do a task for its positive social externalities). Both types of motivation stand in contrast to extrinsic motivation, which may be rooted in individualistic (income, job security, career advancement) or social (status, reputation, recognition) concerns, but which in either case regards effort as palatable only for the extrinsic rewards it earns. A key challenge in assessing the mechanism of an intrinsic incentive is that it may also operate via these extrinsic channels. This is a partly a matter of design and partly an empirical question; I address both in the analysis of effects and mechanisms in Section 3.5.

---

<sup>21</sup>Credit for the technological development of the app is due to Brian DeRenzi.

Table 3.1: Descriptive statistics

	Mean	SD	p25	p75	N
<i>A. ASHA performance</i>					
Average monthly case load	13.13	5.66	9.21	15.70	72
New cases per month	2.89	1.34	2.02	3.52	72
Visits per month	7.40	4.22	4.12	9.98	72
Average share of open cases visited each month	0.46	0.14	0.38	0.54	72
Average gestational week in which case opened	21.67	2.39	19.97	23.31	72
Average visits per case during entire pregnancy	2.57	0.96	1.94	3.06	72
Share of completed cases with outcome reported	0.81	0.08	0.76	0.87	72
<i>B. Client characteristics and pregnancy outcomes</i>					
Age of client (years)	25.19	3.85			4913
Previous pregnancies	2.06	1.80			4576
Institutional delivery (all reported deliveries)	0.77	0.42			3275
Live birth (all reported pregnancies)	0.91	0.28			3835
Premature birth (all reported deliveries)	0.24	0.43			3501
Maternal mortality ratio (all live births, per 100,000)	256.70	5060.80			3506

*Notes:* Panel A includes ASHAs in the generic encouragement condition across all time periods (15 pre-intervention and 10 post-intervention months). Panel B encompasses all pregnant clients in the generic encouragement condition. Denominators are indicated in parentheses. Premature birth is defined as birth at least three weeks before the expected due date. “p25” and “p75” denote the 25th and 75th percentiles, respectively.

Another challenge is that an incentive may function as a production technology; i.e., it may increase output not only by increasing effort but also by increasing the productivity of effort,  $Y(e)$ . For example, an app that increases the frequency with which an ASHA interacts with her work phone may make her more adept at using the phone, which in turn may increase her productivity.<sup>22</sup> To mitigate this confound, I create an analogous app that mimics the user interface of the self-tracking app, but replaces its content with information that is, *a priori*, expected to be lower-powered as an intrinsic incentive. Because the two apps' interfaces are identical, treatment effects on take-up and performance can be attributed to the information content of the apps as opposed to putative motivational or learning effects of the technology with which the information is delivered. There remains the possibility that the information content could alter the production function; I address this in Section 3.5.

#### **3.4.1.1 High-powered intrinsic incentive: Self-tracking**

The self-tracking app enables ASHAs to access data visualizations of their work performance.<sup>23</sup> The data are compiled from the ASHAs' own submissions via their phone-based reporting tool. With one exception (the relative performance graph described below), all of the information contained in the visualizations is generated by the ASHA herself, which highlights the notion of "self-tracking," as opposed to performance feedback, in which the agent is provided with information that is not observable to her.

---

<sup>22</sup>Such an app would need not deliver performance-related information—e.g., a mobile phone game that confers learning benefits (e.g., how to use the phone) that could increase job-specific ability (e.g., how to fill out forms on the phone).

<sup>23</sup>The app is similar in concept to recently developed consumer-oriented mobile apps and wearable technologies that enable users to track and visualize data related to personal activities—e.g., Fitbit, Runkeeper, Apple Watch.

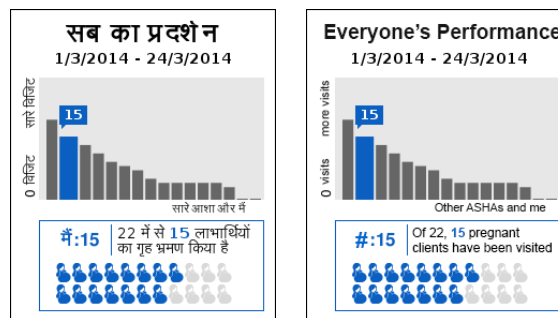
The app has four screens: a menu, a relative performance page, a calendar page, and historical trends page.

Figure 3.1: Self-tracking app: menu page (Hindi sample and English mock-up).



**Menu (Figure 3.1).** The menu page serves as the entry point for accessing the tool. It is accessed by pressing a predefined key on the phone. The menu displays a date interval spanning the first day of the current month to the current day. The calendar month is the primary performance interval for ASHAs; the official ASHA program guideline is that ASHAs should visit all of their pregnant clients at least once per month. Below the date interval are three rows that link to the three other pages, which the ASHA can select using the phone's navigation buttons.

Figure 3.2: Self-tracking app: relative performance page.



**Relative performance (Figure 3.2).** This page features an ordered bar graph of the number of unique pregnant clients visited in the current month, for the ASHA and 15 other ASHAs. The peer ASHAs are chosen randomly from among the 73



ASHAs who belong to the self-tracking treatment condition. The ASHAs are not told the identities of the peers. Anonymity is important for both ethical and theoretical reasons—ethically, to avoid harmful repercussions that may result from publicizing individual ASHAs’ performance, and theoretically, to unlink the informational effect of social comparison from external peer effects, such as these vey repercussions. Each ASHA’s peer group is randomly chosen, such that all peer groups are asymptotically identical but individually unique.<sup>24</sup> In order to avoid serial correlation effects, all peer groups are redrawn at the beginning of each month. The graph updates in response to changes in the ego ASHA’s performance, as well as the performance of peer ASHAs.

The relative performance graph is designed to provide an anonymous social benchmark that makes the information signal about performance more useful.<sup>25</sup> Conceptually, the relationship between social comparison and intrinsic and extrinsic motivation is nuanced. On one hand, publicly identifiable social comparison would be expected to interact with externally-oriented social preferences such as tastes for status and recognition. Private social comparison, however, does not engage these external preferences; status, for example, cannot be conferred on an anonymous entity.<sup>26</sup> On the other hand, private social comparison can interact with internally-oriented social pref-

---

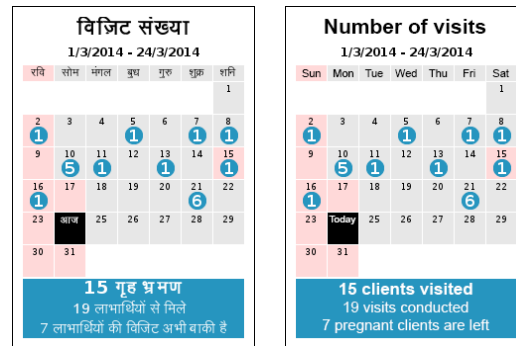
<sup>24</sup>In consequence, each peer group has, in expectation, the same average performance, but any given peer group may be higher- or lower-performing than average (and this may fluctuate on a day-to-day basis).

<sup>25</sup>In a different study population (such as students taking an exam), analogous information could be conveyed by disclosing only the group mean, which is also anonymous. In the ASHA population, piloting exercises revealed that the concept of “average” was difficult to convey to many ASHAs, and thus the visual display format was adopted.

<sup>26</sup>In the absence of peers, motivation around managing one’s self-image could affect effort—e.g., an ASHA who is motivated to do ASHA work because she self-identifies as a prosocial type. However, whereas social image motivation (in which an agent has preferences over how others attribute her behavior) implicates social preferences, self-image motivation can exist in the absence of social preferences, and thus I classify this as a type of intrinsic, not extrinsic, motivation, consistent with Bénabou & Tirole (2006, 2011).

erences, such as competitiveness and taste for winning, or preferences that may not involve social interaction but may be mediated by social information, such as social norms regarding private behavior. If these preferences are in relation to an abstract reference group and do not implicate social interaction, they more closely align with standard definitions of intrinsic rather than extrinsic motivation.<sup>27</sup>

Figure 3.3: Self-tracking app: calendar page.

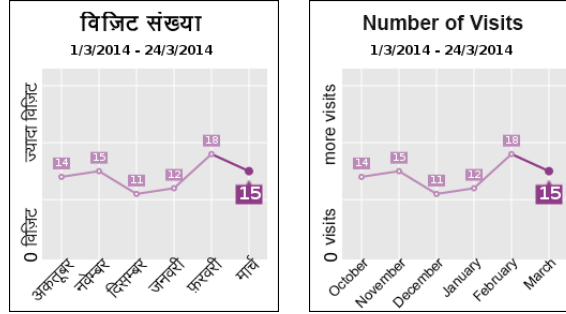


**Calendar (Figure 3.3).** The remaining two pages are more straightforward. The calendar page displays a calendar in the standard seven-day format, which is the local convention.<sup>28</sup> On each day, if any visits to pregnant clients occur, the number of visits is indicated in a circled number. A counting rule restricts the number of times an ASHA can get “credit” for visiting a given client in a given day to one, but there is no restriction on the number of times an ASHA can report visiting a client in a given month.

<sup>27</sup>To illustrate, consider a competitive athlete. While preferences for status and recognition may play a role in driving the athlete’s effort, there may also be a private component—a desire to excel—that is intrinsic. The athlete might set a performance goal that is absolute (running a mile in less than five minutes), relative to her own performance (setting a personal record), or relative to others’ performance (setting a world record). In each case, good performance can be an end in itself, not an instrument for attaining other rewards. Relative social benchmarks, a type of social norm, have been shown to exert substantial influence on behavior even when the behavior is private and does not entail social interaction, such as home energy consumption (Allcott, 2011; Schultz et al., 2007).

<sup>28</sup>Sundays and national holidays are shaded red; the current day is shaded black.

Figure 3.4: Self-tracking app: historical page.



**Historical (Figure 3.4).** This page displays a line graph plotting the total number of clients visited each month for the current month (to date) and the five preceding months, in a rolling manner.

Taken together, the self-tracking app provides the ASHA with the ability to access personalized information about her performance. The data elements conveyed include: total caseload, number of clients visited in the current month, number of clients visited each day in the current month, number of visits conducted in the current month, number of clients visited monthly over the previous six months, and visit performance relative to an abstract peer group. All of these elements are meant to make the effort task and its prosocial impact more psychologically salient to the ASHA, thereby, in theory, amplifying the marginal intrinsic utility of effort.

### 3.4.1.2 Low-powered intrinsic incentive: Generic encouragement

The theoretical framework states that intrinsic incentive regimes  $h$  and  $l$  produce  $\psi_h = 1$  and  $\psi_l = 0$ , respectively. In view of the latter, a counterfactual to the self-tracking app could be to provide no experimental intervention to the control group. However, this would introduce a mechanism confound; any treatment effect could be due to the interactive nature of the app, rather than its information content. The intent of the experiment is to focus on the incentive effect of information itself. To this end,

I develop a counterfactual intervention that preserves the technological interface of the self-tracking app, but provides generic information that is putatively less effective at making the effort task and its prosocial impact salient.

The low-powered app also has four pages; sample pages are shown in Figure 3.5. The content includes generic encouragement messages. A different set of pages is generated on a daily basis, out of an inventory of 52 sets. The menu page points to three further pages: “Responsibilities of an ASHA,” “Advice for Healthy Mothers & Babies,” and “Inspiring Quotes.” Each page includes a statement accompanied by a picture that illustrates the statement. Content for the first two sections is drawn from ASHA training materials. The “Inspiring Quotes” section contains quotes drawn from Hindi-language websites, with many attributed to well-known South Asian cultural figures such as Gandhi and Mother Theresa.

### **3.4.1.3 Audio service**













In the study sample, 28% of the ASHAs are illiterate.<sup>29</sup> To ensure that the interventions would be useful to these ASHAs, the research team developed an automated audio version of each intervention. Analogous to the act of accessing the app on the phone, an ASHA calls a designated phone number from her work phone. An audio recording specific to the phone number is played back to the caller. There is no limit on how often the ASHA can utilize the audio service.

While there are differences in the user experience of the audio and the visual software interface (e.g., navigation is possible only in the latter), the underlying information content is the same. Thus, the theoretical framework is unaffected, and

---

<sup>29</sup>Each ASHA was asked to read a Hindi sentence which stated, “The woman went to the market to buy vegetables.” Twenty-eight percent were able to read no words or only a few words; the remainder were able to read all of the words and are classified as literate.

Figure 3.5: Generic encouragement app: sample pages.

<p><b>रेशमा देवी</b> 24/3/2014 - 30/3/2014</p>  एक आशा के कर्तव्य  स्वस्थ माँ और बच्चे के लिए निर्देश  प्रेरणा दायक विचार	<p><b>एक आशा के कर्तव्य</b></p>  <p>आशा को महीने में कम से कम एक बार अपने लाभार्थी के घर ग्रहभ्रमण करना चाहिए/ अगर लाभार्थी के घर दो साल से छोटा बच्चा, कुपोषित बच्चा या गर्भवती महिला है तो आशा को लाभार्थी के घर जाकर परामर्श देना चाहिए।</p>	<p><b>स्वस्थ माँ और बच्चे के लिए निर्देश</b></p> <p><b>4</b></p>  <p>गर्भावस्था के दौरान चार प्रसव पूर्व जांच करना सुनिश्चित करें और पंजीकरण पहले तीन महीने के अंदर ही करें।</p>	<p><b>प्रेरणा दायक विचार</b></p>  <p>यदि आपके कार्य दूसरों को अधिक सपने देखने, अधिक सीखने, अधिक करने और अधिक बनने को प्रेरित करते हैं तो आप एक नेता हैं।</p>
<p><b>Reshma Devi</b> 24/3/2014 - 30/3/2014</p>  Responsibilities of an ASHA  Advice for Healthy Mothers & Babies  Inspiring Quotes	<p><b>Responsibilities of an ASHA</b></p>  <p>Home visits should take place at least once in a month if not more. Where there is a child below two years of age or any malnourished child or a pregnant woman, she should visit the families at home for counselling them.</p>	<p><b>Advice for Healthy Mothers &amp; Babies</b></p> <p><b>4</b></p>  <p>During pregnancy, four antenatal visits must be ensured, including registration within the first three-month period.</p>	<p><b>Inspiring Quote</b></p>  <p>If your actions inspire others to dream more, learn more, do more, and become more, you are a leader.</p>

in the empirical analysis, unless otherwise stated, I combine usage of the visual and audio systems to yield a composite measure of take-up.

### 3.4.2 Randomization, implementation, and data sources

The 145 ASHAs in Muratganj were randomly assigned to one of the two intrinsic incentive treatments: self-tracking and generic encouragement. Randomization was conducted in May 2014, one month before the launch of the experiment, and was stratified by six variables: Hindi literacy, total client visits conducted over the prior 4 and 12 months, respectively, and scores on three psychometric scales for extrinsic, intrinsic, and prosocial motivation, respectively.<sup>30,31</sup> In June 2014, the self-tracking and generic encouragement apps were installed on ASHAs' phones, and all 145 ASHAs were trained in their use. All ASHAs were told that two different phone-based tools were being piloted, and that the eventual plan was to make both available if so desired by ASHAs. No complaints were raised to the research team or the implementing NGO during the training or at any point thereafter regarding the randomization.

Once the experiment launched, care was taken to preserve the intent of the intervention: to alter the intrinsic utility of effort without altering real or perceived extrinsic returns to effort. To avoid a potential monitoring effect, no efforts were made to affect demand through, e.g., routine follow-up visits, marketing, or interactions during other CommCare activities such as trainings.<sup>32</sup> Instead, the research team

<sup>30</sup>Literacy was directly measured during the baseline survey by asking the ASHAs to read a Hindi sentence. The two visit measures were included to capture both short- and medium-term baseline work performance. The psychometric scales are described in detail in Section C.1.

<sup>31</sup>I used the “T-min-max” re-randomization method with 1,000 draws to carry out the random assignment. For a discussion of this randomization method, see Bruhn & McKenzie (2009).

<sup>32</sup>The one partial exception was an automated SMS system created by the research team, in which a text message is sent to each ASHA every Monday stating either (for self-tracking app users), “Your

interacted with ASHAs only when an ASHA called a research assistant to request assistance with the self-tracking or encouragement app. Such assistance variously involved re-installing the software in the case of accidental deletion, re-saving the phone number used to access the audio service, and addressing questions regarding interpreting the information in the app.

For the first eight months of the experiment, the research team had no contact with ASHAs other than through these troubleshooting visits. At nine months of follow-up, a midline survey was administered, in which 142 out of 145 ASHAs participated. In the course of each ASHA’s interview, the research team rechecked all phone settings related to the use of the visual and audio services, fixed settings as necessary, and documented any steps taken.

In addition to the baseline and midline survey data, this experiment relies on performance data reported by the ASHAs through their phone-based record-keeping tool, and app usage data measured directly. Client visits and app sessions are timestamped with start and end times. Client visits are tied to individual clients, allowing differentiation of initial registration visits and follow-up visits. These and other data sources are described in further detail in Appendix Section C.1.

visits information is available. Please press the shortcut button to access your information,” or (for encouragement app users), “Your advice and encouragement is available. Please press the shortcut button to access your encouragement.” ASHAs were told that these messages were sent automatically to all ASHAs. SMS use is low overall in this setting, and we find no significant Monday fixed effect for app usage.

## 3.5 Analysis

### 3.5.1 Randomization balance

Table 3.2 reports tests of equality for variables measured at baseline across five domains: job performance, job-specific ability, psychological traits, ASHA demographic characteristics, and village characteristics. Of the 18 variables tested, one (whether ASHA reports that her work village is predominantly Muslim) is significant at the 10% level, as would be expected by chance. Jointly, the variables are not significant ( $p = 0.82$ ). In addition, I run  $t$ -tests for differences in means between the two treatment conditions for all 263 numeric variables in the baseline survey. Of these, 4.9% of the differences are significant at the 10% level; 1.9% at the 5% level; and 0.4% at the 1% level. Repeating this exercise for daily home visits for all 455 days in the pre-experimental period,  $t$ -tests reveal that 6.9%, 2.5%, and 0.5% of daily visit counts are significantly different at the 10%, 5%, and 1% levels, respectively. Taken together, these results indicate failure to reject the null hypothesis that the two treatment groups are identical on observables. In the analysis that follows, to the extent possible, baseline differences are controlled for, either with explicit covariates or with fixed-effect estimators that implicitly control for all time-invariant ASHA characteristics.

### 3.5.2 Empirical strategy and average treatment effects

To identify causal effects, I exploit the fact that performance are observed at the daily (or, in the case of earnings, monthly) level. This allows for panel analysis with ASHA fixed effects that control for all time-invariant ASHA characteristics. In addition, the



Table 3.2: Randomization balance

<i>Treatment assignment:</i>	<b>Generic encouragement</b>		<b>Self-tracking</b>		<b>p-value of difference</b>
	(1)		(2)		(3)
<i>Baseline performance (March 2013 - May 2014):</i>					
New clients per month	3.329	[1.457]	3.213	[1.477]	0.636
Visits per month	8.344	[4.296]	8.252	[4.421]	0.899
Monthly earnings (INR)	1989.5	[1459.1]	1974.9	[1360.1]	0.951
<i>Baseline ability:</i>					
Education (years)	9.264	[2.556]	8.795	[3.140]	0.326
Able to read Hindi (0/1)	0.806	[0.399]	0.822	[0.385]	0.802
Score on general health knowledge assessment (0-18)	15.53	[1.34]	15.16	[1.60]	0.141
Score on pregnancy knowledge assessment (0-12)	8.611	[1.369]	8.904	[1.556]	0.231
<i>Psychological traits:</i>					
Extrinsic motivation scale (1-5)	4.208	[0.489]	4.229	[0.566]	0.811
Intrinsic motivation scale (1-5)	4.289	[0.415]	4.279	[0.432]	0.894
Prosocial motivation scale (1-5)	4.499	[0.386]	4.510	[0.381]	0.858
<i>Other ASHA characteristics:</i>					
Age (years)	33.24	[6.91]	33.34	[7.13]	0.928
Belongs to a disadvantaged caste (0/1)	0.806	[0.399]	0.822	[0.385]	0.802
Resides in work village (0/1)	0.903	[0.298]	0.822	[0.385]	0.160
Has electricity at home (0/1)	0.694	[0.464]	0.685	[0.468]	0.902
Has non-mud floor at home (0/1)	0.389	[0.491]	0.384	[0.490]	0.948
Asset index (quintiles 1-5)	2.901	[1.343]	3.068	[1.484]	0.480
<i>Village characteristics:</i>					
Village population (national census)	1435.4	[851.8]	1321.9	[719.7]	0.572
Village is majority Muslim (0/1)	0.278	[0.451]	0.151	[0.360]	0.063*
Number of ASHAs	72		73		
Joint <i>F</i> -test of individual characteristics ( <i>p</i> -value)			0.8199		

*Notes:* Table reports means and standard deviations (in brackets) for the two experimental conditions, as well as a test of equality for each variable, with stars signifying \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ . The first two baseline performance variables (monthly new clients and visits) are obtained from the ASHAs' CommCare form submission data during the 15 months prior the experimental intervention launch (March 2013 - May 2014). Mean monthly earnings pre-intervention is collected from the ASHAs' payment books, in which the government payment office records payments for ASHA work. Education, Hindi literacy, and the health knowledge variables were all obtained from the baseline survey. The psychometric scales and work motivation variables were measured through the baseline survey; see the Data Appendix for more details. ASHA characteristics were measured in the baseline survey. The village population variable is based on 2001 Indian national census data. "Village is majority Muslim" was self-reported by ASHAs in the baseline survey.

availability of pre-experimental data allows for differencing out time-variant, ASHA-specific trends. I estimate the general equation:

$$y_{it} = \beta_0 + \beta_1 A_t + \beta_2 A_t * T_i + \alpha_i + Z_m \gamma_m + u_{it} \quad (3.5.1)$$

where  $y_{it}$  is an outcome of interest for ASHA  $i$  at time  $t$ ;  $A_t$  is an indicator for whether time  $t$  is after the launch of the experiment;  $T_i$  is an indicator for treatment assignment that takes value 0 in the generic encouragement condition and 1 in the self-tracking condition;  $\alpha_i$  is a vector of ASHA fixed effects;  $Z_m$  is a vector of month-year fixed effects; and  $u_{it}$  is the error term. I assume that errors are serially correlated and thus present standard errors clustered at the ASHA level throughout.

Randomization ensures, in expectation, that  $T_i \perp u_{it}$ , and that trends in  $Y_{it}$  between the two treatment groups, in the absence of treatment, would be parallel. Furthermore, spillovers between treatments is unlikely given that ASHAs live in different villages and cover defined, non-overlapping catchment areas and thus are not likely to experience spillovers via either market demand or through direct interactions. Finally, there is virtually no attrition during the follow-up period of the experiment. Of the 145 ASHAs, all participate in app training, and at ten months of follow-up, only 3 ASHAs (one in the generic encouragement condition and two in the self-tracking condition) are no longer working (all three due to outmigration). Under these identifying assumptions,  $\beta_2$  measures the average causal effect of the self-tracking app relative to the generic encouragement app.

### 3.5.2.1 Average treatment effects on take-up and client visits

Table 3.3 presents average treatment effects on take-up of the interventions, on client visits, and on earnings. Columns 1-3 show results for take-up of the main visual soft-

ware, the complementary audio service, and total use of the two modalities combined. All estimates are in units of sessions per day. Three findings are noteworthy. First, despite an experimental protocol that provides little explicit encouragement to use the interventions, demand for both apps is high. The average ASHA in the self-tracking (generic encouragement) treatment uses her app once every 3.75 (3.11) days over the course of the experimental period. To contextualize this, client visits occur every 3.83 and 4.85 days in the self-tracking and generic encouragement conditions, respectively, during the experimental period. Thus, in both conditions, app usage is higher than visit frequency, suggesting that use of the phone for visits is not exclusively driving use of the apps.<sup>33</sup> Second, that take-up of the two apps is similar (with the point estimate favoring the encouragement app) suggests that any treatment effects favoring the self-tracking app cannot be explained by differences in use of the phone (e.g., a learning-by-doing effect that makes frequent phone users more efficient at filling out forms).

The remaining columns of Table 3.3 examine two measures of performance: visits and earnings. Partitioning visits into the initial registration visit and follow-up visits to a given client, the experiment reveals no effect of the self-tracking app on the former (Column 4), but a precisely estimated 36.6% increase in reported follow-up visits in the self-tracking treatment (Column 5). Driven by this effect, self-tracking CHAs report 26.8% more total visits than their counterparts (Column 6). To investigate the time pattern of this effect, Column 7 separates the 10-month experimental period into first and second halves. The treatment effect is concentrated in the first half of the experimental period. Figure 3.6 illustrates this finding graphically by plotting

---

<sup>33</sup>That take-up of the the visual software is higher for the generic encouragement app than for the self-tracking app is likely because the encouragement app provides new content daily, whereas the self-tracking app is informationally static in the absence of visits, which occur less than daily for all ASHAs.

Table 3.3: Average treatment effects on take-up and performance

	Take-up			Performance			
	Visual sessions (1)	Audio sessions (2)	Total sessions (3)	New client visits (4)	Follow-up visits (5)	Total visits (6)	Earnings (INR) (8)
Self-tracking app x Post-intervention period	-0.0461* (0.0258)	-0.00884 (0.0251)	-0.0550 (0.0401)	0.00513 (0.00469)	0.0501*** (0.0150)	0.0553*** (0.0169)	-71.64 (184.3)
Self-tracking app x Second half of post-intervention period						-0.0480*** (0.0182)	
Mean [SD] for generic encouragement app	0.175 [0.182]	0.147 [0.156]	0.322 [0.265]	0.069 [0.344]	0.137 [0.501]	0.206 [0.659]	1550.9 [2218.2]
Unit	daily mean	daily mean	daily mean	day	day	day	month
Pre-intervention units	0	0	0	455	455	455	15
Post-intervention units	286	286	286	286	286	286	8
Baseline controls	Y	Y	Y	N/A	N/A	N/A	N/A
Month-year fixed effects	N/A	N/A	N/A	Y	Y	Y	Y
Adjusted R <sup>2</sup>	0.107	0.091	0.145	0.016	0.047	0.042	0.474
ASHAs	145	145	145	145	145	145	134
Observations	145	145	145	101184	101165	101195	2326

Notes: Table reports OLS estimates for cross-sectional (Columns 1-3) and fixed-effects panel regressions (Columns 4-8). Column headings are dependent variables. Standard errors (clustered at the ASHA level for panel regressions) are in parentheses. The coefficient on the post-intervention time dummy is suppressed in Columns 4-8. Baseline controls in Columns 1-3 include the stratification variables (literacy, baseline monthly visits, extrinsic motivation, intrinsic motivation, and prosocial motivation) and whether the ASHA reports that her village is majority Muslim. The mean in the generic encouragement app condition is reported for the post-intervention period (standard deviation in brackets). The differences in observations in Columns 4-7 are due to varying start dates for ASHAs at the beginning of the pre-intervention period. In Column 8, data for 11 ASHAs are missing due to these ASHAs either no longer working (3 ASHAs) or working but no longer in possession of the government ledger book in which their payments are recorded. The most recent two months of the experiment are also missing in Column 8 due to a lag in payments. Estimates are significant at the \*10%, \*\*5%, and \*\*\*1% level.

the treatment effect on total visits for each month of the experiment. The effect is largest in the first four months of the experiment, and then steadily declines over the next six months, though all point estimates remain positive.

Finally, reinforcing these results as well as the intuition that follow-up visits reflect intrinsic effort, the self-tracking app has no impact on earnings (Column 8). Thus, assuming rational expectations and accurate measurement, the increase in follow-up visits in the self-tracking app treatment cannot be explained by monetary preferences alone.

### 3.5.3 Heterogeneous treatment effects

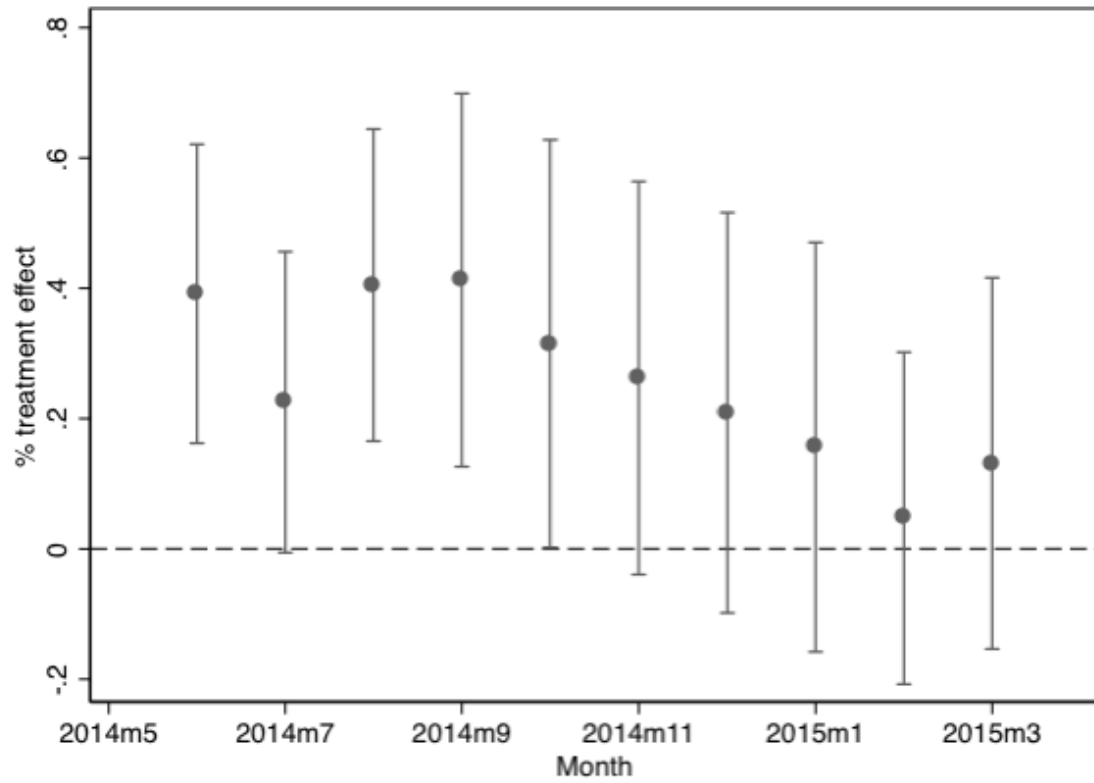
The theoretical framework in Section 3.2 predicts complementarity between intrinsic incentives and intrinsic motivation and that, in the presence of intrinsic aversion, an intrinsic incentive may dampen effort. In this section, I assess both predictions.

Table 3.4 reports the estimates of

$$\ln(y_{it}) = \beta_0 + \beta_1 A_t + \beta_2 A_t T_i + \beta_3 A_t \ln(\theta_i^k) + \beta_4 A_t T_i \ln(\theta_i^k) + \alpha_i + Z_m \gamma_m + u_{it} \quad (3.5.2)$$

where  $\ln(y_{it})$  is log total visits reported by ASHA  $i$  on day  $t$ ;  $A_t$  is the post-intervention indicator;  $T_i$  is the treatment indicator;  $\ln(\theta_i^k)$  is the log motivation of ASHA  $i$  with respect to psychometric dimension  $k \in \{\text{extrinsic, intrinsic, prosocial, social desirability, competitive}\}$ ;  $\alpha_i$  is a vector of ASHA fixed effects;  $Z_m$  is a vector of month-year dummies; and  $u_{it}$  is the error term. The coefficient of interest is  $\beta_4$ , which is the marginal elasticity of effort with respect to psychometric trait  $\theta^k$  in the self-tracking treatment, relative to that in the generic encouragement treatment. In other words, it is the difference in elasticity between two curves—those between total visits

Figure 3.6: Impact of self-tracking app on total visits, by month



*Notes:* Plot shows treatment effects of the self-tracking app on total visits in each month of the experiment, expressed as a percentage of mean visits in the generic encouragement condition in that month. Error bars are 90% confidence intervals.

and  $\theta^k$  in the self-tracking treatment and in the generic encouragement treatment, respectively—and it measures the self-tracking app’s degree of complementarity with a given motivational trait relative to that of the generic encouragement app.

In Table 3.4, each row is a fixed-effects regression of Equation 3.5.2, where only the row variable changes. Column 1 (which reports estimates of  $\beta_3$  in the above equation) shows that, in the control condition, intrinsic and prosocial motivation have negative elasticities of effort as measured by total visits. That is, a 1% increase in intrinsic (prosocial) motivation is associated with a 0.11% (0.09%) decrease in total visits per day. This itself is not remarkable, as motivation is not exogenous and could be correlated with other traits (e.g., ability) that affect performance. More important is the finding in Column 2 that the intrinsic and prosocial motivation elasticities of effort are significantly higher in the self-tracking condition than in the generic encouragement condition. In other words, the self-tracking condition is more effective at eliciting performance the more intrinsically/prosocially motivated an ASHA is; it leverages intrinsic/prosocial motivation. This relationship does not hold for extrinsic motivation, and as two additional placebo tests, it does not hold for social desirability or competitive motivation.<sup>34</sup> The two placebo traits are notable in that they are plausible confounds: the self-tracking app may leverage social desirability or competitiveness, but this is not observed in the data.

Table 3.5 further characterizes these results by partitioning the sample into terciles of psychometric traits and estimating Equation 3.5.1 for each tercile. Specifications 1, 3, and 5 use a pooled causal estimator for the entire post-intervention period, whereas Specifications 2, 4, and 6 estimate each tercile-specific treatment effect in the first half of the experiment, as well as the change in treatment effect in the second half. The

---

<sup>34</sup>Both of these psychometric scales were administered at baseline along with the other psychometric traits. The scale items are listed in Appendix Section C.2.

Table 3.4: Elasticities of effort with respect to psychometric traits

	<b>Total visits</b>	
	Elasticity in generic encouragement treatment	Marginal elasticity in self-tracking treatment
	(1)	(2)
Extrinsic motivation	-0.155 (0.126)	0.0806 (0.169)
Intrinsic motivation	-0.110*** (0.0257)	0.102** (0.0495)
Prosocial motivation	-0.0885*** (0.0242)	0.104** (0.0499)
Social desirability	-0.0110 (0.0301)	0.0472 (0.0544)
Competitive preference	-0.00586 (0.0296)	-0.00989 (0.0501)

*Notes:* Table reports OLS estimates of fixed-effects panel regressions. Standard errors (clustered at the ASHA level) are in parentheses. The unit of analysis is the ASHA-day, and the dependent variable is log total visits. Each row is a specification that interacts the row psychometric trait with the post-intervention dummy to estimate Column 1, and with the post-intervention dummy and the self-tracking treatment to estimate Column 2. Estimates for the post-intervention dummy and the interaction of treatment dummy and post-intervention dummy are suppressed. All psychometric traits are measured on a 1-5 scale, lowest to highest. All specifications include month-year fixed effects. Estimates are significant at the \*10%, \*\*5%, and \*\*\*1% level.



estimates in the odd-numbered columns for intrinsic and prosocial motivation show that the self-tracking app has no effect on total visits in the least motivated tercile of ASHAs with respect to each trait. In contrast, the self-tracking app has positive effects in the middle terciles and even stronger effects in the top terciles of each trait. The largest treatment effect, which is observed for the top tercile of intrinsic motivation, is 46% of the mean in the generic encouragement condition. While the point estimates of the treatment effects for each tercile of extrinsic motivation have a positive slope, it is not significant, as was shown in Table 3.4.

The even-numbered specifications in Table 3.5 illustrate how the treatment effects heterogeneously evolve over time. Across all three motivational traits, the least motivated tercile exhibits a positive treatment effect in the first half of the experiment, but this effect decays to zero in the second half. A similar decay is observed in the middle tercile of extrinsic motivation, whereas the treatment effect in the top tercile of extrinsic motivation increases by an insignificant amount from an insignificant magnitude in the first half of the experiment. In contrast, in the middle and top terciles of intrinsic and prosocial motivation, the treatment effect is positive and precisely estimated in the first half of the experiment, and it persists through the second, half, showing no significant decay. That is, the most intrinsically/prosocially motivated ASHAs respond to the self-tracking app from the outset, and their response persists, whereas less intrinsically/prosocially motivated ASHAs respond at the outset but only transiently.

Taken together, the foregoing analysis of average and distributional effects suggest three main findings. The self-tracking app treatment leads to an average increase in client visits; consistent with theory, this effect interacts positively with intrinsic and prosocial motivation; and this complementarity is the result of the dynamic persistence of treatment response among more intrinsically and prosocially motivated

Table 3.5: Treatment effects on total visits by terciles of psychometric traits

	Total visits					
	Tercile of row variable					
	Bottom		Middle		Top	
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Extrinsic motivation</i>						
Self-tracking app x Post-intervention period	0.0384 (0.0268)	0.0769** (0.0292)	0.0584* (0.0299)	0.0893** (0.0342)	0.0638* (0.0361)	0.0543 (0.0391)
Self-tracking app x Second half of post-intervention period		-0.0810*** (0.0276)		-0.0662* (0.0392)		0.0141 (0.0232)
Adjusted R <sup>2</sup>	0.0426	0.0429	0.0399	0.0400	0.0448	0.0446
ASHAs	58	58	51	51	36	36
Observations	40415	40415	35563	35563	25217	25217
<i>Intrinsic motivation</i>						
Self-tracking app x Post-intervention period	0.0145 (0.0278)	0.0514* (0.0304)	0.0678** (0.0279)	0.0777*** (0.0277)	0.0956*** (0.0300)	0.107*** (0.0358)
Self-tracking app x Second half of post-intervention period		-0.0794*** (0.0271)		-0.0246 (0.0293)		-0.0259 (0.0357)
Adjusted R <sup>2</sup>	0.0475	0.0477	0.0303	0.0302	0.0472	0.0471
ASHAs	59	59	41	41	45	45
Observations	40865	40865	28753	28753	31577	31577
<i>Prosocial motivation</i>						
Self-tracking app x Post-intervention period	0.0236 (0.0294)	0.0538* (0.0305)	0.0808*** (0.0285)	0.0982*** (0.0298)	0.0783** (0.0295)	0.0863** (0.0354)
Self-tracking app x Second half of post-intervention period		-0.0653*** (0.0217)		-0.0352 (0.0353)		-0.0242 (0.0413)
Adjusted R <sup>2</sup>	0.0392	0.0393	0.0478	0.0478	0.0419	0.0417
ASHAs	65	65	42	42	38	38
Observations	45104	45104	29295	29295	26796	26796

Notes: Table reports OLS estimates of fixed-effects panel regressions at the day level. Standard errors (clustered at the ASHA level) are in parentheses. The dependent variable is total visits. Each column estimates the specified coefficient(s) in the specified tercile of the row variable. All specifications include month-year fixed effects. Estimates are significant at the \*10%, \*\*5%, and \*\*\*1% level.

workers. As to the prediction that a high-powered intrinsic incentive may dampen effort for those who are intrinsically unmotivated, the results do demonstrate a decay in client visits over time amongst the least intrinsically/prosocially motivated ASHAs. Whether this treatment effect will actually turn negative remains to be seen.

In the remainder of this sub-section, I examine the robustness of these results with respect to the validity of the data and the impact of the experimental interventions on health outcomes.

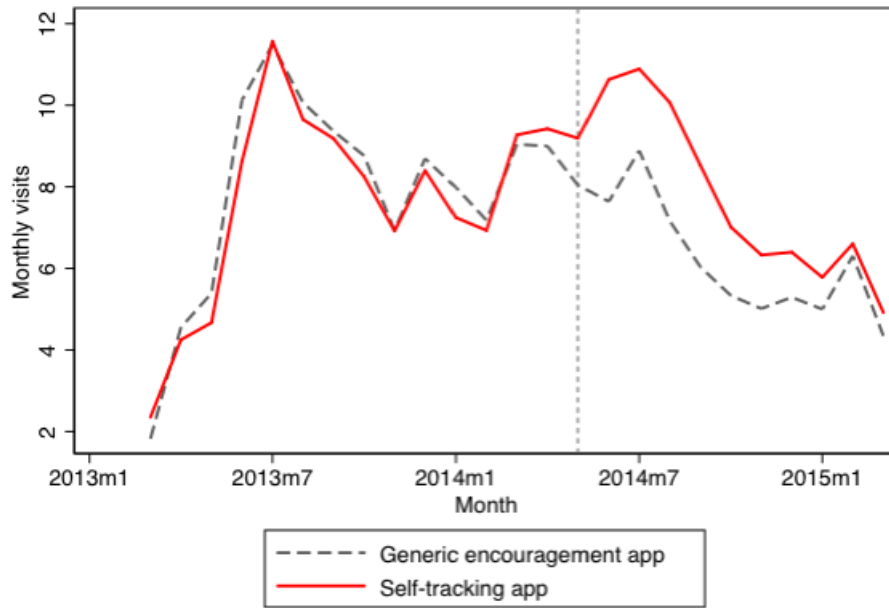
### **3.5.4 Compensating mechanisms**

Table 3.6 examines whether the effect of the self-tracking app on reported visits is compensated by negative effects on other measures of effort. One such mechanism could be poorer targeting of visits both across space—e.g., visiting easy-to-reach clients many times, to the exclusion of other clients—and across time—e.g., visiting clients multiple times in some months but none in others. Column 1 reports estimates of a month-level panel regression in which the dependent variable is the share of pregnant clients visited by an ASHA in a given month. During the experimental period, self-tracking ASHAs visit an 8.72-percentage point greater share of their clients each month, a 23.1% improvement over the control mean. That this effect is approximately equal to the treatment effect on total visits indicates that increased visits do not occur at the expense of coverage. Figure 3.7b graphically illustrates this increase in client coverage.

Columns 2-5 test for fabrication of visits. How long an ASHA spends completing a visit form is tracked by the phone directly, from the time she opens the form to when she completes it. Shorter visit duration may reflect higher productivity—e.g., greater proficiency at typing on the phone—or a socially efficient allocation of time

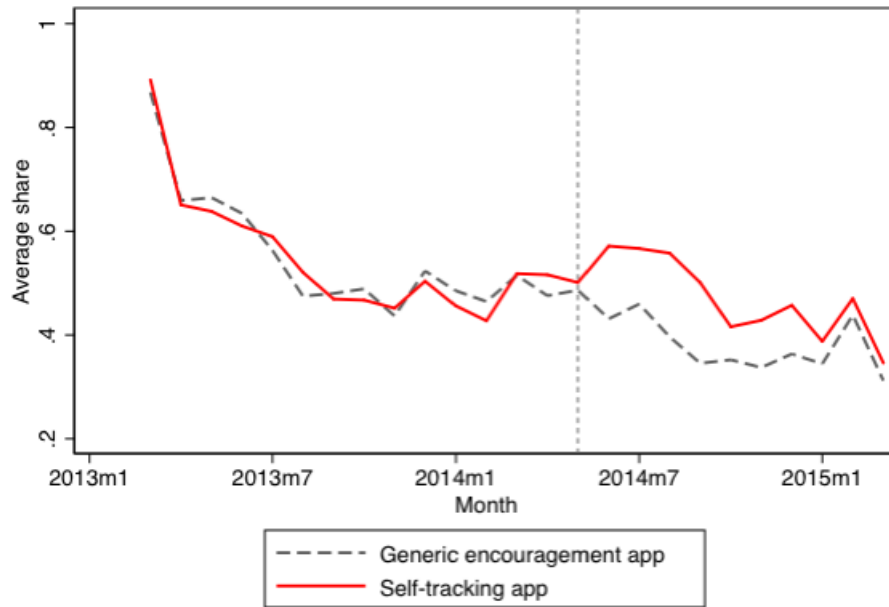
Figure 3.7: Treatment effects on visits and client coverage

(a) Average visits per day



Note: Dashed vertical line denotes last pre-intervention month (May 2014).

(b) Average share of pregnant clients visited each month



Note: Dashed vertical line denotes last pre-intervention month (May 2014).

across clients and visits. Nevertheless, it also raises concern for fabrication. Column 4 shows that self-tracking ASHAs spend 13.5% less time filling out forms; this effect is similar when considering new client visits and follow-up visits separately (Columns 2-3). Taken together, these results suggest two complementary interpretations. First, although average visit duration is 13.5% shorter, because self-tracking CHAs report 26.8% more total visits, they spend more aggregate time “inside” forms. While this does not rule out fabrication, it makes pure fabrication less plausible, since that would be expected to be a time-saving strategy manifesting in part with a higher proportion of very short visits. Second, similar effects on duration are observed for both initial and follow-up visits. This too casts doubt on the extent of fabrication, as the nature of the two types of visits is that it is easier to fabricate a follow-up visit form (which consists of checking off a list of counseling topics, which is difficult to verify) than an initial registration form (which requires typing the name of a client, her husband, her phone number, and so forth, all of which can be verified).

### 3.5.5 Health impacts

Finally, Table 3.6, Columns 5-12 assesses for treatment effects on four health practices and outcomes at the client level, as reported by ASHAs. These include practices such as attending antenatal care visits and receiving tetanus vaccinations, as well as pregnancy outcomes such as institutional delivery and maternal death. Columns 5-12 estimate

$$y_{ijt} = \beta_0 + \beta_1 T_j + \beta_2 E_t + \beta_3 T_j * E_t + Z_i \theta + \alpha_j + u_{ijt} \quad (3.5.3)$$

where  $y_{ijt}$  is the outcome of client  $i$  of ASHA  $j$  during period  $t$ ;  $T_j$  is ASHA  $j$ 's treatment assignment, where  $T_j = 1$  for the self-tracking treatment;  $E_t$  denotes whether

the client's pregnancy was exposed to the experimental period;  $Z_i$  is a vector of client-level controls; and  $\alpha_j$  is a vector of ASHA dummies. The estimation sample is restricted to clients who have completed their pregnancy.  $E_t$  defines as exposed those clients who were registered by an ASHA prior to the launch of the experiment (so as to ensure no endogenous selection) but whose pregnancy concluded after the launch ( $N = 1,779$ ). Those who began and ended their pregnancy before the launch of the experiment are classified as non-exposed ( $N = 4,820$ ). The difference-in-differences estimator of interest is  $\beta_3$ , which, under the parallel trends assumption, identifies the effect of the self-tracking app on client outcomes. Standard errors are clustered at the ASHA level.

In brief, the results are equivocal. Regarding antenatal practices, the self-tracking condition leads to an increase in reported ANC visits but has no effect on ASHAs' reports of how many tetanus vaccines the client received and whether the client has developed a birth plan. The average effect on these practices is not significantly different from zero (Column 8).

Regarding pregnancy outcomes, these too are reported by ASHAs, but using a separate form at the conclusion of the pregnancy. Whether the ASHA submits a pregnancy outcome form is itself an important outcome, as it is likely to be (negatively) correlated with whether the client has been lost to follow-up. On this margin, we observe no significant effect (Column 9). Surprisingly, Column 10 shows that, conditional on delivery, the probability of institutional delivery is 4.8 percentage points lower in the self-tracking group than in the generic encouragement group, in which 78.8% of deliveries occur at a health facility. No effect on the probability of live birth (at home or at a facility) is observed. When pooled, their average effect size is 0.06 standard deviations in favor of the generic encouragement treatment, which is on the cusp of reaching statistical significance ( $p = 0.113$ ).

Table 3.6: Average treatment effects on coverage, visit duration, and health outcomes

	Visit duration (minutes)				Antenatal practices				Pregnancy outcomes			
	Share of clients visited	New client visits	Follow-up visits	All visits	ANC visits (0-3)	Tetanus vaccinations (0-2)	= 1 if has birth plan	Average effect size	= 1 if outcome reported	= 1 if institutional delivery	= 1 if live birth	Average effect size
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Self-tracking app x Post-intervention period	0.0872*** (0.0237)	-1.813* (1.032)	-1.220 (0.914)	-1.539* (0.837)	0.164* (0.0969)	0.0758 (0.0503)	-0.00873 (0.0275)	0.0707 (0.0657)	-0.0167 (0.0214)	-0.0483 (0.0299)	-0.00427 (0.0176)	-0.0605 (0.0382)
Mean [SD] for generic encouragement app	0.376 [0.222]	14.62 [15.35]	12.50 [12.87]	13.121 [13.80]	1.013 [1.021]	0.995 [0.751]	0.945 [0.229]	0.923 [0.266]	0.788 [0.409]	0.886 [0.319]		
Unit	month	visit	visit	visit	client	client	client	client	client	client	client	client
Client controls	N/A	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y
ASHA fixed effects	N/A	Y	Y	Y	Y	Y	y	Y	Y	Y	Y	Y
Adjusted R <sup>2</sup>	0.343	0.084	0.149	0.112	0.180	0.124	0.218	N/A	0.119	0.118	0.047	N/A
ASHAs	145	145	145	145	145	145	145	145	145	145	145	145
Observations	3409	9062	17365	26427	6566	6566	6003	6566	6566	5081	5801	6566

Notes: Table reports OLS estimates of fixed-effect panel regression at the month level (Column 1) and cross-sectional regressions at the level of the visit (Columns 2-4) and client (Columns 5-12). Standard errors clustered at the ASHA level are in parentheses. Column headings are dependent variables. Coefficients on the self-tracking app and post-intervention dummies are suppressed. The mean in the generic encouragement app condition is reported for the post-intervention period (standard deviation in brackets). Visit duration in Columns 2-4 is winsorized at the 1st and 99th percentiles. In Columns 5-12, client-level controls are client age, parity, mobile phone ownership, and gestational week at registration by ASHA. Columns (8) and (12) report average effect sizes (Clingsmith et al., 2009; Kling et al., 2004) of the variables for antenatal practices and pregnancy outcomes, respectively. Estimates are significant at the \*10%, \*\*5%, and \*\*\*1% level.

These results on health impacts raise several questions. Assuming no misreporting of performance data, it is difficult to explain how the self-tracking treatment would reduce pregnant women’s propensity to deliver in a facility. Consistent with this expectation, ASHAs in the self-tracking condition do improve two key antenatal health practices: attending antenatal visits and obtaining immunizations. Alternatively, the results may be biased due to selection on both registering clients and reporting their data. Finally, there may be unobserved beneficial (and detrimental) effects of the self-tracking app that are not captured in the current data.

These explanations cannot currently be disentangled, due to the limited (in both extent and quality) nature of the outcomes data that ASHAs self-report. A much more thorough survey of clients is ongoing and will enrich the analysis of effects of the self-tracking app on health.

### **3.5.6 Causal mechanisms**

The experimental data offer suggestive evidence that, with the caveat that no aggregate health benefit is observed, the high-powered intrinsic incentive—the self-tracking app—is more powerful at stimulating effort than the low-powered incentive, and that it this effect is increasing in intrinsic motivation. The latter interaction effect supports the hypothesis that the treatment effect is mediated by leveraging pre-existing intrinsic motivation. Nevertheless, in this section, I briefly discuss other mechanisms that might explain the treatment effect. In particular, as the utility function (Equation 3.2.1) makes clear, an incentive can increase observed output through three channels: by increasing the marginal intrinsic utility, marginal extrinsic utility, and productivity of effort.



As regards implicit extrinsic incentive effects, two empirical results discussed above cast doubt on the possibility that the treatment effect on client visits is driven by monetary preferences: first, that ASHAs in the self-tracking condition earn no more than their counterparts in the generic encouragement condition, and second, that those who are more extrinsically motivated respond no more strongly to the self-tracking app than those who are less extrinsically motivated.

Besides a financial incentive, another embedded extrinsic incentive could arise if ASHAs interpret the self-tracking app as a monitoring tool or as a way to make good performance more visible to their employer. As suggestive evidence, nine months into the experiment, ASHAs across both treatments were asked whether their direct supervisors, the auxiliary nurse midwife (ANM) cadre, ever viewed their apps. ASHAs interact with their supervising ANM at least monthly for village health and nutrition days, and the ANMs could conceivably use either app, but especially the self-tracking app, as a supervisory tool. In fact, the proportion of ASHAs who report that their ANM views their app is balanced between the self-tracking (20.5%) and generic encouragement (26.4%) conditions.

Finally, in addition to increasing marginal returns to effort, an incentive can also increase returns for a given level of effort by altering an agent's production function such that the effort cost of producing a given level of output is reduced. These alterations could include a reminder effect that decreases the effort required to remember to do a task. Although plausible in principle, this reminder mechanism is inconsistent with the design of both the self-tracking and generic encouragement apps. That is, the software is "pull"-based in that the ASHA must access the software intentionally; there is no "pushing" of app content to the ASHAs. The software is therefore unlikely to function as a reminder given that the ASHA is already thinking about ASHA work when she accesses it.

## 3.6 Conclusion

It is not from the benevolence of the butcher, the brewer, or the baker that we expect our dinner, but from their regard for their own interest. We address ourselves not to their humanity, but to their self-love. . . .

—ADAM SMITH, *The Wealth of Nations*, 1776

Adam Smith’s canonical passage about the power of self-interest in creating efficient market outcomes has yet to be refuted. But as Smith (1759) himself argued, humans are also driven by motivations other than self-interest. Perhaps in no economic setting is this more true than in the prosocial sector, where agents often cite intrinsic rewards arising from enjoyable, purposeful work as a principal job benefit. Given this motivation, there may be value in addressing ourselves not only to their self-love, but also to their humaneness.<sup>35</sup>

This experiment provides evidence that a technology designed to leverage intrinsic motivation does indeed enhance intrinsic effort, in a manner that fits neatly into a standard utility maximization framework. The experiment sets forth a conceptual framework for thinking about extrinsic and intrinsic motivation and incentives, and it demonstrates the feasibility of developing and testing intrinsic incentives for public service delivery in the field. I now conclude by posing two broad sets of questions for further research.

First, in what ways do intrinsic incentives interact with preferences and institutional contexts? For example, this paper has argued that the effect of an intrinsic incentive depends in part on the distribution of intrinsic motivation in the agent pop-

---

<sup>35</sup>The observation applies to the demand side context as well. For example, most people are intrinsically motivated to be healthy, but efforts to engage in healthy behaviors often fall short due to self-control problems. At a time when demand-side financial incentives such as conditional cash transfers are gaining popularity (Rawlings, 2005; Volpp et al., 2009), there may be a role for intrinsic incentives.

ulation. But it may also depend on other factors, especially the background extrinsic incentive regime, for which, in this experiment, there is no variation to enable identification. Are extrinsic and intrinsic rewards substitutes or complements? What is their relative cost-effectiveness? As much as the cross-effect of extrinsic incentives on intrinsic motivation has been studied, what is the effect of intrinsic incentives on extrinsic motivation? Can intrinsic incentives crowd out or crowd in extrinsic motivation?

Second, what other intrinsic incentives might be effective? The theoretical framework posits that there is a function  $\psi(\cdot)$  that modulates intrinsic returns to effort. The self-tracking app is designed to increase marginal intrinsic utility by making the effort task more salient. Hence, task salience may belong in the  $\psi$  function. But other incentive technologies may as well, such as an organizational culture that fosters a sense of prosocial meaning, a production system that encourages social bonding, or a gamification technology that makes mundane work take on the aspect of a game. And the effect of these technologies is likely to interact with heterogeneous preferences. Thus, although a simple device, the  $\psi(\cdot)$  function points to a rich line of inquiry—dating at least as far back as the ancient Greek concepts of *hedonia* and *eudaimonia*—on what makes us happy. While this question has historically been the purview of other disciplines (including the humanities), it also lies at the heart of economics—implicating, as it does, the very notion of utility and the preferences that shape it.

# Bibliography

- Abowd, J. M. & Kramarz, F. (1999). The analysis of labor markets using matched employer-employee data. *Handbook of labor economics*, 3, 2629–2710.
- Aghion, P. & Tirole, J. (1997). Formal and Real Authority in Organizations. *Journal of Political Economy*, 105(1), 1–29.
- Alatas, V., Banerjee, A., Hanna, R., Olken, B., & Purnamasari, R. (2015). Self-targeting: Evidence from a field experiment in indonesia. *Journal of Political Economy*. *forthcoming*.
- Allcott, H. (2011). Social Norms and Energy Conservation. *Journal of Public Economics*, 95(9-10), 1082–1095.
- Amabile, T. M. et al. (1994). The work preference inventory: Assessing intrinsic and extrinsic motivational orientations. *Journal of Personality and Social Psychology*, 66(5), 950–967.
- Angrist, J. & Lavy, V. (2009). The effects of high stakes high school achievement awards: Evidence from a randomized trial. *The American Economic Review*, 99(4), 1384–1414.
- Ariely, D., Kamenica, E., & Prelec, D. (2008). Man’s search for meaning: The case of Legos. *Journal of Economic Behavior & Organization*, 67(3-4), 671–677.
- Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, 63(4), 596.
- Aron, A. et al. (2004). Including others in the self. *European Review of Social Psychology*, 15(1), 101–132.
- Ashraf, N., Bandiera, O., & Jack, B. K. (2014). No margin, no mission? A field experiment on incentives for public service delivery. *Journal of Public Economics*, 120, 1–17.

- Ashraf, N., Bandiera, O., & Lee, S. S. (2015). Do-gooders and Go-Getters: Career Incentives, Selection, and Performance in Public Service Delivery.
- Ashraf, N., Berry, J., & Shapiro, J. M. (2010). Can higher prices stimulate product use? evidence from a field experiment in zambia. *American Economic Review*, 100(5), 2383–2413.
- Azmat, G. & Iriberry, N. (2010). The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics*, 94(7), 435–452.
- Bandiera, O., Barankay, I., & Rasul, I. (2013). Team incentives: Evidence from a firm level experiment. Forthcoming, Journal of the European Economic Association.
- Banerjee, A. V., Glennerster, R., & Duflo, E. (2008). Putting a Band-Aid on a Corpse: Incentives for Nurses in the Indian Public Health Care System. *Journal of the European Economic Association*, 6(2-3), 487–500.
- Baqui, A., Kumar, V., Mohanty, S., Kumar, A., Misra, R. P., Santosham, M., & Awasthi, S. (2008). Effect of community-based behaviour change management on neonatal mortality in Shivgarh, Uttar Pradesh, India: A cluster-randomised controlled trial. *The Lancet*, 372(9644), 1151–1162.
- Barankay, I. (2012). Rank incentives: Evidence from a randomized workplace experiment. Working Paper, Wharton School, University of Pennsylvania.
- Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, 9(1-2), 9–30.
- Barrick, M. R., Stewart, G. L., & Piotrowski, M. (2002). Personality and job performance: Test of the mediating effects of motivation among sales representatives. *Journal of Applied Psychology*, 87(1), 43–51.
- Basinga, P., Gertler, P. J., Binagwaho, A., Soucat, A. L. B., Sturdy, J., & Vermeersch, C. M. J. (2011). Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation. *Lancet*, 377(9775), 1421–8.
- Bellah, R. N. et al. (1988). *Habits of the Heart: Individualism and Commitment in American Life*. University of California Press, Berkeley, CA.

- Bénabou, R. & Tirole, J. (2002). Self-confidence and personal motivation. *The Quarterly Journal of Economics*, 117(3), 871–915.
- Bénabou, R. & Tirole, J. (2003). Intrinsic and Extrinsic Motivation. *Review of Economic Studies*, 70(3), 489–520.
- Bénabou, R. & Tirole, J. (2006). Incentives and Prosocial Behavior. *American Economic Review*, 96(5), 1652–1678.
- Bénabou, R. & Tirole, J. (2011). Identity, morals, and taboos: Beliefs as assets. *Quarterly Journal of Economics*, 126(2), 805–855.
- Berglas, S. (1985). Self-handicapping and self-handicappers: A cognitive/attributional model of interpersonal self-protective behavior. *Perspectives in Personality*, 1, 235–270.
- Besley, T. & Ghatak, M. (2005). Competition and incentives with motivated agents. *American Economic Review*, 95(3), 616–636.
- Besley, T. & Ghatak, M. (2008). Status incentives. *The American Economic Review*, 98(2), 206–211.
- Bhandari, N., Paul, V. K., Sachdev, H. S., Mavalankar, D., Ramachandran, P., Sankar, M. J., & Sreenivas, V. (2011). Reproductive health, child health and nutrition in India: Meeting the challenge. *The Lancet*, 377, 332–349.
- Bruhn, M. & McKenzie, D. (2009). In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal: Applied Economics*, 1(4), 200–232.
- Brunnermeier, M. & Parker, J. A. (2005). Optimal expectations. *The American Economic Review*, 95(4), 1092–1118.
- Callen, M., Gulzar, S., Hasanain, A., Khan, Y., & Rezaee, A. (2015). Personalities and Public Sector Performance: Evidence from a Health Experiment in Pakistan. (pp. 1–69).
- Campbell, O. M. R. & Graham, W. J. (2006). Strategies for reducing maternal mortality: getting on with what works. *Lancet*, 368(9543), 1284–99.
- Charness, G. & Grosskopf, B. (2001). Relative payoffs and happiness: An experimental study. *Journal of Economic Behavior and Organization*, 45, 301–328.

- Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K., & Rogers, F. H. (2006). Missing in Action: Teacher and Health Worker Absence in Developing Countries. *Journal of Economic Perspectives*, 20(1), 91–116.
- Clingingsmith, D., Khwaja, A. I., & Kremer, M. (2009). Religion and Tolerance in India's Global Gathering. *Quarterly Journal of Economics*, 124(3), 1133–1170.
- Dal Bó, E., Finan, F., & Rossi, M. A. (2013). Strengthening state capabilities: The role of financial incentives in the call to public service. *The Quarterly Journal of Economics*, 128(3), 1169–1218.
- Das, J. & Hammer, J. (2007). Money for nothing: The dire straits of medical practice in Delhi, India. *Journal of Development Economics*, 83(1), 1–36.
- Das, J. & Hammer, J. (2014). Quality of Primary Care in Low-Income Countries: Facts and Economics. *Annual Review of Economics*, 6, 525–553.
- Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125(6), 627–68.
- Deci, E. L. & Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. Perspectives in Social Psychology. New York: Plenum Press.
- Deserranno, E. (2014). Financial incentives as signals: Experimental evidence from the recruitment of health workers. Working paper.
- Dewatripont, M., Jewitt, I., & Tirole, J. (1999). The economics of career concerns, part i: Comparing information structures. *Review of Economic Studies*, 66, 183–198.
- Dewey, K. G. & Adu-Afarwuah, S. (2008). Systematic review of the efficacy and effectiveness of complementary feeding interventions in developing countries. *Maternal & Child Nutrition*, (pp. 24–85).
- Duckworth, A. L. et al. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92(6), 1087–1101.
- Duflo, E., Hanna, R., & Ryan, S. P. (2012). Incentives work: Getting teachers to come to school. *The American Economic Review*, 102(4), 1241–1278.
- Dur, R., Bradler, C., Neckermann, S., & Non, A. (2013). Employee recognition and performance: A field experiment. ZEW Discussion Paper No. 13-017.

- Dwenger, N., Kleven, H., Rasul, I., & Rincke, J. (2014). Extrinsic and Intrinsic Motivations for Tax Compliance: Evidence from a Field Experiment in Germany.
- Fehr, E. & Schmidt, K. (1999). A theory of fairness, competition and cooperation. *Quarterly Journal of Economics*, 114(3), 817–868.
- Ferraz, C. & Finan, F. (2011). Motivating politicians: The impacts of monetary incentives on quality and performance.
- Food and Nutrition Technical Assistance Project (2011). Anthropometry: Assessing children under 5 pocket reference.
- Freeman, R. & Gelber, A. (2010). Prize structure and information in tournaments: Experimental evidence. *American Economic Journal: Applied Economics*, 2(1), 149–64.
- Fryer, R. (2011). Financial incentives and student achievement: Evidence from randomized trials. *The Quarterly Journal of Economics*, 126, 1755–1798.
- Fryer, R. G. (2013). Teacher incentives and student achievement: Evidence from New York City public schools. *Journal of Labor Economics*, 31(2), 373–407.
- Gagliarducci, S. & Nannicini, T. (2011). Do better paid politicians perform better? disentangling incentives from selection. *Journal of the European Economic Association*.
- Gebauer, J. & Lowman, D. (2008). *Closing the Engagement Gap: How Great Companies Unlock Employee Potential for Superior Results*. Towers, Perrin, Forster & Crosby, Inc.
- Glewwe, P., Ilias, N., & Kremer, M. (2010). Teacher Incentives. *American Economic Journal: Applied Economics*, 2(July), 205–227.
- Government of India (2011a). *Annual Health Survey 2010-11*. Technical report.
- Government of India (2011b). *India Population and Housing Census 2011*. Technical report.
- Government of India (2015). ASHA Status of Selection and Training.
- Government of Zambia (2010). *National Community Health Worker Strategy in Zambia*. Technical report, Ministry of Health.



- Government of Zambia (2011). *National Human Resources for Health Strategic Plan 2011–2015*. Technical report, Ministry of Health.
- Government of Zambia (2014). *Zambia Demographic and Health Survey 2013-14: Preliminary Report*. Technical report.
- Grant, A. M. (2007). Relational Job Design and the Motivation To Make a Prosocial Difference. *Academy of Management Review*, 32(2), 393–417.
- Grant, A. M. (2008). Does intrinsic motivation fuel the prosocial fire? motivational synergy in predicting persistence, performance, and productivity. *Journal of Applied Psychology*, 93(1), 48–58.
- Hays, R. D., Hayashi, T., & Stewart, A. L. (1989). A Five-item Measure of Socially Desirable Response Set. *Education and Psychological Measurement*, 49, 629–636.
- Holzer, H., Katz, L. F., & Krueger, A. B. (1991). Job queues and wages. *The Quarterly Journal of Economics*, 106(3), 739–768.
- International Institute for Population Sciences & Macro International (2007). *National Family Health Survey (NFHS-3), 2005-06: India: Volume I*. Technical report.
- Jones, E. & Berglas, S. (1978). Control of attributions about the self through self-handicapping strategies: The appeal of alcohol and the role of underachievement. *Personality and Social Psychology Bulletin*, 4(2), 200–206.
- Kandel, E. & Lazear, E. (1992). Peer pressure and partnerships. *Journal of Political Economy*, 100(4), 801–817.
- Kelley, H. (1971). *Attribution in Social Interaction*. Morristown, New Jersey: General Learning Press.
- Kling, J. R., Liebman, J. B., Katz, L. F., & Sanbonmatsu, L. (2004). Moving to Opportunity and Tranquility: Neighborhood Effects on Adult Economic Self-Sufficiency and Health from a Randomized Housing Voucher Experiment. *Working Paper*, (August).
- Kolstad, J. R. & Lindkvist, I. (2013). Pro-social preferences and self-selection into the public health sector: evidence from an economic experiment. *Health policy and planning*, 28(3), 320–7.

- Kolstad, J. T. (2013). Information and Quality When Motivation is Intrinsic: Evidence from Surgeon Report Cards. *American Economic Review*, 103(7), 2875–2910.
- Köszegi, B. (2002). Emotional agency. *The Quarterly Journal of Economics*, 121(1), 121–155.
- Köszegi, B. (2006). Ego utility, overconfidence, and task choice. *Journal of the European Economic Association*, 4(4), 673–707.
- Kremer, M., Miguel, E., & Thornton, R. (2009). Incentives to learn. *The Review of Economics and Statistics*, 91(3), 437–456.
- Lagarde, M. & Blaauw, D. (2013). Pro-social preferences and self-selection into rural jobs: Evidence from South African nurses.
- Lazear, E. (1989). Pay equality and industrial politics. *Journal of Political Economy*, 97(3), 561–580.
- Lazear, E. & Oyer, P. (2012). Personnel economics. *The Handbook of Organizational Economics*, (pp. 479).
- Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review*, 90(2), 1346–1361.
- Lee, S. S. (2014). A Field Experiment on Leveraging Intrinsic Motivation in Public Service Delivery.
- Lehmann, U. & Sanders, D. (2007). Community health workers: What do we know about them? the state of the evidence on programmes, activities, costs and impact on health outcomes of using community health workers. *World Health Organization*, 2, 1–42.
- Leonard, K. L., Masatu, M. C., & Vialou, A. (2013). Getting Doctors to Do Their Best: The Roles of Ability and Motivation in Health Care Quality. *The Journal of Human Resources*, 42(3), 682–700.
- Leuven, E., Oosterbeek, H., & van der Klaauw, B. (2010). The effect of financial rewards on students' achievement: Evidence from a randomized experiment. *Journal of the European Economic Association*, 8(6), 1243–1265.
- Lim, S. S., Dandona, L., Hoisington, J. a., James, S. L., Hogan, M. C., & Gakidou, E. (2010). India's Janani Suraksha Yojana, a conditional cash transfer programme to increase births in health facilities: an impact evaluation. *Lancet*, 375(9730),

2009–23.

- Maestad, O., Torsvik, G., & Aakvik, A. (2010). Overworked? On the relationship between workload and health worker performance. *Journal of Health Economics*, 29(5), 686–98.
- Major, B., Testa, M., & Bylsma, W. H. (1991). *Social Comparison: Contemporary Theory and Research*. Hillsdale, NJ: Erlbaum.
- Marinescu, I. & Wolthoff, R. (2013). Wages, applications, and skills.
- Mashek, D., Cannaday, L., & Tangney, J. (2007). Inclusion of community in self scale: A single-item pictorial measure of community connectedness. *Journal of Community Psychology*, 35, 257–275.
- Mathauer, I. & Imhoff, I. (2006). Health worker motivation in Africa: The role of non-financial incentives and human resource management tools. *Human Resources for Health*, 4(1), 24.
- Milkovich, G. & Newman, J. (1996). *Compensation*. Homewood, IL: Irwin, 4 edition.
- Miller, G. et al. (2012). Effectiveness of provider incentives for anaemia reduction in rural china: a cluster randomised trial. *BMJ: British Medical Journal*, 345.
- Moldovanu, B. & Sela, A. (2001). The optimal allocation of prizes in contests. *The American Economic Review*, 91, 542–558.
- Moldovanu, B., Sela, A., & Shi, X. (2007). Contests for status. *Journal of Political Economy*, 115, 338–363.
- Müller, W. & Schotter, A. (2010). Workaholics and dropouts in organizations. *Journal of the European Economic Association*, 8(4), 717–743.
- Muralidharan, K. & Sundararaman, V. (2011). Teacher performance pay: Experimental evidence from India. *Journal of Political Economy*, 119(1), 39–77.
- Mwangome, M. K., Fegan, G., Mbunya, R., Prentice, A. M., & Berkley, J. A. (2012). Reliability and Accuracy of Anthropometry Performed by Community Health Workers Among Infants Under 6 Months in Rural Kenya. *Tropical Medicine and International Health*, 17(5), 622–629.
- Myatt, M., Khara, T., & Collins, S. (2006). A Review of Methods to Detect Cases of Severely Malnourished Children in the Community for their Admission into

- Community-based Therapeutic Care Programs. *Food and Nutrition Bulletin*, 27(3 Supplement), S7–S23.
- Nalbantian, H. & Schotter, A. (1997). Productivity under group incentives: An experimental study. *The American Economic Review*, 87(3), 314–341.
- Oster, E., Shoulson, I., & Dorsey, E. R. (2013). Optimal expectations and limited medical testing: Evidence from huntington disease. *American Economic Review*, 103(2).
- Oyer, P. & Schaefer, S. (2011). Chapter 20 - personnel economics: Hiring and incentives. In *Handbook of Labor Economics*, volume 4. Elsevier.
- Rasul, I. & Rogger, D. (2015). Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service.
- Rawlings, L. B. (2005). Evaluating the Impact of Conditional Cash Transfer Programs. *The World Bank Research Observer*, 20(1), 29–55.
- Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 3(2), 136–146.
- Schein, E. H. (1985). *Organizational Culture and Leadership: A Dynamic View*. J-B US non-Franchise Leadership. San Francisco: Jossey-Bass Inc.
- Schmuck, P. & Schultz, W., Eds. (2002). *Psychology of Sustainable Development*, chapter Inclusion with Nature: The Psychology of Human-Nature Relationships, (pp. 62–78). Kluwer Academic.
- Schultz, P. W. (2002). Inclusion with nature: The psychology of human-nature relations. *Psychology of Sustainable Development*, (pp. 61–78).
- Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, 18(5), 429–34.
- Smith, A. (1759). *The Theory of Moral Sentiments*. London: A. Millar.
- Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. Oxford: Clarendon Press.
- Spencer, M., Rosland, A.-M., Kieffer, E. C., Sinco, B., Valerio, M., Palmisano, G., Anderson, M., Guzman, J. R., & Heisler, M. (2011). Effectiveness of a community

- health worker intervention among African American and Latino adults with type 2 diabetes: A randomized controlled trial. *American Journal of Public Health*, 101(12), 2253–2260.
- Staiger, D. O. & Rockoff, J. E. (2010). Searching for effective teachers with imperfect information. *The Journal of Economic Perspectives*, 24(3), 97–117.
- Stone, R. (2004). Self-handicapping, tournaments, and ego utility. *Unpublished manuscript*.
- Tran, A. & Zeckhauser, R. (2012). Rank as an inherent incentive: Evidence from a field experiment. *Journal of Public Economics*, 96, 645–650.
- Volpp, K. G., Troxel, A. B., Pauly, M. V., Glick, H. a., Puig, A., Asch, D. a., Galvin, R., Zhu, J., Wan, F., DeGuzman, J., Corbett, E., Weiner, J., & Audrain-McGovern, J. (2009). A randomized, controlled trial of financial incentives for smoking cessation. *The New England journal of medicine*, 360(7), 699–709.
- Wageman, R. (2001). How leaders foster self-managing team effectiveness: Design choices versus hands-on coaching. *Organization Science*, 12(5).
- Weick, K. E. (1995). *Sensemaking in Organizations*. Thousand Oaks, CA: Sage Publications.
- World Health Organization (2006). The world health report 2006: working together for health. geneva: World health organization.
- Wrzesniewski, A., Mccauley, C., Rozin, P., & Schwartz, B. (1997). Jobs, Careers, and Callings: People’s Relations to Their Work. *Journal of Research in Personality*, 33(31), 21–33.
- Yariv, L. (2002). I’ll see it when I believe it—A simple model of cognitive consistency. Cowles Foundation Discussion Paper No. 1352.

# Appendix A

## Do-Gooders and Go-Getters

### A.1 Differences in sorting vs. differences in recruitment

#### A.1.1 Methodology

The goal of this section is to assess whether CHAs in treatment and control differ because career incentives attract different types, because recruitment panels choose different candidates, or both. To do so, we first test whether applicants differ along the dimensions discussed in Section 1.5.1 and compare them to the candidates chosen by the recruitment panels. To aid the comparison, we also test whether recruitment panels put different weights on these traits when choosing which candidates to nominate.

Recruitment panels have five members: the district health official, a representative from the health post's associated health center, and three members of the local neighborhood health committee. Recruitment panels are exposed to the salience policy as they see the same posters as the candidates. This notwithstanding, they know much more about the actual job attributes and who would be suitable for the posi-

tions. Indeed, contrary to the applicants (whose only source of information was the recruitment poster), the two more senior panel members—the district health official and the health center representative—are employees of the Ministry of Health, and hence familiar with career progression rules regardless of salience policy. The salience policy treatment is likely not as powerful, or perhaps entirely moot, for them.<sup>1</sup>

Table A.5 presents descriptive statistics for the 1585 candidates who interviewed for the CHA jobs (Part I) and for the 334 candidates who are chosen by the panels (Part II). The final 314 CHA trainees differed from the 334 nominees in two ways: (i) to obtain gender balance, GRZ replaced all male nominees (i.e., men ranked 1 or 2 by the interview panels) with female reserves (i.e., women ranked 3 to 5) when available, resulting in 68 changes (22% of the total), and (ii) 13 applicants who were ranked “top 2” declined, and were replaced by reserves. By the time training commenced, twenty spots remained empty.

The data is drawn from MoH’s administrative data on the applicants’ high school test scores and from a survey that we asked candidates to fill in at the interview stage. We mostly use the same measures as in Table 1.8, except for the psychometric scales that were too complex to be administered at the interview stage. As in Table 1.8, we report mean values in the two treatment groups and the p-value of the difference from a regression of the outcome of interest on the career treatment and the stratification variables, with errors clustered at the level of randomization, the district. To shed light on the differences between Parts I and II, Table A.6 estimates the probability that candidate  $i$  in health post  $h$  is chosen by the recruitment panels as follows:

---

<sup>1</sup>Further analysis, available upon request, shows that treatment does not affect panel composition.

$$s_{ih} = \sum_{j \in J} \alpha_j^c C_h X_i^j + \sum_{j \in J} \alpha_j^s (1 - C_h) X_i^j + \sum_{j \in J} \beta_j \bar{X}_h^j + \gamma N_h + \zeta_{ih}$$

where  $s_{ih} = 1$  if  $i$  is one of the two nominated candidates and 0 otherwise; and  $C_h$  equals 1 if health post  $h$  is in the career incentives treatment and 0 if it is in the control group.  $X_i^j$  are individual characteristics, and the set  $J$  includes variables that are affected by salience policy (skills, pro-social preferences, career preferences) as well as age and gender, as GRZ requires giving preference to women. The coefficients of interest are  $\alpha_j^c$  and  $\alpha_j^s$ , which measure the weight given to trait  $j$  in the career and control groups, respectively. Differences, if any, could be due to the fact that panels think that a given trait is more important for a career (community) job, or to the fact that panels in the two treatments face different pools. To account for this, we control for the average traits of the applicants in the same health post  $\bar{X}_h^j$  for all  $j \in J$ . To measure the strength of competition, we include the number of interviewed candidates in the same health post  $N_h$ . As in earlier specifications, we control for the stratification variables and cluster standard errors at the district level. Table 1.4 A.6 reports the estimates of  $\alpha_j^c$  and  $\alpha_j^s$  for all  $j \in J$  and the p-value of the test of equality. We estimate the model with and without the characteristics of the applicant pool  $\bar{X}_h^j$ .

### A.1.2 Results

Table A.5, Panel A.I shows that making career incentives salient attracts more qualified candidates; thus, the differences we see among CHAs in Table 1.8 are at least partly due to differences in the applicant pools. Applicants in the career treatment have a higher total score (p=.019), and have a stronger scientific background (p=.006), which is directly relevant to medical practice. Table A.6 shows that the strongest determinant of appointment is ability in both treatment and control groups: panels



are between 17 and 23 percentage points more likely to appoint candidates at the top of the O-level exam score distribution within their health post. In the average health post, 21% of candidates are appointed; being at the top of the O-level exam score distribution doubles the probability of being selected. Panel A.II, Table A.5, confirms that the recruitment process screened in the most skilled applicants, as both total scores and the number of O-levels in science are higher for the selected CHAs than they are for the average applicant, and the difference between treatments is not precisely estimated. Recruitment panels were thus able to reduce differences in observable measures of skill, but as we know from Table 1.8, unobservable differences remained and CHAs recruited with career incentives had significantly higher test scores during the training program.

Panel B reports motivations and preferences. We see that the differences in career ambitions reported in Table 1.8 were already present in the applicant pool. Panel B.I shows that the share of applicants who aspire to be in a highly-ranked position (environmental health technician, clinical officer, or doctor) within the Government in 5-10 years' time is higher in the career treatment. The difference between treatment and control groups is 6 percentage points (32% of the control group mean) and precisely estimated ( $p=.026$ ). Our main measure of social preferences at the interview stage is based on the adapted "Inclusion of Others in Self (IOS) scale" Aron et al. (2004), which measures the extent to which individuals perceive community and self-interest as overlapping. IOS has been validated across a wide variety of contexts, and adapted versions are found to be strongly correlated with environmental behavior Schultz (2002) and connectedness to the community Mashek et al. (2007). The measure is coded as 0-1, where 1 implies highest overlap.<sup>2</sup> Panel B.I shows that

---

<sup>2</sup>Applicants are asked to choose between four pictures, each showing two circles (labeled "self" and "community") with varying degrees of overlap, from non-overlapping to almost completely over-

84% of the applicants in both treatments perceive their interests to be aligned with the community's, suggesting that career incentives do not displace this type of pro-social preference in the applicant pool. Table A.6 shows that recruitment panels in both treatment and control are more likely to appoint applicants with career ambitions and with pro-social preferences. As a consequence, appointed candidates in Panel B.II have both stronger career ambitions and stronger pro-social preferences. The differences between treatment and control reflect the differences in the applicant pool, and these in turn determine the differences we observe in Table 1.8: CHAs in the treatment group have stronger career ambitions, but the same level of pro-sociality.

Interestingly, panels face no trade-off between skills, career ambitions and pro-sociality in either group. In particular, applicants with top O-level scores have stronger career ambitions and the same level of pro-sociality, and this holds in both the treatment and control group. Similarly, there is no trade-off between career ambitions and pro-sociality in either group.

Turning to demographics, Panel C.I shows no difference in either gender or age in the applicant pool, in contrast with the fact that selected CHAs in the treatment group are older and more likely to be male. Table A.6 shows that recruitment panels in both treatment and control are about 9pp more likely to appoint women as directed by GRZ, yet the share of women drops by 2pp from applicant to nominated candidates in the treatment group and increases by 5pp in the control group. To shed light on this we note that recruitment panels in the two groups face a different trade-off between gender and skills: among the candidates with top O-level scores, the share of women is 25% in the control group and 17% in the treatment group ( $p=.025$ ). This creates a difference in gender balance between nominated candidates that gets further lapping. This variable equals 1 if the respondent chooses the almost completely overlapping picture, 0 otherwise.

reinforced by MoH's affirmative action policy, bringing the share of women among deployed candidates to 44% in the treatment group and 57% in the control group, as seen in Table 1.8 . Regarding age, Table A.6 shows that this is the only dimension where panels seem to differ: treatment panels put a small positive weight on age (1 SD increase in age increases the probability of nomination by 7pp) while control panels do not, and the difference is precisely estimated. The trade-off between age and skill is also different in the two groups as applicants with top O-level scores are younger in the control group (25.7 vs 26.5,  $p=.09$ ) but not in the treatment group. Taken together, these imply that nominated and selected CHAs in the treatment group are on average one year older than those in the control group.

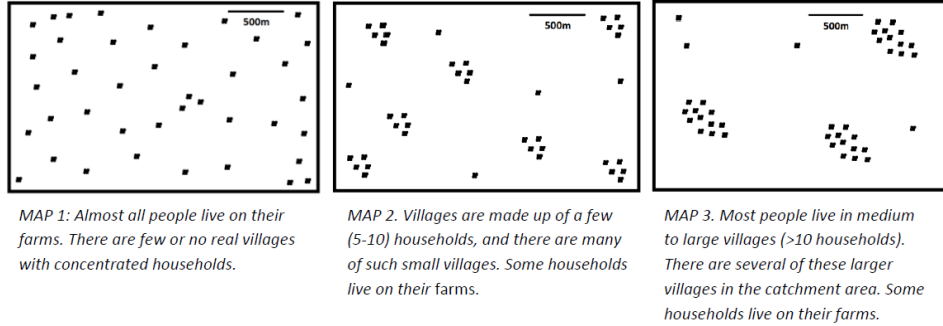
Ultimately, the evidence in this Section shows that career incentives attract applicants who differ on the key dimensions of skill and career ambition, but not the weight that recruitment panels put on these attributes, so that appointed CHAs differ on these traits because they came from different pools, rather than having been chosen differently by the recruitment panels.

## A.2 Data Appendix

In this section, we describe each of the variables used in our analysis, including its source, unit of measurement, and data source. Because we used a number of different data sources, we describe each of them below. We collect data at each stage of the program: application, selection, training, and performance in the field. Each variable indicates which data source it is generated from. A description of each source, including the sample, can be found in Section A.2.5.

### A.2.1 Eligible population and catchment area characteristics

- *Number of staff in health post* (source: district health officials survey, by phone)
  - Total number of nurses, environmental health technicians, and clinical officers assigned to the health post, as reported by district health officials we surveyed by phone.
- *Geographical distribution of households in catchment area* (CHA survey, in person, at refresher training) - CHAs were shown stylized maps accompanied by the description above and asked to choose the one that most closely resembled the catchment area of their health post. Questions were asked to each CHA individually so that two CHAs from the same health post could give different answers. For the 5 out of 161 cases in which the two CHAs gave different answers, we used the information provided by supervisors to break the tie.



- *Poor cell network coverage* (source: attempted phone calls) - We attempted to call all CHAs after deployment. We made daily calls for 118 consecutive days. The health post was classified as having poor coverage if we did not manage to reach either of its two CHAs during this period.

## A.2.2 Experiment Validation

- Relative weight variables are derived from a survey question (CHA survey, in person, at training) that asked the trainees to allocate 50 beans between different potential motivations for applying to the CHA position: “good future career,” “allows me to serve the community,” “earns respect and high status in the community,” “pays well,” “interesting job,” “allows me to acquire useful skills,” and “offers stable income.”
- *Expects to be employed in MoH in 5-10 years* (source: CHA survey, in person, at interview) - Circled any combination of being a “Community Health Worker,” “nurse,” “environmental health technician,” “clinical officer,” or “doctor” in response to the question, “When you envision yourself in 5-10 years’ time, what do you envision yourself doing?”

## A.2.3 Performance in Service Delivery

### Household Visits

Source: SMS Receipts

- *Unique households visited*
- *Number of visits per household*
- *Average visit duration, in minutes*

MINISTRY OF HEALTH HOUSEHOLD VISIT RECEIPT		CHA ID: <input type="text"/>	<input type="text"/>	<input type="text"/>		
1	START TIME:	<input type="text"/>	:	<input type="text"/>		
	END TIME:	<input type="text"/>	:	<input type="text"/>		
	DATE:	<input type="text"/>	/	<input type="text"/>	/	20
2	Client's Name	<input type="text"/>				
	Client's Village	<input type="text"/>	<input type="text"/>	<input type="text"/>		
	Household ID	<input type="text"/>	<input type="text"/>	<input type="text"/>		
	Client's Phone Number (if available)	<input type="text"/>	<input type="text"/>	<input type="text"/>		
3	I, the Client, certify that this receipt is truthful and accurate.	<input type="text"/>				
	CLIENT'S SIGNATURE	<input type="text"/>				

Source: HMIS (monthly reports)

Each reported variable is the sum of each indicator's monthly values from September 2012 to January 2014.

- *Number of households visited*
- *Number of women and children visited per household visit*
- *Number of patients seen at HP*
- *Number of community mobilization meetings*

## Time Use

Source: CHA survey, in person, at refresher training

- *Number of hours worked in a typical week* - CHAs were asked “In a typical week, how many total hours do you spend doing CHA work? Please count work that you do at the health post and in the village, including moving from household to household.”
- *Frequency of out-of-hours calls in a typical week* - CHAs were asked “In a typical week, how often do you have to leave your house at night and do CHW work due to emergencies like a pregnancies or accidents?” Possible responses were “5-7 days per week,” “3-4 days per week,” “1-2 days per week,” “2-3 times per month,” “Once per month,” “Sometimes, but less than once per month,” and “Never.”
- *Share of time allocated to* - To obtain time allocations, CHAs were asked to allocate 50 beans between different activities. The instructions were as follows:  
*Please use the beans to show how much time you spend doing each activity. If you spend more time in an activity, you should place more beans on the card. If you never do an activity, you should place no beans on the card. Place the beans any way you would like. For instance, you can place all beans on one card, or 0 beans on any card.*  
*Household visits - Now I would like you to think about household visits specifically. Here are some cards that list different activities you may do during household visits.*
  - *greeting household members*
  - *assessing and referring sick household members*

- reviewing and discussing the household’s health profile and goals
- asking questions about household health behaviors and knowledge
- providing health counseling
- doing household inspections (waste disposal, latrines, etc.)
- documentation (filling registers/books and sending visit receipts via SMS)

*Health Post - Now here are some cards that list different activities you may do at the HEALTH POST OR RURAL HEALTH center.*

- seeing sick patients at the OPD
- dispensing medications from the pharmacy
- helping with ANC visits
- cleaning and maintaining the facility
- assisting with deliveries and other procedures when needed
- documentation (filling registers/books and sending monthly reports through HMIS)

*In the Community - Now here are some cards that list different activities you may do as a CHA.*

- campaigns for polio, measles, child health, and other health issues
- health talks and other community mobilization activities
- school health talks and other school activities
- meeting with NHC and volunteer CHWs for planning



## A.2.4 CHAs' observable traits

### Skills

- *Average test score at training [0-100]*- Average score in 11 tests on basic medical practices taken during the training program.
- *O-levels total exam score* (source: MOH application files) - This variable is constructed as the sum of inverted O-levels scores (1=9, 2=8, and so on) from all subjects in which the applicant wrote the exam, so that larger values correspond to better performance.
- *O-levels passed in biology and other natural sciences* (source: MOH application files) - Includes biology, chemistry, physics, science and agricultural science.

### Applicants' Preferences and Motivations

- *Donation to local hospital (dictator game)* (source: baseline survey) - In the modified dictator game, trainees were given 25,000 Kwacha (approximately USD 5; half of a CHA's daily earnings) and invited to donate any portion (including nothing) to the local hospital to support needy patients. This donation decision occurred privately and confidentially in concealed donation booths. Previous work has found dictator games adapted for specific beneficiary groups predictive of performance on pro-social tasks Ashraf et al. (2014) and choices of public sector nurses to locate to rural areas Lagarde & Blaauw (2013).

I am happy to inform you that we have recently received a small donation from an outside donor to support the Community Health Assistants. In a moment, you will each receive an equal portion of this outside donation.

While the money is yours to keep, the donor has also requested that we provide you with an opportunity for you to share this gift with the community. This is an opportunity to support people in this community who are sick but are unable to afford the health care that they need. As you know, there are many such people in the communities from where you come from and also here in Ndola. They get sick, but because they are very poor, they are not able to get the health care that they need.

Because we want to protect your privacy, we have set up a donation booth in the next room. There you will see a collection box where you can deposit your donation, if you choose to donate. You do not have to give anything if you don't want to. No one here will know if you decide not to give anything. Your donation will be recorded, but we will not have access to this information. Once everyone has had an opportunity to give, IPA will collect any donations made to this cause, and we will donate the total amount to Ndola Central Hospital to directly support patients who are unable to pay for their medicines and treatment.

In a moment, we will give you the money, and you will come to this desk where you will be able to donate to help needy patients if you wish.

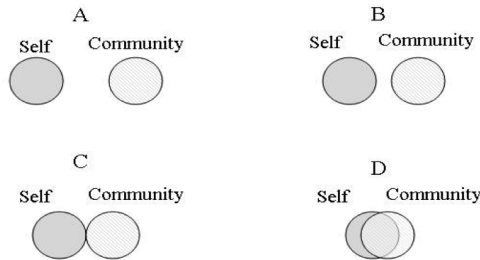
I am happy to announce now that the donor is able to provide each of you with 25,000 Kwacha.

In a moment, I will ask each of you to come to the registration table one-by-one. When you come to the table, that is when I will give you the money. I will also give you an envelope in case you want to support the patients at Ndola Central Hospital.

If you want to give any amount of money to help needy patients in the community, place the money in the envelope. Then seal the envelope, and place that envelope in the "Help Needy Patients in the Community" box. Please be sure to place the money INSIDE the envelopes before placing it in the cash box. Do not put any loose bills into the cash box. Whatever money you have remaining, you can keep in your main envelope.

- *Main goal is "service to community" vs. "career advancement"* (source: baseline survey) - Asked of all trainees: "In terms of your new CHA position, which is more important to you?" with two possible responses: "serving community" and "promoting career."
- *Perceives community interests and self-interest as overlapping* (source: CHA survey, in person, at interview) - Based on the "Adapted Inclusion of Others in Self (IOS) scale" Aron et al. (2004) which measures the extent to which individuals perceive community- and self-interest as overlapping. The Inclusion of Other in the Self scale was originally designed by Dr. Art Aron and colleagues (Aron et al. (1992)) as a measure of self-other inclusion and relationship closeness. The Continuous IOS makes use of the basic design of the original IOS, but allows for (a) the measure to be embedded within a web-based questionnaire, (b) the output values to be continuously scaled, and (c) modifications in the appearance and behavior of the measure. IOS has been validated across a wide variety of contexts, and adapted versions are found to be strongly correlated with environmental behavior (Inclusion of Nature in the Self, Schmuck & Schultz (2002)) and connectedness to the community (Inclusion of Community

in Self, Mashek et al. (2007)). The measure is coded as 0-1, where 1 implies highest overlap. Applicants are asked to choose between sets of pictures, each showing two circles (labeled “self” and “community”) with varying degrees of overlap, from non-overlapping to almost completely overlapping. This variable equals 1 if the respondent chooses the almost completely overlapping picture (D), 0 otherwise.



- *Aims to be a higher-rank health professional in 5-10 years* (source: CHA survey, in person, at interview) - Circled any combination of being an “environmental health technician,” “clinical officer,” or “doctor” in response to the question, “When you envision yourself in 5-10 years’ time, what do you envision yourself doing?”

### Psychometric Scales

Each measure (source: baseline survey) takes on a value between 1 and 5 and represents, among the statements listed below, the extent to which the applicant agreed, on average. Levels of agreement are 1 (strongly disagree), 2 (disagree), 3 (neither agree nor disagree), 4 (agree), and 5 (strongly agree). The psychometric scales came from validated scales used in employment surveys on pro-social motivation and career orientation. Each variable is the average of the item scores within each psychometric scale. For instance, in a scale with three items, the variable value equals the sum of

levels of agreement for all items divided by three. It represents the average level of agreement with the included items.

- *Career orientation* - Adapted from Wrzesniewski et al. (1997). In contrast to *Calling* below, individuals with high career orientation tend to have a deeper personal investment in their work and mark their achievements not only through monetary gain, but through advancement within the occupational structure. This advancement often brings higher social standing, increased power within the scope of one's occupation, and higher self-esteem for the worker Bellah et al. (1988). This scale consists of the following items: "I expect to be in a higher-level job in five years," "I view my job as a stepping stone to other jobs," and "I expect to be doing the same work as a CHA in five years" (reverse-scored).
- *Pro-social motivation (pleasure-based)* - Adapted from Grant (2008) and consists of the following items: "Supporting other people makes me very happy," "I do not have a great feeling of happiness when I have acted unselfishly" (reverse-scored), "When I was able to help other people, I always felt good afterwards," and "Helping people who are not doing well does not raise my own mood" (reverse-scored).
- *Desire for positive pro-social impact* - Adapted from Grant (2008). This measure provides an index of the degree to which an individual desires and benefits psychologically from the positive impact of her work on others. The scale consists of the following items: "It is important to me to do good for others through my work," "I care about benefiting others through my work," "I want to help others through my work," "I want to have positive impact on others through my work," "I get motivated by working on tasks that have the potential to benefit others," "I like to work on tasks that have the potential to benefit others," "I prefer to

work on tasks that allow me to have a positive impact on others,” “I do my best when I’m working on a task that contributes to the well-being of others,” “It is important to me to have the opportunity to use my abilities to benefit others,” “It is important to me to make a positive difference in people’s lives through my work,” “At work, I care about improving the lives of other people,” and “One of my objectives at work is to make a positive difference in other people’s lives.”

- *Affective commitment to beneficiaries* - Adapted from Grant (2008) and answers the following question: “How much do I care about/committed to the beneficiaries of my work?” The scale consists of the following items: “The people who benefit from my work are very important to me,” and “The people who benefit from my work matter a great deal to me.”

### A.2.5 Data Sources

- **Source: Application** (sample: all applicants) - Applications were submitted from August-September 2010. The initial application stage was comprised of the initial application form, which includes fields for gender, date of birth, village of residence, educational qualifications, and previous health experience (position, organization, start and end years). The application form also included a question asking through what means the applicant first learned of the CHA job opportunity: recruitment poster, facility health worker, community health worker, government official, word-of-mouth, or “other.”
- **Source: Interview Candidate Questionnaire** (sample: subset of applicants called for an interview) - Ranking questionnaires were filled and collected from September to October 2010. If applicants met the basic criteria noted above, they were invited for interviews, and asked to complete a questionnaire on the interview

day. The questionnaire (written in English) included a series of questions about the interviewee's demographic background, community health experience, social capital, and work preferences and motivations. Notably, we included a measure employed by social psychologists, "Inclusion of Others in Self" from Aron et al. (2004) to measure connection with the community. The questionnaire stated that the answers would not be used for selection purposes but rather as part of a research project, although we cannot rule out that panelists could have seen the questionnaire or referred to it when making their decisions.

- **Source: Ranking Sheet** (sample: members of interview panels) - Ranking sheets were filled and collected from September to October 2010. Each panel consisted of five members: the district health officer, a representative from the health center, and three neighborhood health committee members. Once all interviews were completed, every member of the selection panel completed a private and individual ranking sheet by ranking their top ten candidates. This ranking exercise occurred *before* panel members formally deliberated and discussed the candidates. After interviewing all candidates and deliberating, interview panels were requested to complete and submit a consensus-based "Selection Panel Report" that included fields for the two nominated candidates as well as three alternates.
- **Source: Baseline Survey** (sample: all trainees) - The baseline survey was conducted in June 2011 and consisted of five components:
  1. Questionnaire- Conducted one-on-one by a surveyor and collected information on the trainees' socio-economic background and livelihoods, previous experience with health work, motivations to apply, and expectations of the program.

2. Psychometric scales- A self-administered written exercise which gathered alternative information on motivations to apply, determinants of job satisfaction, and other character traits.
3. Modified dictator game- An experimental game whereby students received a small donation and were given the opportunity to give some of it back for a good cause. It explored the altruistic nature of the students.
4. Coin game- An experimental game that explored the risk-taking behavior of the students.
5. Self-assessment- A three-hour exam with multiple choice questions to determine the knowledge on health matters that each student had prior to the training.

- **Source: Catchment Area Survey** (sample: all deployed CHWs and supervisors) - Just prior to graduation in July 2012, all CHWs and supervisors were given a short survey that asked about characteristics of their health posts, including population density, rainy-season information, and general community health measures.
- **Source: Time Use Survey** (sample: all deployed CHWs) - This survey was conducted in April/May 2013 in Ndola, Zambia. The respondents were pilot CHAs who reported to Ndola for a supplemental in-service training to introduce new tasks as part of a revised CHA scope of work. The survey was administered by Innovations for Poverty Action, in partnership with the Ministry of Health, the CHA Training School, and the Clinton Health Access Initiative.
- **Source: SMSs** (sample: all deployed CHAs) - All CHAs carry with them receipt books for each visit, which require the signature of the client visited. The information on these receipts—consisting of the data, time, and duration of the

visit, as well as the client’s phone number—is then SMS’ed in real time to the MoH and our central data-processing facility. 5% of these visits are audited.

### A.3 District Instruction Appendix

The CHA program was introduced differently to health centers depending on the treatment group. In each district, the district health official was given a package that contained a script, a memo from the Permanent Secretary, and detailed instructions about the CHA recruitment process. In addition, district health officials received “health center packages” for each participating health center in the district, which contained a set of posters and application forms and instructions for the health center representative on how to post posters and collect applications. The district health officials were to visit each health center and meet with the staff and neighborhood health committee members to introduce the program and distribute the health center packages, using the script provided to them in their packages. The script was only provided to the district health officials, and was addressed directly to them. It is unlikely that the applicants or health center staff were able to read this script themselves.

The following script was given to district health officials in the career-incentives treatment group:

*To Health center and Neighborhood Health Committee: I would like to you let you know about a new government program to strengthen the country’s health workforce. Applications are currently being accepted for a new Community Health Worker position. This is an opportunity for qualified Zambians to obtain employment and to advance their health careers. Opportunities for training to advance to positions such as Nurse and Clinical*



*Officer may be available in the future. Successful applicants will receive 1 year of training, both theoretical and practical. All training costs, including transportation, meals and accommodation during the one-year training program, will be covered by the Ministry of Health. Please encourage all qualified persons to apply so that they can benefit from this promising career opportunity.*

The district health officials in the control group received the following script:

*To Health center and Neighborhood Health Committee: I would like to you let you know about a new government program to improve health care services in your community. Applications are currently being accepted for a new Community Health Worker position. This is an opportunity for local community members to become trained and serve the health needs of their community. The new CHWs will work at the Health Post and community level in coordination with an affiliated Health center. Successful applicants will receive 1 year of training, both theoretical and practical. All training costs, including transportation, meals and accommodation during the one-year training program, will be covered by the Ministry of Health. Please encourage all qualified persons to apply so that they can benefit from this promising community service opportunity.*

Table A.1: Experimental checks: reasons to apply

	<b>treatment</b>	<b>control</b>	p-value of the difference
Weight given to the following reasons at the application stage [0,1]			
Good future career	.165 (.157)	.120 (.112)	.002
Pays well	.031 (.092)	.025 (.057)	.442
Interesting job	.150 (.162)	.152 (.140)	.784
Allows to acquire useful skills	.181 (.168)	.160 (.136)	.214
Allows to serve the community	.396 (.226)	.432 (.239)	.050
Earns respect and status in the community	.037 (.094)	.057 (.109)	.048
Offers stable income	.027 (.057)	.024 (.054)	.469
Expects to be employed in MoH in 5-10 years	.924 (.022)	.900 (.026)	.728

Notes: Columns 1 and 2 show means and standard deviations in parentheses, Column 3 reports the p-value of the test of equality of means based on standard errors clustered at the district level. All variables are drawn from a survey administered at the beginning of the training program. To measure the "Weight given to the following reasons," CHAs were given 50 beans and asked to allocate them on cards, listing different reasons in proportion to the importance of each reason when applying. The cards were scattered on a table in no particular order.

Table A.2: Validation of household visit measures

dependent variable	Number of visits from HMIS records		"=1 if HH reports a visit by CHA"		HH satisfaction: level of CHA's knowledge		HH satisfaction: level of CHA's caring		HH satisfaction: level of CHA's effort		HH satisfaction: overall CHA's services	
	HMIS	HH survey	HH	HH survey	HH	HH survey	HH	HH survey	HH	HH survey	HH	HH survey
source	HMIS	HH survey	HH	HH survey	HH	HH survey	HH	HH survey	HH	HH survey	HH	HH survey
level	Health post	HH	HH	HH	HH	HH	HH	HH	HH	HH	HH	HH
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Number of visits (in 00s) reported by CHA between 9/12 and 1/14	0.767*** (0.0672)	0.0210** (0.00832)	0.0342** (0.0149)	0.0394** (0.0189)	0.0358* (0.0210)	0.0402** (0.0194)						
Mean (SD) of dependent variable	643.6	0.438	4.302	4.314	4.309	4.329						
Adjusted R-squared	0.473	0.014	0.012	0.012	0.011	0.014						
N	145	1286	1214	1254	1247	1255						

Note: OLS estimates, standard errors clustered at the health post level in Columns 2-6. The dependent variable in Column 1 is the total number of visits done by the two CHAs in the health post drawn from HMIS administrative data. The dependent variables in Columns 2-6 are drawn from a HH survey administered to 16 HHs in each of in 47 communities where CHAs are active. Satisfaction measures range from 1 (very dissatisfied) to 5 (very satisfied).

Table A.3: Treatment effect on time use

<i>Panel A: Time allocation during household visits</i>						
share of time allocated to:	(1)	(2)	(3)	(4)	(5)	(6)
counseling		inspections	filing in receipts and forms	asking questions about health behaviors and knowledge	discussing health profile and goals	visiting sick household members
Career incentives	.006 (.012)	.007 (.015)	-.016 (.010)	-.011 (.009)	-.003 (.012)	.010 (.009)
Mean of dependent variable in social treatment	0.207	0.196	0.146	0.137	0.122	0.100
Area characteristics	no	no	no	no	no	no
R-squared	.030	.041	.049	.026	.014	.027
N	292	292	292	292	292	292

<i>Panel B: Time allocation during work at the health post</i>						
share of time allocated to:	(1)	(2)	(3)	(4)	(5)	(6)
seeing sick patients		filling in forms	dispensing medications	helping with ante natal care visits	cleaning and maintaining the health post	
Career incentives	-.002 (.011)	-.050*** (.018)	.006 (.012)	.019 (.019)	.019 (.013)	
Mean of dependent variable in social treatment	0.262	0.228	0.207	0.160	0.104	
Area characteristics	no	no	no	no	no	
R-squared	.051	.104	.091	.095	.133	
N	271	271	271	271	271	

**Notes:** System estimates (SURE), bootstrapped standard errors clustered at the district level in parenthesis. All regressions include the stratification variables (province dummies and share of high school graduates in the district). All 298 participants in the refresher training program were given 50 beans and asked to allocate the beans to show how much time they spent doing each activity within each task. They were instructed to place more beans on a card if they spent more time on an activity, to place no beans if they never do an activity, and to place the beans any way they would like, including placing all beans on one card, or 0 beans on any card. Panel A activities are: greeting household members, assessing and referring sick household members, reviewing and discussing the household's health profile and goals, asking questions about health behaviors and knowledge, providing health education and counseling, doing household inspections (waste disposal, latrines, etc.), and documentation (filling registers/books and sending SMS visits). The omitted category in Panel A is "greetings." The sample in Panel A covers the 292 out of 298 CHAs who reported spending time doing visits. Panel B activities are: seeing sick patients in the health post, dispensing medications from the pharmacy, helping with ANC visits, cleaning and maintaining the facility, assisting with deliveries and other procedures when needed, and documentation (filling registers/books and sending monthly reports through DHIS2). The omitted category in Panel B is "assisting with deliveries." The sample in Panel B covers the 271 out of 298 CHAs who reported spending time at the health post. Area characteristics include: number of staff in the health post, geographical distribution of households in the catchment area, and an indicator variable that equals 1 if the CHA reports to have good cell network coverage most of the time or all the time.

Table A.4: Psychometric tests

	treatment	control	p-values
<b>Average Scores:</b>			
Social Desirability	.353 (.019)	.397 (.022)	.100
Autonomy	2.244 (.048)	2.102 (.046)	.065
Internal Motivation	4.392 (.055)	4.372 (.063)	.851
Extrinsic Motivation	3.189 (.039)	3.230 (.038)	.215
Intrinsic Motivation	3.706 (.031)	3.749 (.034)	.448
Calling Orientation	4.049 (.040)	4.063 (.041)	.451
Status Striving	3.502 (.063)	3.412 (.054)	.305
Accomplishment Striving	4.285 (.033)	4.332 (.036)	.148
Consistent Interest	2.266 (.051)	2.255 (.055)	.589
Grit	2.083 (.036)	2.063 (.039)	.477
Persistent Effort	1.900 (.046)	1.887 (.048)	.734
Proactive Personality	3.582 (.056)	3.591 (.056)	.820
Personal Prosocial Identity	4.257 (.049)	4.319 (.051)	.375
Company Prosocial Identity	4.382 (.049)	4.502 (.043)	.030
Perceived Prosocial Impact	4.090 (.053)	4.141 (.055)	.303
Perceived Antisocial Impact	1.678 (.068)	1.701 (.073)	.698
Perceived Social Worth	4.100 (.057)	4.087 (.066)	.830

Notes: Scores are calculated as averages of a series of questions scaled 1 to 5, except for Social Desirability (RAND). Autonomy scales are taken from questions in Wageman, 1995. Internal Motivation is from Edmonson, 1999. Extrinsic Motivation and Intrinsic Motivation are from Amabile et al., 1994. Calling Orientation is from Wreszniewski et al., 1997. Status Striving, and Accomplishment Striving are from Barrick et al., 2002. Consistent Interest, Grit, and Persistent Effort are from Duckworth et al., 2007. Proactive personality is from Claes et al., 2005. Personal Prosocial Identity and Company Prosocial Identity are from Grant et al., 2008. Perceived Prosocial Impact, Perceived Antisocial Impact, and Perceived Social worth are from Grant et al., 2008b/c.

Table A.5: Applicants vs. nominated candidates

	Part I: Applicants (N=1585)		Part II: Nominated Candidates (N=334)	
	treatment	control	treatment	control
	p-values		p-values	
<b>Panel A: Skills</b>				
O-levels total exam score	24.8 (9.81)	23.3 (9.35)	27.14 (10.95)	25.65 (8.81)
		.019		.173
O-levels passed in biology and other natural sciences	1.44 (.858)	1.24 (.888)	1.55 (.890)	1.47 (.805)
		.006		.642
<b>Panel B: Motivation and preferences</b>				
Aims to be a higher-rank health professional in 5-10 years	.246 (.431)	.188 (.391)	.354 (.479)	.253 (.436)
		.026		.031
Perceives community interests and self-interest as overlapping	.839 (.367)	.842 (.364)	.865 (.342)	.887 (.317)
		.975		.685
<b>Panel C: Demographics and socio-economic status</b>				
Gender (=1 if female)	.292 (.016)	.294 (.016)	.273 (.447)	.345 (.477)
		.458		.237
Age	26.0 (.200)	26.2 (.205)	27.85 (6.63)	26.64 (5.99)
		.745		.040

Notes: For each part of the Table, Columns 1 and 2 show means and standard deviations in parentheses while Column 3 reports the p-values of the null hypothesis that the career treatment effect equals zero conditional on stratification variables and with standard errors clustered at the district level. Ordinary levels or O-levels are administered by the Examinations Council of Zambia (ECZ) to 12th-grade students, the highest grade in the Zambian secondary education system. O-levels total exam score is constructed as the sum of inverted O-levels scores (1=9, 2=8, and so on) from all subjects in which the applicant wrote the exam, so that larger values correspond to better performance. O-levels passed in biology and other natural sciences equals the number of O-levels passed in biology, chemistry, physics, science and agricultural science. Aims to be a higher-rank health professional in 5-10 years: equals 1 if the candidate chooses any combination of being an "environmental health technician," "clinical officer," or "doctor" in response to the question, "When you envision yourself in 5-10 years' time, what do you envision yourself doing?" Perceives interests as overlapping: Adapted Inclusion of Others in Self scale (Aron et al., 2004). Applicants are asked to choose between sets of pictures, each showing two circles (labeled "self" and "community") with varying degrees of overlap, from non-overlapping to almost completely overlapping. This variable equals 1 if the respondent chooses the almost completely overlapping picture, 0 otherwise.

Table A.6: Effect of career incentives on candidate selection by panels

	=1 if nominated	p-value	=1 if nominated	p-value
High relative exam score X treatment	0.235*** (0.0405)		0.201*** (0.0399)	
High relative exam score X control	0.174*** (0.0369)	.256	0.148*** (0.0349)	.304
Aims to be a higher-rank health worker in 5-10 years X treatment	0.111*** (0.0363)		0.138*** (0.0404)	
Aims to be a higher-rank health professional in 5-10 years X control	0.0778** (0.0309)	.489	0.109*** (0.0378)	.565
Perceives interests as overlapping X treatment	0.0203 (0.0397)		0.0126 (0.0446)	
Perceives interests as overlapping X control	0.0981** (0.0392)	.169	0.0729* (0.0388)	.266
Female X treatment	0.0913** (0.0354)		0.113*** (0.0393)	
Female X control	0.0854*** (0.0311)	.901	0.0926*** (0.0361)	.689
Age X treatment	0.0125*** (0.00355)		0.0134*** (0.00423)	
Age X control	0.00320 (0.00280)	.043	0.00463 (0.00301)	.079
Number of interviewees in health post	-0.0103*** (0.00358)		-0.00256 (0.00338)	
Applicant pool controls				
Adjusted R-squared	0.149		0.141	
N	1269		1230	

Notes: OLS estimates. All regressions include the stratification variables (province dummies and share of high school graduates in the district) and standard errors clustered at the district level. Independent variables are interacted with each treatment (social and career incentives). High relative exam score equals 1 if the applicant's exam score is one of the 3 highest (4 in case of tie) among applicants to the same health post. Aims to be a higher-rank health professional in 5-10 years equals 1 if the candidate chooses any combination of being an "environmental health technician," "clinical officer," or "doctor" in response to the question, "When you envision yourself in 5-10 years' time, what do you envision yourself doing?". Perceives interests as overlapping: Adapted Inclusion of Others in Self scale (Aron et al., 2004). Applicants are asked to choose between sets of pictures, each showing two circles (labeled "self" and "community") with varying degrees of overlap, from non-overlapping to almost completely overlapping. This variable equals 1 if the respondent chooses the almost completely overlapping picture, 0 otherwise. Number of interviewees in health post: total candidates interviewed per health post. Applicant pool controls include the following variables, all computed over applicants to the same health post: top 3 (4 in case of tie) exam scores; the share of applicants who aims to be a higher-rank health professional in 5-10 years; the share of applicants who perceive interests as overlapping; the share of applicants who are female; the average age.

Table A.7: The effect of career incentives on aggregate health outputs-robustness checks

	Panel A. Placebo test						
	dependent variable: total over each quarter 2011:1-2014:2 women giving birth at postnatal (0-6 weeks) the health center						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	women giving birth at the health center	visits	children under 5 visited	children under 5 weighed	children under 1 receiving BCG vaccinations	children under 1 receiving polio vaccinations	children under 1 receiving measles vaccinations
Career incentives	-3.700 (11.18)	-12.96 (9.580)	-17.60 (175.7)	-18.42 (157.7)	11.92 (13.55)	-4.133 (10.58)	2.792 (10.34)
After	0.705 (4.055)	15.23** (4.617)	89.93 (78.79)	152.6** (74.58)	2.451 (5.338)	3.685 (4.130)	-3.144 (3.838)
Career incentives*After	12.33** (5.409)	8.964 (9.920)	369.4** (120.3)	339.6** (140.0)	8.490 (10.28)	12.18** (5.590)	13.16 (8.518)
Placebo After	7.191** (2.277)	-0.0192 (4.663)	-65.29 (82.07)	-97.48 (68.14)	-8.328* (4.936)	-10.79* (5.579)	3.733 (4.517)
Career incentives*Placebo After	4.554 (5.778)	-1.037 (5.441)	-108.4 (139.0)	-117.7 (135.7)	-2.230 (7.921)	6.373 (8.089)	-3.138 (8.711)
Area characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of dependent variable in control in year 1	46.7	49.9	1312.8	1261.5	89.8	73.9	73.6
Adjusted R-squared	0.336	0.203	0.245	0.247	0.147	0.144	0.107
Number of facilities	89	118	123	123	121	120	121
Number of observations	1269	1528	1618	1610	1518	1530	1535
Panel B. Health post fixed effects.							
dependent variable: total over each quarter 2011:1-2014:2 women giving birth at postnatal (0-6 weeks) the health center							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	women giving birth at the health center	visits	children under 5 visited	children under 5 weighed	children under 1 receiving BCG vaccinations	children under 1 receiving polio vaccinations	children under 1 receiving measles vaccinations
After	5.082 (4.194)	15.46** (5.205)	63.77 (65.98)	106.9 (65.64)	-1.094 (4.786)	-1.299 (3.893)	-1.024 (3.844)
Career incentives*After	13.48** (6.485)	8.906 (9.933)	306.7** (107.1)	278.8** (118.9)	8.349 (8.843)	15.40** (5.325)	11.58 (7.826)
Area characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of dependent variable in control in year 1	46.7	49.9	1312.8	1261.5	89.8	73.9	73.6
Adjusted R-squared	0.819	0.664	0.618	0.591	0.496	0.565	0.445
Number of facilities	89	118	123	123	121	120	121
Number of observations	1269	1528	1618	1610	1518	1530	1535

Notes: OLS Estimates, standard errors clustered at the district level. Data source is the Health Management and Information System (HMIS) available monthly from January 2011 until June 2014. Health center and health post staff are required to submit monthly reports that summarize their activities at the health post/community level. These are aggregated at the quarter level in the regressions. The variable in Column (1) is defined at the health center level because health centers are equipped for child births and health posts are not. The variables in columns (2)-(7) are defined at the health post level if this reports data at the health center otherwise. After=1 after September 2011, halfway through the period before the CHAs started working, Placebo After=1 after September 2011, halfway through the period before the CHAs started working. All regressions include the stratification variables (province dummies and share of high school graduates in the districts). Area characteristics include: number of staff in the health post, geographical distribution of households in the catchment area, and an indicator variable that equals 1 if the CHA reports to have good cell network coverage most of the time or all the time.



# Appendix B

## Awards Unbundled

Figure B.1: Outline of one-year Community Health Assistant training curriculum

Course	Module	Teaching Days	Topics
0	1: Health care system in Zambia	4	<ul style="list-style-type: none"> <li>• Organization and functions of the health care system in Zambia</li> <li>• Roles and responsibilities of a Community Health Assistant</li> <li>• Code of conduct</li> <li>• Gender equality</li> <li>• Community mobilization and networking</li> </ul>
	2: Behavioral sciences	19	<ul style="list-style-type: none"> <li>• Introduction to psychology</li> <li>• Mental health and common psychiatric conditions</li> <li>• Introduction to sociology</li> <li>• Family and community</li> </ul>
	3: Health promotion	9	<ul style="list-style-type: none"> <li>• Introduction to health promotion</li> <li>• Communication skills</li> </ul>
	4: Environmental health	31	<ul style="list-style-type: none"> <li>• General principles of infection prevention</li> <li>• Water supply</li> <li>• Excreta disposal</li> <li>• Solid waste management</li> <li>• Food hygiene and safety</li> <li>• Housing and health</li> <li>• Insect and rodent control</li> </ul>
1	5: Epidemiology	34	<ul style="list-style-type: none"> <li>• Infectious disease epidemiology</li> <li>• Epidemic investigation and management</li> <li>• Epidemiological surveillance</li> <li>• Data collection</li> </ul>
	6: Reproductive and child health	35	<ul style="list-style-type: none"> <li>• Introduction to reproductive health</li> <li>• Introduction to child health</li> <li>• School health services</li> </ul>
2	7: Anatomy and physiology	20	<ul style="list-style-type: none"> <li>• Introduction to the human body</li> <li>• Musculoskeletal system</li> <li>• Cardiovascular system</li> <li>• Respiratory system</li> <li>• Digestive system</li> <li>• Urinary system</li> <li>• Special senses</li> </ul>
3	8: Basic procedures	20	<ul style="list-style-type: none"> <li>• Occupational safety and health</li> <li>• Lifting and moving patients</li> <li>• History and physical exam skills</li> <li>• Assessment of hygiene, nutrition, physical activity, pain, and vital signs</li> <li>• Wound care</li> <li>• Palliative care</li> </ul>
	9: Common medical conditions	38	<ul style="list-style-type: none"> <li>• Common conditions (malaria, diarrhea, respiratory infection, HIV, tuberculosis, anemia, etc.)</li> <li>• Oral health</li> </ul>
4	10: Diagnostic procedures	4	<ul style="list-style-type: none"> <li>• Malaria rapid diagnostic testing</li> <li>• HIV testing</li> <li>• Sputum collection for TN testing</li> </ul>
	11: First aid	9	<ul style="list-style-type: none"> <li>• Principles of first aid</li> <li>• Bandaging</li> <li>• Lifting and moving patients</li> <li>• Cardio-pulmonary resuscitation</li> <li>• Handling of selected emergencies (toxic ingestion, bites and stings, fractures, burns, drowning, foreign body ingestion)</li> </ul>

Figure B.2: Control group

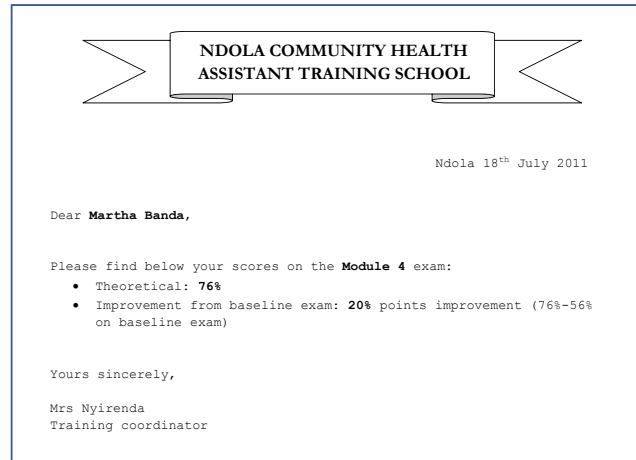
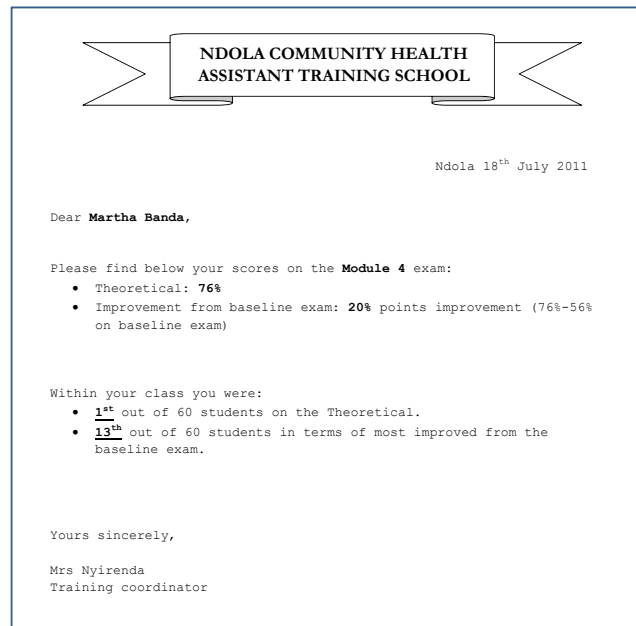


Figure B.3: Treatment #1



All names are fictional.

Figure B.4: Treatment #2

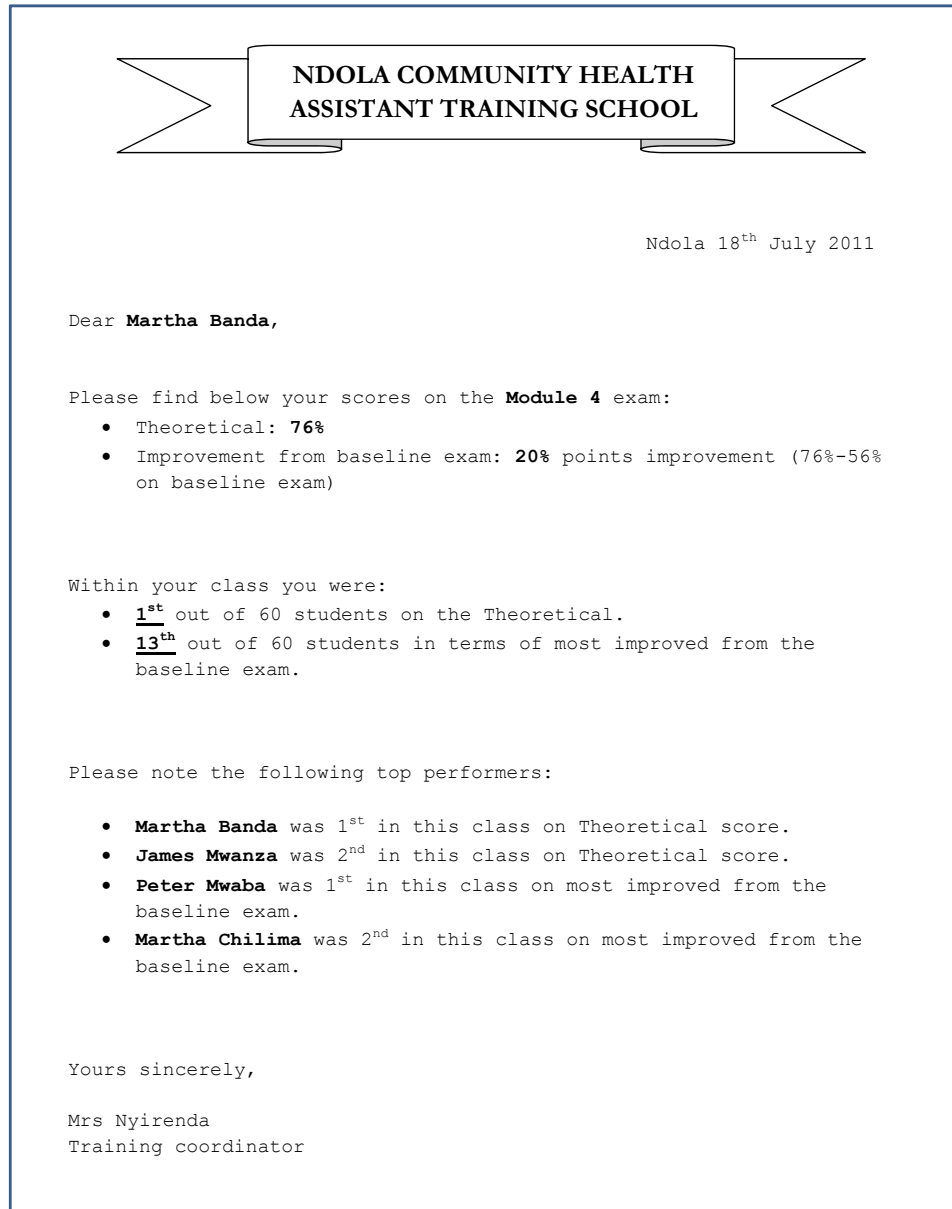


Figure B.5: Treatment #3

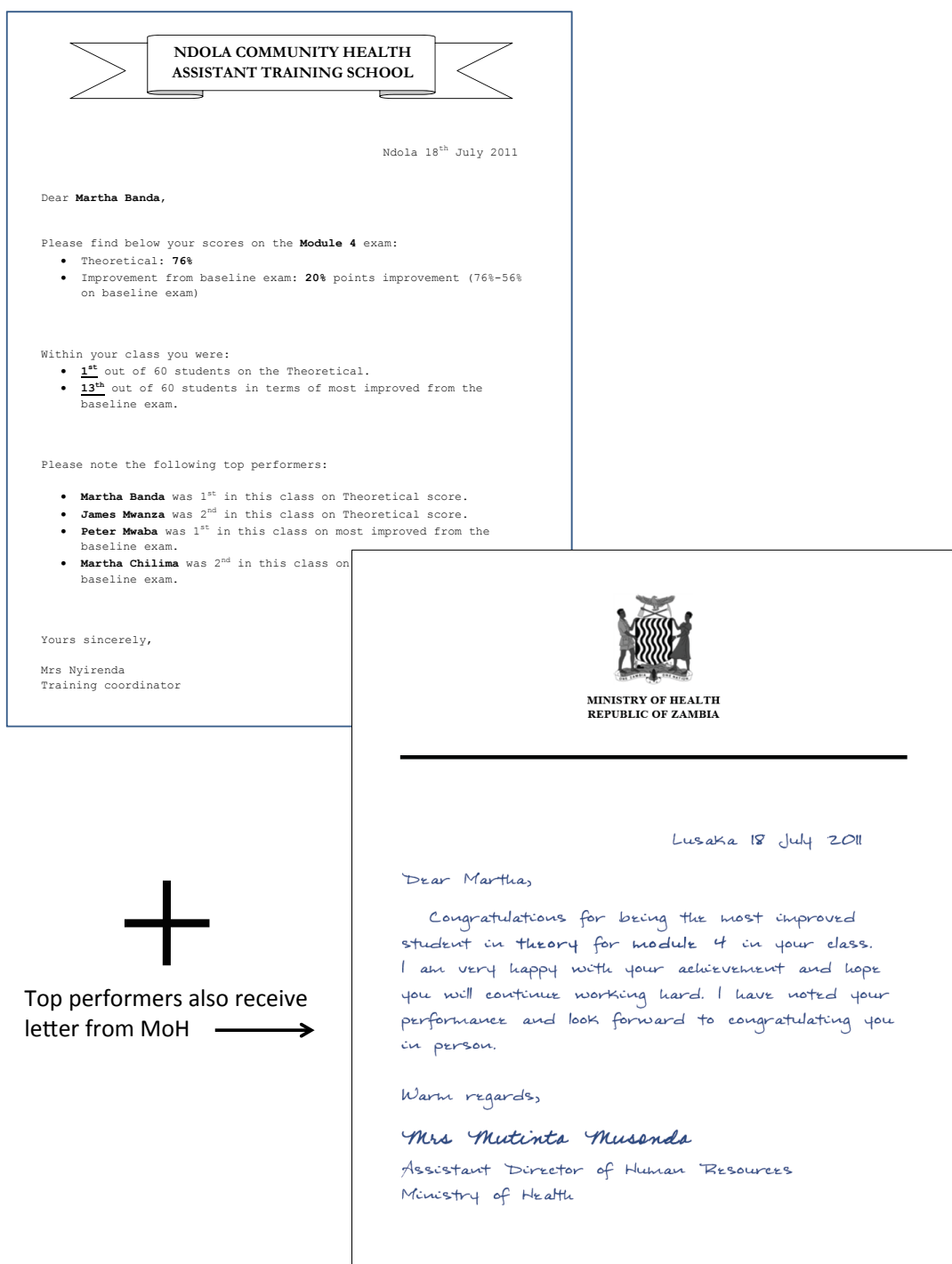


Figure B.6: Treatment #4



Ndola 18<sup>th</sup> July 2011

Dear **Martha Banda**,

Dear **John Banda**,

Please find below your scores on the **Module 4** exam:

- Theoretical: **76%**
- Improvement from baseline exam: **20%** points improvement (76%-56% on baseline exam)

Within your class you were:

- **3<sup>rd</sup>** out of 60 students on the Theoretical.
- **13<sup>th</sup>** out of 60 students in terms of most improved from the baseline exam.

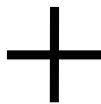
**John Banda** was

Please **Martha Banda** following top performers:

- **Mary Phiri** was 1<sup>st</sup> in this class on Theoretical score.
- **James Mwanza** was 2<sup>nd</sup> in this class on Theoretical score.
- **Peter Mwaba** was 1<sup>st</sup> in this class on most improved from the baseline exam.
- **Martha Chilima** was 2<sup>nd</sup> in this class on most improved from the baseline exam.

Yours sincerely,

Mrs Nyirenda  
Training coordinator



One of top performers also featured in a newsletter sent to their community →

CHA NEWSLETTER | Issue 1

## “Why I want to be a CHA”

By Martha Banda

**LET’S GET TO KNOW MARTHA!**

Martha Banda is a 27-year-old CHA student representing Lubanda Health Post, Itzhi-tezhi District, Southern Province. Her counterpart at this health post is Andrew Phiri.

She is married and a mother of two children. She completed secondary school at Kabulonga Girls High School in 2005. Before joining the CHA program, Martha was volunteering as an HIV/AIDS educator in her community.

*In this section, we wish to provide an opportunity for CHA students to be known to the community. **Martha Banda** from Lubanda Health Post shares her thoughts on why she wants to be a Community Health Assistant.*

Zambia has faced many challenges in its health sector but I have become inspired by the fact that the government wants to train community health workers. I joined

the program because of my passion for patients, my love for others, and because I want to reduce the gap between my people and the hospital. I plan to encourage my community to use healthy resources like boreholes for clean drinking water and community resources such as the health post to learn more about nutritious diets.

I like telling people about health issues and I want to be a Community Health Assistant so that I can help my friends in my community, and so we can work together. I like caring for people who are sick – cooking for

them, washing them, and giving them words of encouragement.

My first step to being a Community Health Assistant, after finishing my training, will be to gain the respect of my community, first and foremost. I will start by talking with my village headman and, after getting his blessing, I will begin to call the community for meetings where I plan to give health talks on how to care for children under-five and educate the community on treating their water with chlorine prior to drinking it. I am looking forward to serving my community!

Martha Banda, age 27 years

Table B.1: Quantile treatment effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Quantile	10th	20th	30th	40th	50th	60th	70th	80th	90th
T1: Private Social Comparison	-0.504*** (0.176)	-0.392*** (0.142)	-0.282* (0.158)	-0.320*** (0.115)	-0.304** (0.135)	-0.256** (0.105)	-0.184** (0.0875)	-0.0635 (0.117)	-0.0118 (0.124)
T2: T1 + Public Social Comparison	-0.704*** (0.171)	-0.445*** (0.140)	-0.292* (0.156)	-0.428*** (0.114)	-0.381*** (0.133)	-0.329*** (0.103)	-0.275*** (0.0857)	-0.174 (0.114)	-0.152 (0.115)
T3: T2+ Employer Recognition Award	0.0681 (0.157)	-0.00494 (0.137)	0.170 (0.152)	0.0644 (0.113)	-0.0805 (0.133)	-0.0786 (0.103)	-0.0341 (0.0862)	0.0105 (0.117)	-0.0440 (0.117)
T4: T2+Social Visibility Award	-0.0257 (0.168)	0.00956 (0.153)	0.0694 (0.172)	0.117 (0.126)	0.0218 (0.147)	-0.0151 (0.115)	0.0161 (0.0965)	0.136 (0.130)	0.0677 (0.135)
Net effect of Public Social Comparison (T2-T1)	-0.200 (0.180)	-0.053 (0.141)	-0.010 (0.156)	-0.108 (0.114)	-0.076 (0.133)	-0.074 (0.104)	-0.090 (0.087)	-0.110 (0.117)	-0.140 (0.128)
Net effect of Employer Recognition Award (T3-T2)	0.772*** (0.164)	0.441*** (0.137)	0.462*** (0.152)	0.492*** (0.111)	0.300** (0.131)	0.251** (0.102)	0.241*** (0.086)	0.184 (0.116)	0.108 (0.118)
Net effect of Social Visibility Award (T4-T2)	0.679*** (0.172)	0.455*** (0.149)	0.362** (0.169)	0.545*** (0.124)	0.402*** (0.145)	0.314*** (0.114)	0.291*** (0.097)	0.309 (0.131)	0.220 (0.138)
Number of observations (trainee-courses)	1149	1149	1149	1149	1149	1149	1149	1149	1149
Adjusted R-squared	0.1528	0.1392	0.1317	0.1335	0.1336	0.1341	0.1363	0.1283	0.1183

Notes: Table reports OLS estimates, weighted by course duration, with standard errors clustered at the trainee level in parentheses. Dependent variable is normalized exam score, normalized by the mean and standard deviation of the control group for each exam. Individual controls include: score in the baseline test for each of the four courses, English test score, gender, age, previous experience in the health sector, employment status at the time of application, district recruitment strategy.

\*\*\* Significant at the 1 percent level.

\*\* Significant at the 5 percent level.

\* Significant at the 10 percent level.

Table B.2: Heterogeneous treatment effects by baseline test score and period

	Bottom tercile of baseline test score			Middle tercile of baseline test score			Top tercile of baseline test score					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Heterogeneous effects by period			Heterogeneous effects by exam period			Heterogeneous effects by exam period			Heterogeneous effects by exam period		
	All periods	First period	Following periods	Test of equality: (2) = (3)	All periods	First period	Following periods	Test of equality: (6) = (7)	All periods	First period	Following periods	Test of equality: (10) = (11)
T1: Private Social Comparison	-0.672** (0.298)	-0.578** (0.263)	-0.737* (0.394)	$p = 0.656$ (0.271)	-0.233 (0.271)	-0.119 (0.286)	-0.316 (0.328)	$p = 0.519$ (0.213)	0.010 (0.213)	-0.018 (0.233)	0.081 (0.234)	$p = 0.618$ (0.234)
T2: T1 + Public Social Comparison	-0.812*** (0.260)	-1.065*** (0.261)	-0.640** (0.314)	$p = 0.124$ (0.199)	-0.179 (0.199)	-0.258 (0.256)	-0.126 (0.209)	$p = 0.562$ (0.203)	-0.113 (0.203)	0.059 (0.226)	-0.225 (0.230)	$p = 0.186$ (0.230)
T3: T2+ Employer Recognition Award	-0.280 (0.216)	-0.146 (0.249)	-0.372 (0.244)	$p = 0.335$ (0.250)	0.157 (0.250)	0.075 (0.329)	0.213 (0.246)	$p = 0.600$ (0.159)	-0.005 (0.159)	0.210 (0.207)	-0.154 (0.170)	$p = 0.060$ (0.170)
T4: T2+Social Visibility Award	-0.102 (0.309)	-0.102 (0.391)	-0.150 (0.336)	$p = 0.891$ (0.223)	0.204 (0.223)	-0.108 (0.275)	0.286 (0.226)	$p = 0.064$ (0.177)	-0.031 (0.177)	-0.203 (0.234)	0.028 (0.191)	$p = 0.294$ (0.191)
Net effect of Public Social Comparison (T2-T1)	-0.140 (0.334)	-0.487 (0.298)	0.097 (0.428)	$p = 0.115$ (0.243)	0.054 (0.243)	-0.138 (0.243)	0.190 (0.320)	$p = 0.303$ (0.214)	-0.153 (0.214)	0.076 (0.229)	-0.306 (0.247)	$p = 0.079$ (0.247)
Net effect of Employer Recognition Award (T3-T2)	0.532** (0.255)	0.919*** (0.278)	0.268 (0.284)	$p = 0.007$ (0.214)	0.336 (0.214)	0.333 (0.300)	0.339 (0.220)	$p = 0.982$ (0.166)	0.107 (0.166)	0.152 (0.200)	0.071 (0.195)	$p = 0.705$ (0.195)
Net effect of Social Visibility Award (T4-T2)	0.710** (0.342)	0.963** (0.408)	0.490 (0.375)	$p = 0.182$ (0.183)	0.383** (0.183)	0.150 (0.222)	0.412** (0.202)	$p = 0.223$ (0.173)	0.082 (0.173)	-0.261 (0.226)	0.253 (0.200)	$p = 0.032$ (0.200)
Trainees controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Number of clusters (trainees)	92	92	92	107	107	107	107	107	107	107	107	107
Number of observations (trainee-courses)	350	350	350	401	401	401	401	398	398	398	398	398
Adjusted R-squared	0.186	0.228	0.228	0.079	0.079	0.111	0.111	0.157	0.157	0.207	0.207	0.207

Notes: Table reports OLS estimates, weighted by course duration, with standard errors clustered at the trainee level in parentheses. Dependent variable is normalized exam score, normalized by the mean and standard deviation of the control group for each exam. Individual controls include: score in the baseline test for each of the four courses, English test score, gender, age, previous experience in the health sector, employment status at the time of application, district recruitment strategy. Columns (2) and (3), (6) and (7), and (10) and (11) report coefficients estimated in the same regression where we include all treatments interacted with an indicator variable that takes value 0 in the first period and 1 thereafter.

\*\*\* Significant at the 1 percent level.

\*\* Significant at the 5 percent level.

\* Significant at the 10 percent level.

# Appendix C

## Intrinsic Incentives

### C.1 Data sources

This experiment draws on several data sources. Data on ASHA performance come from two sources. The first is the CommCare mobile phone application that all ASHAs use, as described in Section 3.3.2. ASHAs self-report client visits, and visits data are at the client-day level. Since each visit is associated with a client, it is possible to track initial vs. follow-up visits to the client. In addition, ASHAs collect and report basic health and demographic data about their clients (expected delivery date, age, parity, marital status). At the conclusion of the client's pregnancy, the ASHA fills out a case completion form reporting the maternal outcome (survival, death), the fetal outcome (miscarriage, stillbirth, live birth), and delivery location (home, health facility).

The other source of performance data is earnings. ASHA payments are recorded in a receipt book that ASHAs keep in their possession. To be paid, the ASHA visits the district hospital and submits a request for payment. Based on delivery records, the payment is transferred into the ASHA's bank account and recorded in the ASHA's



receipt book. The receipt books only record ASHA-related payments (i.e., no personal transactions). The raw data in the receipt books are at the transaction level, but these do not differentiate between payments for deliveries vs. other activities, or for specific deliveries. The compensation data used in this paper are thus aggregated at the month level.

For data on take-up of the experimental interventions, I measure usage directly. The software has the ability to track usage by users in a highly granular manner, with a timestamp recorded for every page view across the different pages in the software. For the analysis in this paper, one unit of take-up is defined as any contiguous series of page-views with a maximum timestamp interval of ten minutes. Thus, if a user views 10 pages every two minutes over the course of 20 minutes, this is counted as one session. If she views two pages over the course of 20 minutes, this is counted as two sessions. The intention behind this rule is to capture how often ASHAs use the experimental intervention, rather than how intensively they do so. However, the results of the experiment are robust to using individual page-views as the unit of take-up. The audio service also tracks incoming phone calls, whether the automated response call is answered, and how much content is played to the recipient before the recipient hangs up. In this paper, any instance of the system's automated response call being received by the ASHA is counted as one session.

I collect survey data about ASHA traits, beliefs, and preferences through a baseline survey conducted the launch of the experiment and a midline survey conducted in month 9 of the experiment. The baseline survey includes modules for personal and household demographics; an assessment of health- and pregnancy-related knowledge; a series of psychometric scales; and job motivations and preferences. The psychometric scales are adapted from previously published validated scales designed to measure extrinsic, intrinsic, and prosocial preferences Amabile et al. (1994); Grant (2008);

Wrzesniewski et al. (1997). The statements constituting each scale are listed in Section C.2. During the survey, the enumerator read each scale item aloud to the respondent, and the respondent was asked to state whether she strongly agreed, agreed, neither agreed nor disagreed, disagreed, or strongly disagreed with each statement.

The midline survey was administered during the ninth month of the experiment. It queried each ASHA's work habits, motivations, time use across tasks, and comprehension and ability to use the experimental app to which she was assigned. A final module asked about each ASHA's knowledge of the treatment assignment and performance of each of her peers within her subcenter.

## **C.2 Psychometric scales: list of statements**

The psychometric scales used in this experiment are adapted from validated scales. During the actual baseline survey, the items were interspersed with one another along with items from other scales. The extrinsic and intrinsic motivation scales draw primarily from Amabile et al. (1994), and the prosocial motivation scale draws primarily from Grant (2008). Each of the three scales also draws from items in Wrzesniewski et al. (1997). The particular items were chosen *ex ante* based on consultation with native Hindi speakers regarding which items could be translated into Hindi most clearly and would be most likely to be understood by the ASHA respondents. All three of these scales were pre-specified in a pre-analysis plan Lee (2014).

The final two scales were administered in the same exercise. The social desirability scale is adapted from Hays et al. (1989). The competitive preference scale is adapted from Barrick et al. (2002) and Amabile et al. (1994).

*Extrinsic motivation:*

- I want to be in a higher-level job in five years.
- My primary reason for working is financial—to support my family and lifestyle.
- I think a lot about how much money I have and how much I can make.
- I often think about salary and promotions.

*Intrinsic motivation:*

- I enjoy talking about health to others.
- The more difficult the problem, the more I enjoy trying to solve it.
- I enjoy doing work that is new to me.
- I want to find out how good I can be at my work.
- What matters most to me is enjoying what I do.

*Prosocial motivation:*

- ASHA work makes the world a better place.
- I want to help others through my work.
- At work, I care about improving the lives of other people.
- I do more work than is required of me to help my clients to be healthy.

*Social desirability motivation:*

- I am always respectful and considerate even to people who are rude and unfriendly.
- No matter who I am talking to, I am always a good listener.
- I sometimes feel annoyed when I don't get my way (reverse-coded).

*Competitive motivation:*

- I often compare my ASHA work against other ASHAs' work.
- I try to be the highest performing ASHA.
- To me, success means doing better than other ASHAs.