Ordeal Mechanisms and Information in the Promotion of Health Goods in Developing Countries: Evidence From Rural China

Xiaochen Ma[1], Sean Sylvia[2], Matthew Boswell, and Scott Rozelle[3]

Draft Oct 20th, 2013

Abstract: The cost-effectiveness of policies that provide subsidized health goods is often compromised by the fact that many individuals do not use the goods that are provided to them. Cost-sharing strategies can improve targeting efficiency by inducing self-selection, but have been shown to significantly dampen overall take-up (which is often the primary policy goal). As a potential solution, applying an *ordeal* mechanism to the distribution of subsidized health goods has been proposed as a way to balance the dual goals of improving targeting efficiency while at the same time maintaining take-up. Another commonly used policy to promote health goods is health training. Despite the work that has been done (separately) on ordeal mechanisms and health training programs, there is a gap in the literature: that is, the effect of ordeal mechanisms without and without health training programs. In this paper, we report the results of a randomized field experiment designed to provide eyeglasses to myopic students of primary schools in rural western China. We test the performance of an ordeal mechanism with and without health training. We emphasize three findings. First, both with and without health training, the ordeal modestly improved targeting efficiency and reduced program costs compared to direct free distribution. Second, in the short run, the training program caused the ordeal mechanism to screen out some individuals who would have used eyeglasses if provided for free. Third, in the medium run, the ordeal mechanism only screened out individuals who would not have used eyeglasses both with the training program and without.

Keywords: Ordeal Mechanism; Information; Health; Developing Countries

JEL Codes: C93, D12, H42, D80

Affiliations: University of California, Davis: Xiaochen Ma; University of Maryland, College Park: Sean Sylvia; Stanford University: Matthew Boswell and Scott Rozelle. We would like to acknowledge the great effort of 300 enumerators from the Center for Chinese Agricultural Policy, Chinese Academy of Sciences; Renmin University of China; Xibei University; and Shaanxi Normal University. We give special thanks to staff from Zhongshan Ophthalmic Center, Sun Yat-sen University, for their invaluable guidance and advice. We are also grateful for financial and technical support from OneSight, Luxottica-China, Essilor and CLSA. All errors are our own.

1. Introduction

Policy makers seeking to promote health goods through subsidization in developing countries are often faced with the dual challenge of a) increasing the overall take-up rate among individuals who can benefit from a good and b) doing so in a cost effective way. While for some goods, the issue of cost-effectiveness is almost moot since private and social benefits easily outweigh the costs (vaccinations, for example); for other health goods, however, higher costs of provision make cost-effectiveness a more salient consideration. This is particularly so when the provision of an otherwise effective health good does not necessarily lead to benefits, such as when recipients simply do not use the good provided. The literature on health goods in developing countries provides numerous such examples: insecticide-treated mosquito nets (Cohen and Dupas, 2010), water purification technologies (Dupas et al 2013) or improved cookstoves (Miller and Mobarak, 2013). For these technologies, cost-effectiveness can be improved by targeting these subsidized goods to individuals who are most likely to use—and therefore, benefit—from them.

One way to potentially improve the cost-effectiveness of providing a health good is through cost-sharing (Cohen and Dupas 2010, Ashraf et al. 2010). Cost-sharing can improve targeting in two ways, through a.) a selection effect (or the effect by which a higher price skews the distribution of "buyers" towards individuals with a greater propensity to use the good); and b.) a sunk cost effect (the a psychological effect whereby individuals are more likely to use a good because they paid a positive price for it, regardless of the intrinsic welfare the good brings).

The empirical evidence on the positive aspects of cost-sharing (i.e., increased use and, therefore, reduced waste) is not very encouraging (Cohen and Dupas 2010, Ashraf et al. 2010). There is also little evidence that higher prices lead to greater use. Indeed, charging even a modest price considerably dampens the purchase of health goods and screens out a large number of individuals who would otherwise use them (Cohen and Dupas 2010, Ashraf et al. 2010- which, of course, compromises another of the primary goals of policy makers).

A recently-proposed method of balancing these dual goals in developing countries (increasing overall take-up and doing so cost-effectively) is through a so-called *ordeal* mechanisms (Alatas et al., 2013; Dupas et al., 2013). Traditionally, ordeal mechanisms have been used to target government transfers such as welfare and unemployment (Nicols et al., 1971; Nicols and Zeckhauser, 1982; Besley and Coate, 1992; Alatas et al., 2013). These mechanisms attempt to cheaply target resources to the needy by imposing different costs on targeted and non-targeted individuals. This approach may then lead individuals to self-select, with targeted individuals selecting themselves into the program (and non-targeted individuals not selecting themselves in). This self-selection thereby eliminates the need for costly (or impossible) collection of information for targeting. Because the utility cost imposed by the ordeal is different for rich and poor (poor people may be more liquidity constrained and less time constrained than relatively better-off people), requiring that applicants undergo such an ordeal (which may be as simple as requiring applicants to go through time-consuming application procedures) can deter the relatively wealthy from applying for benefits thus reserving government resources for the needy. Putting a mechanism in place that leads individuals to self-select may not only be cheaper than alternative targeting approaches but might also lead to more efficient targeting (Alatas et al., 2013).

The potential cost of ordeal mechanisms, however, is that they can screen out individuals who would have otherwise demanded the good. Of course, this negative effect is similar to what has been observed for price instruments (Cohen and Dupas 2010, Ashraf et al. 2010). If the ordeal mechanism reduces access to those that need the good, then it could be counter productive to the other primary goal of the program—increasing overall take-up of a health good among those who could benefit from it.

How well ordeal mechanisms function to screen out only individuals who would not use a good if acquired for free depends on a number of factors. Critical among these is individuals' perceptions of the relative benefits and costs of using a particular health good. Because take-up and use occur sequentially, these perceptions are not only important at the time individuals choose to acquire a good but also how these perceptions evolve with experience. Particularly in the case of experience goods (such as eyeglasses or bednets), perceptions may differ at these two points in time because: a.) the two decisions are made under two different information sets; b.) individuals gradually update their private perceptions of benefit and cost. If individuals systematically undervalue a health good relative to their realized net benefit ex-post (after having acquired a good) the cost of an ordeal may screen out too many people. Indeed, incomplete or imperfect information that leads individuals to undervalue a good is often cited as a major factor contributing to surprisingly low take-up and use rates of health technologies in developing countries, even when they are provided for free (Dupas, 2010).

Given evidence that individuals tend to underestimate the value of many health goods, a common approach to increase overall take-up (independent of targeting and cost-effectiveness) is through health training or “social marketing” campaigns. One part of the sizeable literature on health education and training shows that it can effectively increase take-up and even increase the effect that subsidies have on the take-up of health goods (Jalan and Somanathan 2008, Luoto et al. 2012, Ashraf et al., 2013). On the other hand, some research has shown health education and training campaigns to be largely ineffective (Shi et al 2012).

Despite the work that has been done on ordeal mechanisms and health training programs separately, as far as we know, there is no research on how health training programs affect the functioning of an ordeal mechanism. Specifically, does training help to keep the overall rate of take-up from being dissuaded by the ordeal imposed? Does training help to screen out those who would tend to not use the good anyhow and induce those who would use the health good to be more decisive in doing so? Given that ordeal mechanisms are often (or easily) implemented together with health training programs, understanding how they interact with one another to influence the take-up and use of health goods can inform how to design policy approaches that can more effectively address the dual goals of increasing overall take-up and doing so in a cost-effective way.

In this paper, we report the results of a randomized field experiment designed to test the performance of an ordeal mechanism with and without an additional education campaign. We emphasize three findings. First, both with and without health training, the ordeal only modestly improved targeting efficiency and reduced program costs compared to direct free distribution. Second, in the short run, despite the overall rise in targeting efficiency, the training program caused the ordeal mechanism to screen out some individuals who would have used eyeglasses if provided for free. This appears to be due to the health training program having had a much larger effect on use when the good was directly distributed for free compared to when individuals needed to go through the ordeal to acquire the good. Third, targeting due to the ordeal mechanism was more efficient in the medium run. Both with and without the training program, the ordeal only screened out individuals who would not have used the good if freely delivered.

Our study makes several contributions to the existing literature. First, we contribute to the nascent literature on the application of ordeal mechanisms to the take-up of health goods in developing countries (the only other similar study that we are aware of is Dupas et al, 2013). Second, we present the first evidence of the role of information in determining the "success" of ordeal mechanisms. Third, we contribute to the literature on the role of education campaigns in increasing the take-up of under-utilized health technologies, specifically how information affects the demand curve defined on non-monetary cost (distance).

The rest of the paper is organized as follows. Section 2 describes the research context, presents the experimental design and the data collection. Section 3 presents the results. Section 4 concludes.

2. Context and Experimental Design

2.1 Context

A series of World Health Organization-supported studies suggest that approximately 10 to 15 percent of school-aged children in the developing world are have common vision problems (Maul et al., 2000; Murthy et al., 2002; He et al., 2007). In most cases children's vision problems can be easily corrected by timely and proper fitting of quality eyeglasses (World Health

Organization [WHO], 2006). Unfortunately, studies in a variety of developing countries document that 35 to 85 percent of individuals with refractive error do not have eyeglasses (Bourne et al., 2004, Ramke et al., 2007).

In rural areas of China the prevalence of vision problems among children is among the highest in the world (He, Huang, Zheng, Huang, & Ellwein, 2007; He et al., 2004). However, recent investigations in rural China demonstrate that fewer than one third of children needing glasses have and wear glasses (Li et al., 2010).

Although the inability of many families to afford eyeglasses is likely a part of the reason, two significant factors contributing to such low use rates in China are lack of awareness and mis-information (Li et al., 2008). In many cases, people are not aware of their vision problems since there are very access of vision check (Li et al., 2010; Yi et al., 2013). There are also commonly held (but mistaken) views in many countries, including China, that wearing eyeglasses will harm one's vision, i.e., it cause one's vision to deteriorate even faster (Li et al., 2010; Yi et al., 2013). In our survey, for example we find around half of the students were not aware their vision problems; in addition, when students being asked "do you believe wearing glasses will harm your vision," more than 40% of the students answered "Yes." (See Appendix Table 2 for details)

2.2 Sampling

Our experiment took place in two adjoining provinces of western China: Shaanxi and Gansu.[4] In each of the provinces, one prefecture was included in the study. A map of these regions is provided in Figure 1. From each prefecture, a list of all rural primary schools was obtained. This list formed the sampling frame and 252 schools were randomly selected for inclusion in the study. To minimize the possibility of contamination, we first selected townships

[4] Shaanxi's GDP per capita of USD6108 was ranked 14th among China's administrative regions in 2012, and was very similar to that for the country as a whole (USD 6091) in the same year, while Gansu was the second-poorest province in the country (per capita GDP USD3100) (China National Statistics Bureau, 2012).

and selected only one school per township. Within schools, our data collection efforts (discussed below) focused on 4th and 5th grade students. From each grade, one class was randomly selected and surveys and eye exams were given to all students in these classes.

2.3 Experimental Design

Following the baseline survey and vision tests, schools were randomly assigned to one of the six cells in a 3 by 2 experimental design (shown in Figure 2). Schools were first randomized into one of three *provision* groups (free distribution; ordeal and control). Half of the schools assigned to each provision group were then assigned to receive a health training program. To improve power, randomization was stratified by county and by the number of children in the school found to need eyeglasses. In total, this yielded a total of 45 strata. Our analysis takes this randomization procedure into account (Bruhn and McKenzie 2009).

The three provision groups are as follows:

Free distribution: In this group each student diagnosed with poor vision was given a free pair of eyeglasses as well as a letter to their parents informing their prescription. This pair of glasses was delivered to the hands of students by a team of one optometrist and two enumerators.

Ordeal: In this group, each student diagnosed with poor vision was given a voucher. Their prescription was also printed in the voucher. This voucher was redeemable for one pair of free glasses at an optical store that was in the county seat. The distance from each students home and the county seat varied in great deal among our sample, ranging from 1 kilometer to 105 kilometers (average distance of 1st distance quartile is 9 kilometers and that of 4th distance quartile is 62 kilometers, most 7 times further). In most cases in order to undertake the ordeal, the parent and student would have to pay for a round trip bus ticket and take the time to pick up the glasses. To eliminate the opportunity for arbitrage, student information (student name, school name and county name) was printed on each voucher and students were required to show their ID to redeem the voucher.

Control: Students in the control group were given a letter to their parents informing their myopia status and prescription, but without a free delivered eyeglasses or a voucher.

In each of these groups, half of the schools were assigned to receive an training program:

Training program*:* The training program involved a short documentary-type film, a set of cartoon-based pamphlets for students (regardless of whether they had poor vision or not), and a lecture/handout for the parents and teachers. All these materials address the importance of wearing glasses (correct vision problem and help school performance) and correct some of most commonly wrong briefs (wearing glasses will harm one's vision; doing eye exercise can help treating myopia).

2.4 Data Collection

Survey Data

A baseline survey was conducted in September 2012. The baseline survey collected detailed information on schools, students and households. The school survey collected information on school infrastructure and characteristics of school administrators and teachers. A student survey was given to all students in selected 4^{th} and 5^{th} grade classes. This student survey collected information on basic background characteristics of students as well as information relevant to vision health and eye glasses (whether or not have eye glasses, knowledge about myopia, etc). Household surveys were also given to these same students, which they took home and filled out with their parents. The household survey collected information on households (e.g., parents education levels) that children would likely have difficulty answering.

Nearly identical surveys were conducted at endline in May 2013.

Eye Exams

At the same time as the school survey, a two-step eye exam was administered to all students in the randomly selected classes in all sample schools. First, visual acuity screenings (or E-chart exams) were administered by a team of two trained staff. Visual acuity was tested separately for each eye at a distance of 4 meters using Early Treatment Diabetic Retinopathy

Study (ETDRS) charts (Precision Vision, La Salle, IL, USA) in a well-lighted, indoor area of the school. Students who failed the visual acuity screening test (cutoff is defined by VA of either eye less than or equal to 6/12 or 20/40) were enrolled in a second vision refraction test which was carried out at each school 1-2 days after the first. This second vision test was conducted by a team of one optometrist, one nurse and one assistant staff. Cycloplegic automated refraction with subjective refinement was underwent to determine prescriptions for children needing glasses.

Unannounced Checks on Short and Medium Term Take-up and Use of Glasses

Two rounds of unannounced checks were conducted to collect information on the take-up and use of eyeglasses. The short term check was done in early November 2012 (approximately 3 weeks after glasses and vouchers were distributed). In this round, a team of two enumerators visited each of the 252 schools without any prior announcement of their visit. Once at the school, enumerators entered classes while classes were in session and recorded the number of students wearing glasses. One limitation of these data is that it was not recorded for individual students. Instead, enumerators only collected information on the number of students in each class wearing glasses. After this count was finished, students with poor eyesight were given a short survey that included questions about a) whether they owned eyeglasses and how they were acquired and b) if they owned eyeglasses how often and when they wore them.

A second round—medium term check—was conducted in line with the endline survey in May 2013 (seven months after eyeglasses and vouchers were distributed). As the endline survey was more involved and time consuming, school principals in each school were informed the arrival of the survey team. However, they were not told specifically that we would check to see if children were wearing eyeglasses. In line with the first round, a team of 2 enumerators was sent into schools in advance of the rest of the survey team to conduct classroom checks. In this second round of checks, enumerators were given a list of children diagnoses with poor vision to record individual-level information on the use of glasses for each student.

2.5 Baseline Characteristics and Balance Check

Baseline characteristics of sample students are shown in Table 1. Of the 19,934 students given eye examinations at baseline, 3,177 (16%) were found to be myopic and need glasses. Only these students are included in the analysis sample.

On average, only 15 percent (463 out of 3177) of students who needed glasses in had glasses at baseline. In line with the finding in literature, besides the liquidity constrain, misinformation plays an important role of this low rate. For example, nearly 40 percent of myopic children believed that wearing eyeglasses could deteriorate one's vision and less than half knew they were myopic. Few of these students–only about one third–had family members that wore glasses. Generally, the education level of parents was low: only 11% of mothers and only 16% of fathers had attended high school. Ten percent of students in the sample had parents who had both migrated elsewhere for work and 22% boarded at school.

The last two columns of Table 1 show that these characteristics are well-balanced across the treatment groups. The first part of the table shows tests across the provision groups (free distribution, ordeal and control) and the second part of the table shows tests between students in training program and no training program schools. These tests were conducted by regressing baseline characteristics on treatment arm dummy variables, controlling for randomization strata fixed effects and clustering errors at the school level. According to this table, the sample is well-balanced in both dimensions: only 2 out of 13 coefficients were significant at 10% or higher.

2.6 Attrition

The overall attrition rate after seven months was low (less than four percent of our sample students dropped between baseline and endline). Further, there is no differential attrition across treatment groups. Results of this check are in Appendix Table A.1.

3. Results

3.1 Take-up and Voucher Redemption

We distinguish take-up and use. Take-up is defined by ownership, i.e., if students own one pair of eyeglasses (regardless of how it was obtained). Use is defined by whether students actually wear glasses. We begin the discussion of results by examining the effect of the ordeal (vouchers) on the take-up of eyeglasses. Because 100% of students in "free" group obtained glasses by design, we exclude this group from this part of the analysis.

In this section, we are interested in three main questions. First, what percent of voucher recipients redeemed vouchers? Another way of asking the question in this part of the analysis is that we seek to understand what percentage of individuals was screened out by the ordeal. Second, how did the training program affect voucher redemption? In this part of the analysis we seek to understand if information affects that impact of the ordeal. Third, how does take-up in the case of ordeal group compare to the case of no subsidy of any kind (the control group)?

3.1.1 Voucher Redemption Rates

Figure 4 plots short and medium terms take-up rates for each experimental cell. Panel A and B shows the data collected in short term and medium term, respectively. Beginning with the case of no training program (columns on the left side of each graph), we find that the rates of voucher redemption are high. In the short term (after only 3 weeks) 84% of students had glasses in the ordeal group. Surprisingly, the medium term data shows an take-up rate only slightly higher – suggesting that if individuals were going to redeem the voucher, they did so immediately. In the medium term, the take-up rate in the ordeal group with no training program was 89%. Therefore, in the case of our rural China voucher treatment arm, the ordeal only screened out 11% of those who would have acquired glasses if given for free.

With the training program, take-up rates in the ordeal group were not much higher. The short run rate in the ordeal group with training was 89%. The medium run rate was 92%. The training program therefore led to raw differences of only 5 percentage points in the short term and 3 percentage points in the medium term.

3.1.2 Determinants of Voucher Redemption

Table 2 looks more closely at the effect of training and distance on voucher redemption.

This table shows estimates from the following linear probability model (using only observations from the ordeal group):

$$\text{Voucehr Redemption}_i = \alpha + \beta_1 * \text{Training}_i + \beta_2 * \text{Distance}_i + \beta_3 * \text{Distance}_i * \text{Training}_i + \text{Controls}_i + \text{Strata Fixed Effects}_i + \varepsilon_i \quad (1)$$

where $\text{Voucehr Redemption}_i$ is 1 if the student had redeemed the voucher. This is regressed on a dummy for being assigned to the training program group (Training_i), distance from the school to the county seat (Distance_i, where vouchers were redeemed), and the interaction between the training group dummy and distance) controlling for baseline characteristics of student (awareness of their myopia status, misinformation about wearing glasses) and both individual SES variables (gender, grade level, school boarding status) as well as family SES variables (parental migrant status, parental education and household wealth) as well as randomization strata fixed effects. Standard errors are clustered at the school level.

In the table, the first three columns show variants of this regression with voucher pickup by the short term check as the dependent variable. Columns (4) to (6) show estimates for pickup by the endline survey. In addition to estimating the effect of distance assuming linearity in columns (1) and (4), columns (2), (3) , (5) and (6) include dummies for the 2nd, 3rd and 4th quartiles of the distance variable.

In line with the take-up results plotted in Figure 4 above, we see that the average effect of the training program (irrespective of distance) has only a modest effect on voucher redemption (Table 2, columns 1 and 4, row 1). This effect is significant at the short term but fades away by the enline survey. Moreover, as would be expected, distance has a strong negative effect on voucher redemption. This correlation with distance also appears to be non-linear (columns 2 and 5): redemption rates decline significantly only when schools are in the 3rd or 4th quantiles of the distance measure. Interestingly, columns (3) and (6) show that training has a strong influence on those living the furthest away. For these individuals, the training program compensates almost fully for the reduction in demand due to the cost of the ordeal (distance). This effect of flattening the demand curve (defined over distance) is consistent with Ashraf et al. (2013) who find training flattens the demand curve defined over prices.

3.1.3 Voucher Redemption vs. Self-purchase

In this section, we compare take-up in the ordeal group to take-up in the control group. This comparison provides an indication of how many more individuals received glasses with the voucher than would have otherwise (left to purchase glasses on their own). To do so, we run the following regression using observations from the ordeal and control groups:

$$\mathrm{Take-up_i} = \alpha + \beta_1 * \mathrm{Ordeal_i} + \beta_2 * \mathrm{Training_i} + \beta_3 * \mathrm{Ordeal_i} * \mathrm{Training_i} + \mathrm{Controls_i} + \mathrm{Strata\ Fixed\ Effects_i} + \varepsilon_i \quad (2)$$

where $\mathrm{Take-up_i}$ is 1 if an individual owns glasses and 0 otherwise; $\mathrm{Ordeal_i}$ is an indicator for assignment to the ordeal group; $\mathrm{Training_i}$ is the an indicator variable of whether the student is in the group of receiving the training program; $\mathrm{Controls_i}$ are the same as in equation (1).

Columns (1) and (2) in Table 3 report the estimates based on short term data. Among students given no training program, the voucher intervention is estimated to boost the take-up of glasses by 58 percentage points from the base of 24 percent in pure controlled group (almost a 250% increase). Absent the voucher, training program has a small effect on take-up similar to the effect seen on voucher redemption. In the short term, training has a positive influence on the increase in take-up due to the voucher, though this is not significant.

The data collected in the medium term shows some interesting differences with the short run (column (3) to (4) in Table 2). First, the magnitude of impact of the voucher alone on improving glasses take-up decreases by 12 percentage points (or 21%) (the $\widetilde{\beta_1}$ decreases from 0.58 to 0.46). This is because students in control group have increased their glasses ownership by 19 percentage points (an 80% increase) instead of that in ordeal group drops over time. The impact of training alone also slightly increased about 3 percentage points (or 60%) after one semester (the $\widetilde{\beta_2}$ increases from 0.05 to 0.08).

3.2 Use

In this section, we compare the use rates across all treatment groups, in which use is defined by whether students actually wear glasses. Comparing short- and medium-run use rates, particularly between the free and ordeal groups provides an indication of how well the ordeal mechanism functioned. If use rates are higher in the free group than in the ordeal group, this suggests that the ordeal screened out individuals who would have used glasses if given for free. If they are similar, however, this suggests that the ordeal mechanism screened out only those individuals who would not have used eyeglasses if given for free.

Figure 5 shows the use rates based on the unannounced checks for each of the experimental groups in the short term (Panel A) and in the medium term (Panel B). In the short term data, we see that the functioning of the mechanism varies with and without the training program. With no training, use in the ordeal group is actually slightly *higher* than the free group. One possible explanation for this is that the cost of the ordeal (travelling to redeem the voucher) induces a sunk cost effect (Thaler 1980; Arkes and Blumer 1985). With the training program, however, we see that this result is reversed. With training, use in the free group is higher than in the ordeal group. This implies that, with training, the ordeal screens out individuals who would have used glasses if given for free. Comparing this set of outcomes with the outcomes from the no training group, the reverse in results appears to be due to the fact that training had larger impact in free group than in the ordeal group. In Panel (B), the pattern found in the medium-term results is similar to the short term but overall use rates are smaller.

To compare use rates more formally, we estimate the following model:

$$\text{Use}_i = \alpha + \beta_1 * \text{Training}_i + \beta_2 * \text{Ordeal}_i + \beta_3 * \text{Free}_i + \beta_4 * \text{Training}_i * \text{Ordeal}_i + \beta_5 * \text{Training}_i * \text{Free}_i + \text{Controls}_i + \text{Strata Fixed Effects}_i + \varepsilon_i \qquad (3)$$

where the right hand side is the same as above. Use_i , however, differs between the short and medium run results due to the protocol change in data collection. For the short term estimates, because the data is collected in school level, Use_i is the proportion of myopic children in a school wearing glasses at the time of the short term check. For medium-run estimates, since data is counted in individual level, Use_i is 1 if student I was wearing glasses at the time of the medium-run check and 0 otherwise. $Controls_i$ are the same as in equation (1).

Columns (1) and (2) of Table 4 show the short run use estimates. First, comparing the coefficients on Ordeal and Free shows that with no training, use in the ordeal group is indeed higher than use in the free group but this difference is not significant. We estimate this difference to be 8.0 percentage points (difference between β_2 and β_3 in column 2, p-value: 0.18) when covariates are included. With the training program, this relationship is reversed. We estimate that the training program boosted use in the free group by 16.6 percentage points (the sum of β_1 and β_5, p-value: 0.00) while it had no effect in the ordeal group. As a result, in the presence of the training, use in the ordeal group is 9.4 percentage points (p-value: 0.09) lower than the free group. Thus, in the short run, the ordeal does not screen out individuals who would have used eyeglasses when there is no training program but does when jointly implemented with an training program.

Turning to the medium-run results, we find that without the training program use in the voucher and free groups remains comparable. In the medium run, however, the impact of the training program on use in the free group is gone. As a result, the difference between the ordeal and free groups with the training program becomes insignificant (5.8 percentage points; p-value 0.21).

Overall, the results suggest that in the short run, the training program caused the ordeal mechanism to screen out some individuals who would have used eyeglasses if provided for free. In the medium run, however, the ordeal mechanism only screened out individuals who would not have used eyeglasses both with the training program and without.

4. Conclusion

Ordeal or self-selection mechanisms have been proposed as a way to improve the cost-effectiveness of programs that provide subsidized health goods in developing countries (Dupas et al 2013). By sorting out individuals who would have not used a good even if provided for free, ordeal mechanisms can save considerable program resources. There is potential, however, that such mechanisms can sort out too many individuals, as is found in the empirical evidence of cost-sharing strategies, a result that can be counter productive to increasing overall take-up.

In this paper, we examined how a simultaneously-implemented health training program--in addition to subsidization, another commonly-used policy tool to boost take-up rates of health goods in developing countries--affected the performance of an ordeal mechanism. To do so, we implemented a cluster randomized experiment in which schools were first randomized into one of three groups: a.) a control group, b.) an ordeal group in which children with poor eyesight were given vouchers redeemable for free eyeglasses in the county seat, or c.) a group in which all children were delivered free eyeglasses. Half of the schools in each group were then randomly assigned into a cross-cutting health training program that provided information on the benefits of wearing glasses and addressed common misinformation related to wearing glasses.

We find that—in the short run—the ordeal mechanism worked well when not combined with a training program, that is, it only screened out individuals who would not have used the eyeglasses if given them for free. When paired with the training program, however, the ordeal screened out individuals in the short term who would have used glasses if provided them for free. In the medium run, the ordeal performed well (only screening out those who would not have used glasses) both with and without the training program.

While it is difficult to say what lead to these results, we do observe that the training program had a much larger effect on use when eyeglasses were provided for free (at least in the short run). This suggest that results may have been driven by differences in how individuals respond to product-specific health training program when that product was provided to them freely or if they incurred a cost to acquire it.

Reference

Arkes, H, C., Blumer, 1985. “The Psychology of Sunk Cost,” *Organizational Behavior and Human Decision Processes*, 35 (1985), 124–140.

Ashraf, N , K., Jack, E., Kamenica. 2013. Information and subsidies: Complements or substitutes? *Journal of Economic Behavior & Organization* 88 (2013) 133–139

Bourne, R. R., Dineen, B. P., Huq, D. M. N., Ali, S. M., & Johnson, G. J. 2004. Correction of refractive error in the adult population of Bangladesh: meeting the unmet need. *Investigative Ophthalmology & Visual Science, 45*(2), 410-41

Cohen, J., P., Dupas. 2010 Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment *The Quarterly Journal of Economics* (2010) 125 (1): 1-45

Dupas, P. 2013. Short-Run Subsidies and Long-Run Adoption of New Health Products: Evidence from a Field Experiment. *Econometrica, Forthcoming*

Dupas, P. 2011. "Do Teenagers Respond to HIV Risk Information? Evidence from a Field Experiment in Kenya," *American Economic Journal: Applied Economics,* American Economic Association, vol. 3(1), pages 1-34, January.

Dupas, P., V., Hoffmann, M., Kremer, A., Zwane. 2013. Ordeal mechanisms and preventive health products: Evidence on short-term take-up and habit formation. *Working paper*

He, M., Huang, W., Zheng, Y., Huang, L., & Ellwein, L. B. (2007). Refractive error and visual impairment in school children in rural southern China. *Ophthalmology*, *114*(2), 374-382.

He, M., Zeng, J., Liu, Y., Xu, J., Pokharel, G. P., & Ellwein, L. B. (2004). Refractive error and visual impairment in urban children in southern China. *Investigative Ophthalmology & Visual Science*, *45*(3), 793-799.

Jalan, J. and E. Somanathan, 2008. “The Importance of Being In-formed: Experimental Evidence on Demand for Environmental Quality.” *Journal of Development Economics,* 87(1):14–2

Kremer, M.,, E., Miguel, S., Mullainathan, C., Null, A. P. Zwane. 2011. “Social Engineering: Evidence from a Suite of Take-up Experiments in Kenya”, *Working Paper*

Li, L., Lam, J., Lu, Y., Ye, Y., Lam, D. S., Gao, Y., .Congdon, N. (2010). Attitudes of students, parents, and teachers toward glasses use in rural China. *Archives of ophthalmology, 128*(6), 759.

Luo, R, Y., Shi, L., Zhang, C., Liu, S., Rozelle, B., Sharbono. 2009. Malnutrition in China's Rural Boarding Schools: the Case of Primary Schools in Shaanxi Province. *Asia Pacific Journal of Education*, Vol. 29 no. 4, page(s) 481–501

Shi, Y, F.,Chang, X., Su, R., Luo, L., Zhang, S., Rozelle. 2012 Parental training, anemia and the impact on the nutrition of female students in China's poor rural elementary schools. *China Agricultural Economic Review*, 05/2012; 4(2):151-167.

Luoto, J., D. Levine and J. Albert, 2012. Information and Persuasion: Achieving Safe Water Behaviors in Kenya." *Working Paper*
Maul, E., Barroso, S., Munoz, S. R., Sperduto, R. D., & Ellwein, L. B. (2000). Refractive error study in children: results from La Florida, Chile. *American Journal of Ophthalmology*, *129*(4), 445-454.

Miguel, E., and M. Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities", *Econometrica* 72(1): 159-217

Miller, G., and A. M. Mobarak. 2013. Gender Differences in Preferences, Intra-Household Externalities, and Low Demand for Improved Cookstoves. *Working paper*

Murthy, G., Gupta, S. K., Ellwein, L. B., Muñoz, S. R., Pokharel, G. P., Sanga, L., & Bachani, D. (2002). Refractive error in children in an urban population in New Delhi. *Investigative Ophthalmology & Visual Science*, *43*(3), 623-631.

Nichols, A. L., R. Zeckhauser. 1982. "Targeting Transfers through Restrictions on Recipients", *The American Economic Review Papers and Proceedings*, Vol. 72 (2): 372-377

Nichols, D., E. Smolensky and T.N. Tideman. 1971. "Discrimination by waiting time in merit goods." *The American Economic Review* 61(3):312–323.

Ramke, J., Du Toit, R., Palagyi, A., Brian, G., & Naduvilath, T. (2007). Correction of refractive error and presbyopia in Timor-Leste. *British Journal of Ophthalmology*, 91(7), 860-866.

Thaler, R, 1980. "Toward a Positive Theory of Consumer Choice," *Journal of Economic Behavior and Organization*, 1 (1980), 39–60.

Yi, H.., L., Zhang, X., Ma, N., Congdon, Y., Shi, X., Pang, J., Zeng, L., Wang, M., Boswell, S., Rozelle, 2013. Poor Vision among China's Rural Primary School Students: Prevalence, Correlates and Consequences. *working paper*.

Figure 1. Study Region

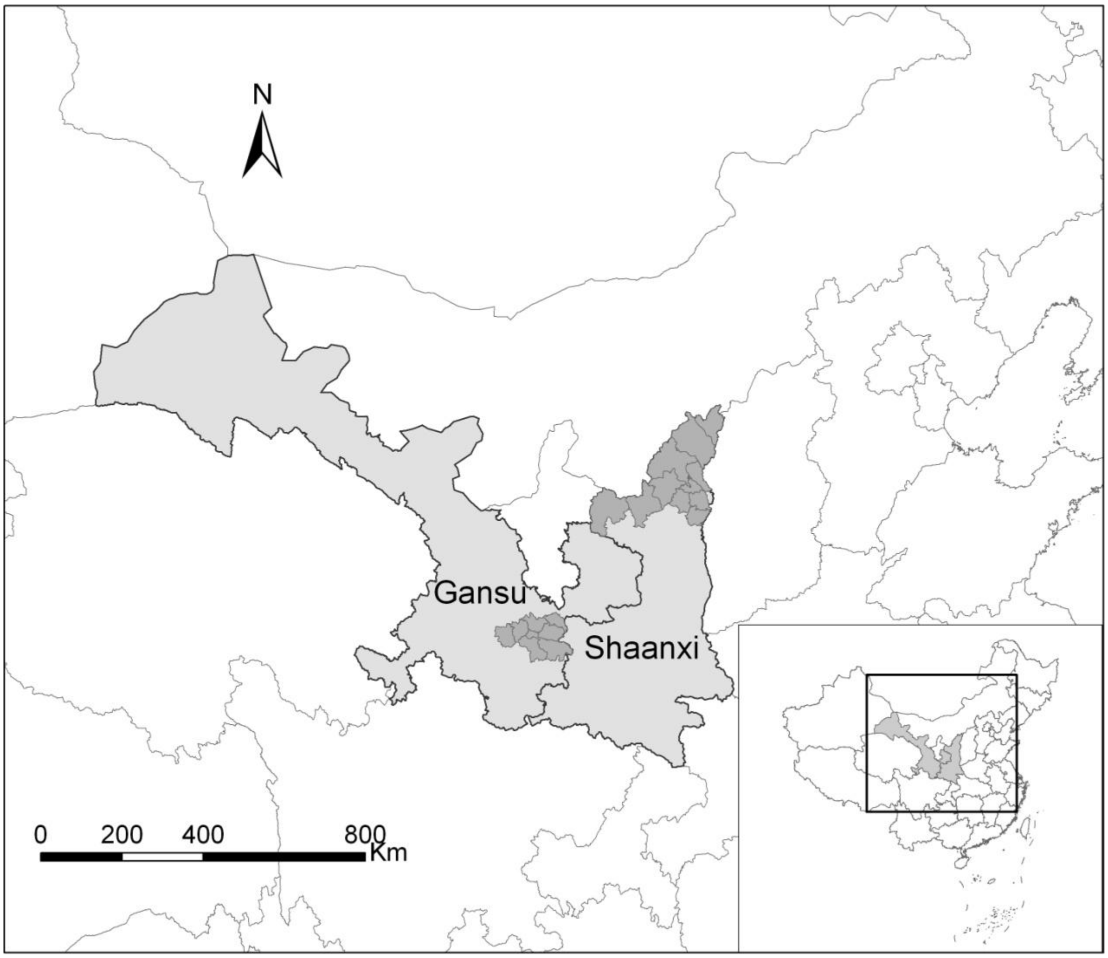


Figure 2: Experimental Design

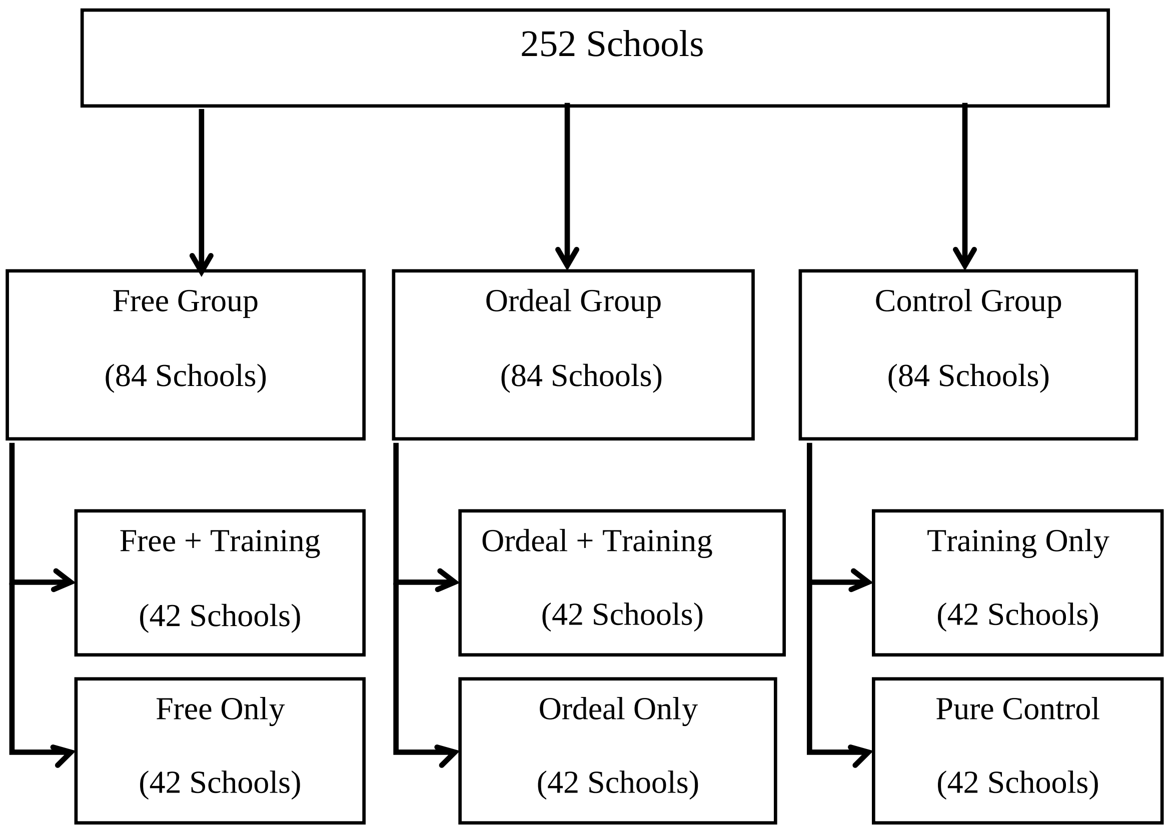


Figure 3: Data Collection and Intervention Timeline

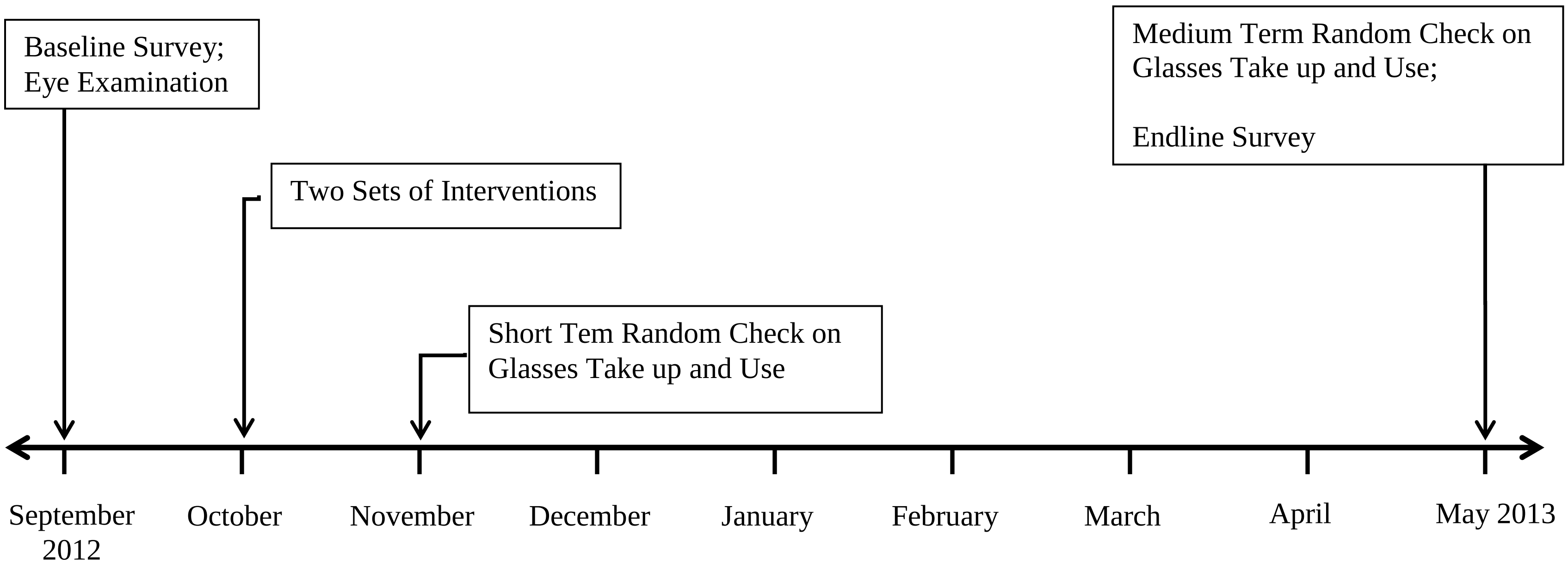
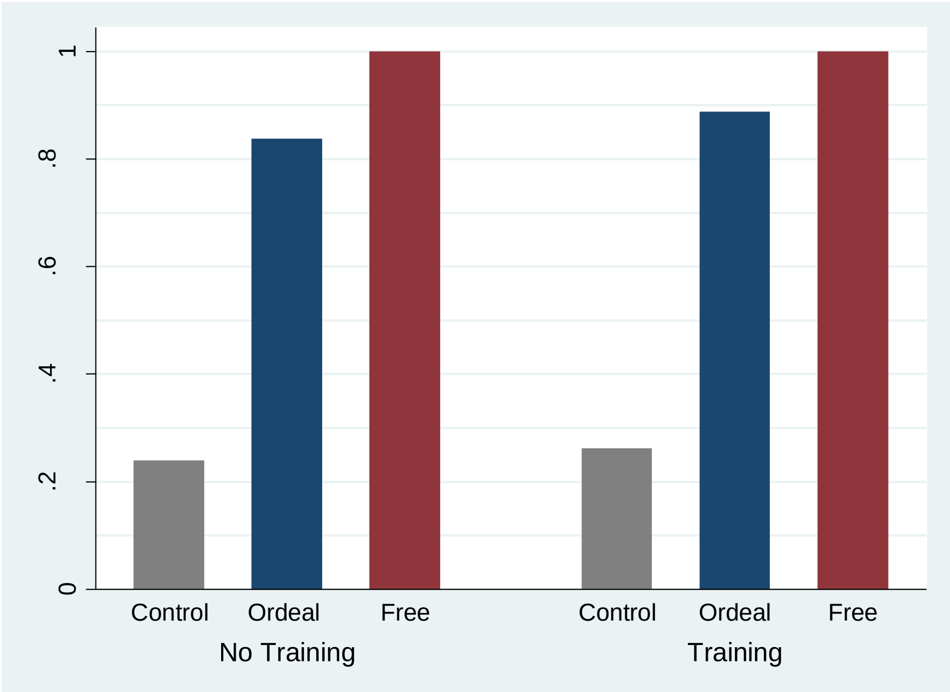
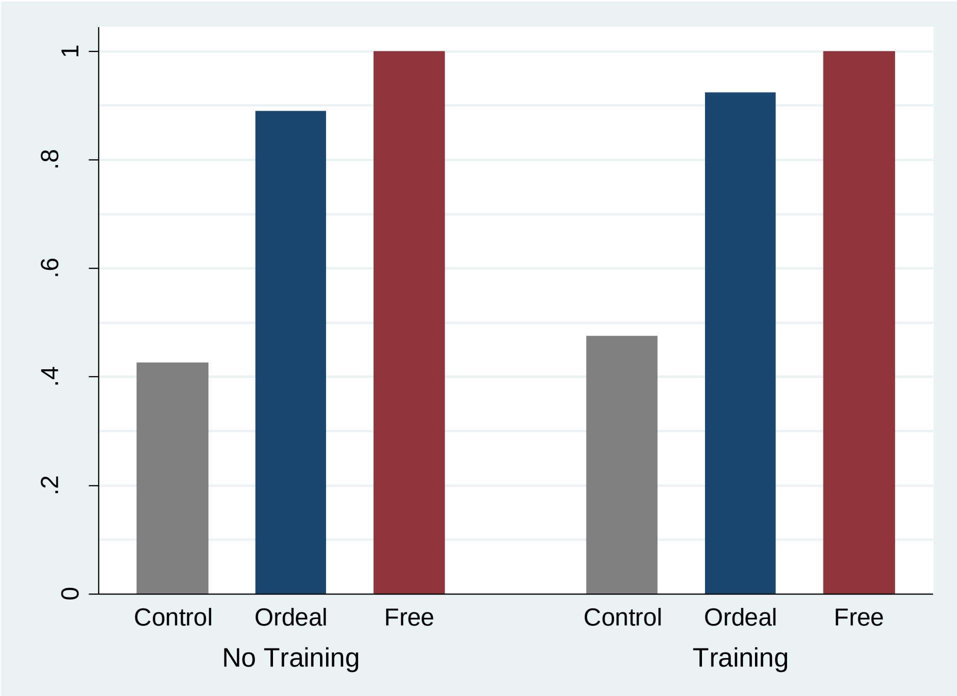


Figure 4: Glasses Take-up Across Treatment Groups Over Time

Panel A. Short Term (Three Weeks Follow up)



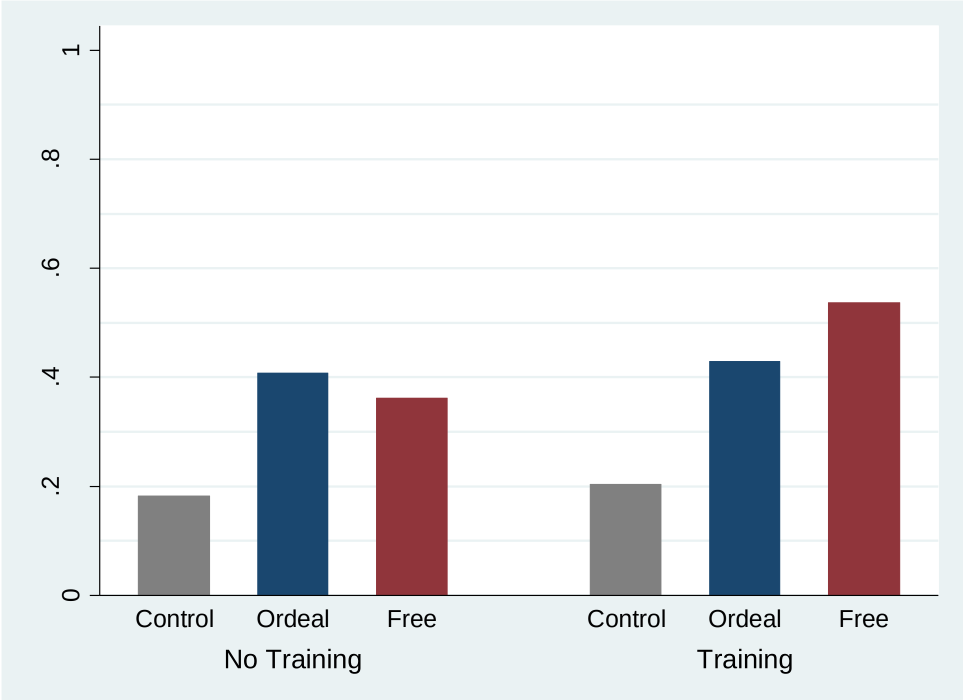
Panel B. Medium Term (Seven Months Follow Up)



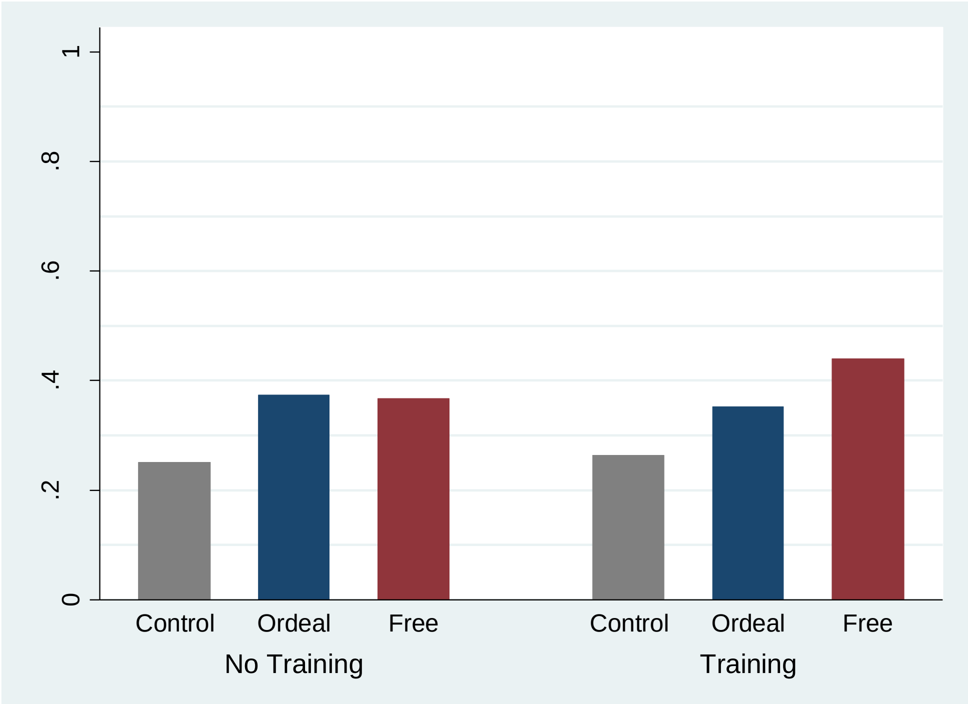
NOTES: Figures show mean values of glasses take-up. Estimated treatment effects on glasses take-up are reported in Table 2.

Figure 5: Glasses Use Across Treatment Groups Over Time

Panel A. Short Term (Three Weeks Follow up)



Panel B. Medium Term (Seven Months Follow Up)



NOTES: Figures show mean values of glasses use. Estimated treatment effects on glasses use are reported in Table 4.

Table 1. Baseline Characteristics Among Treatment Arms

	Total (1)	Control (2)	Ordeal (3)	Free (4)	P value (2) = (3)	P value (2) = (4)
Baseline own glasses (0/1)	0.15	0.14	0.14	0.16	0.98	0.42
	(0.35)	(0.35)	(0.35)	(0.36)		
Baseline severity of myopia	0.56	0.57	0.57	0.54	0.91	0.20
	(0.50)	(0.50)	(0.50)	(0.50)		
Believes Wearing Glasses Will	0.40	0.41	0.37	0.42	0.17	0.69
Harm Vision (0/1)	(0.49)	(0.49)	(0.48)	(0.49)		
Awareness of myopia (0/1)	0.46	0.45	0.47	0.46	0.56	0.87
	(0.50)	(0.50)	(0.50)	(0.50)		
Family glass (0/1)	0.34	0.34	0.32	0.36	0.31	0.51
	(0.48)	(0.48)	(0.47)	(0.48)		
Number of classmates wearing	1.64	1.36	1.47	2.04	0.71	0.10*
glasses	(2.04)	(1.74)	(1.70)	(2.45)		
Gender (0/1)	0.49	0.50	0.48	0.49	0.35	0.68
	(0.50)	(0.50)	(0.50)	(0.50)		
Grade5 (0/1)	0.60	0.60	0.62	0.60	0.39	0.80
	(0.49)	(0.49)	(0.49)	(0.49)		
Boarding (0/1)	0.22	0.23	0.19	0.25	0.41	0.66
	(0.42)	(0.42)	(0.39)	(0.43)		
Parent migration (0/1)	0.10	0.11	0.10	0.09	0.68	0.31
	(0.30)	(0.31)	(0.30)	(0.29)		
Father high school (0/1)	0.16	0.16	0.14	0.17	0.38	0.47
	(0.36)	(0.36)	(0.35)	(0.38)		
Mother high school (0/1)	0.11	0.10	0.08	0.14	0.29	0.03**
	(0.31)	(0.30)	(0.28)	(0.35)		
Household asset	0.00	0.04	0.04	0.06	0.99	0.30
	(1.00)	(0.97)	(0.99)	(1.03)		
Observation	3177	1036	988	1153	2024	2189

Notes: All standard deviations in parentheses. Tests for equality between study arms account for clustering at the school level.

Table 1 Cont. Baseline Characteristics Among Treatment Arms

	No Training	Training	P value
	(1)	(2)	(1) = (2)
Baseline Glasses (0/1)	0.16	0.13	0.17
	(0.37)	(0.34)	
Baseline Severity of Myopia	0.56	0.55	0.59
	(0.50)	(0.50)	
Believes Wearing Glasses Will Harm Vision (0/1)	0.39	0.41	0.30
	(0.49)	(0.49)	
Awareness of Myopia (0/1)	0.48	0.45	0.22
	(0.50)	(0.50)	
Family Member Wearing Glass (0/1)	0.34	0.35	0.44
	(0.47)	(0.48)	
Number of Classmates Wearing Glasses	1.64	1.65	0.99
	(1.98)	(2.10)	
Gender (0/1)	0.50	0.48	0.40
	(0.50)	(0.50)	
Grade5 (0/1)	0.61	0.60	0.34
	(0.49)	(0.49)	
Boarding (0/1)	0.21	0.23	0.70
	(0.41)	(0.42)	
Parent Migration (0/1)	0.10	0.10	0.83
	(0.30)	(0.30)	
Father Education	0.16	0.16	0.86
	(0.37)	(0.36)	
Mother Education	0.10	0.12	0.12
	(0.30)	(0.33)	
Household asset	(0.03)	0.03	0.44
	(1.01)	(0.99)	
Observation	1529	1648	3177

Notes: All standard deviations in parentheses. Tests for equality between study arms account for clustering at the school level.

Table 2. Determinants of Voucher Redemption in Ordeal Group

	(1)	(2)	(3)	(4)	(5)	(6)
		Short Term			Medium Term	
Dep. Variable		Voucher Redemption (1 = Yes)			Voucher Redemption (1 = Yes)	
Training	0.051*	0.019	-0.029	0.037	-0.001	-0.038
	(0.028)	(0.034)	(0.044)	(0.024)	(0.031)	(0.036)
Distance	-0.003***			-0.003***		
	(0.001)			(0.001)		
2^{nd} Distance Quartile		0.051	0.018		0.059	0.047
		(0.060)	(0.104)		(0.046)	(0.099)
3^{rd} Distance Quartile		-0.144**	-0.176***		-0.151***	-0.179***
		(0.056)	(0.056)		(0.036)	(0.044)
4^{th} Distance Quartile e		-0.134**	-0.269***		-0.082**	-0.179***
		(0.054)	(0.071)		(0.035)	(0.055)
Training * 2^{nd} Distance Quartile			0.023			0.003
			(0.124)			(0.114)
Training * 3^{rd} Distance Quartile			-0.054			-0.015
			(0.114)			(0.136)
Training * 4^{th} Distance Quartile			0.221***			0.161**
			(0.070)			(0.062)
Controls Included	YES	YES	YES	YES	YES	YES
Observation	947	947	947	947	947	947
*Mean of Dep. Var.*in No Training Ordeal Group		0.84			0.89	
adj. R^2	0.133	0.136	0.141	0.122	0.131	0.133

Notes: Standard deviations clustered at school level are reported in parentheses. Strata fixed effects are included. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3. Impact of Treatments on Take-up (Ordeal and Control Groups)

	(1)	(2)	(3)	(4)
	Short Term		Medium Term	
Dep. Variable	Take-up (1 = Own Glasses)		Take-up (1 = Own Glasses)	
Training	0.024	0.050*	0.055	0.076**
	(0.029)	(0.026)	(0.040)	(0.037)
Ordeal	0.585***	0.581***	0.456***	0.459***
	(0.031)	(0.029)	(0.036)	(0.033)
Ordeal*Training	0.035	0.041	-0.019	-0.016
	(0.043)	(0.043)	(0.052)	(0.051)
Controls Included	NO	YES	NO	YES
Observation	1950	1950	1950	1950
Mean of Dep. Variable in Pure Control Group	0.24		0.43	
adj. R^2	0.407	0.540	0.264	0.358

Notes: Standard deviations clustered at school level are reported in parentheses. Strata was included.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4. Impact of Treatment on Glasses Use

	(1)	(2)	(3)	(4)
	Short Term		Medium Term	
Dep. Variable	Use (1 = Wear Glasses)		Use (1 = Wear Glasses)	
Training	0.012	0.019	0.027	0.036
	(0.037)	(0.038)	(0.036)	(0.036)
Ordeal	0.229***	0.245***	0.123***	0.127***
	(0.053)	(0.053)	(0.044)	(0.040)
Free	0.153***	0.164***	0.109***	0.105***
	(0.050)	(0.052)	(0.038)	(0.037)
Ordeal*Training	-0.009	-0.027	-0.038	-0.038
	(0.072)	(0.070)	(0.062)	(0.060)
Free*Training	0.169**	0.147**	0.049	0.042
	(0.067)	(0.066)	(0.056)	(0.057)
Controls Included	NO	YES	NO	YES
Observation	252	252	3054	3054
Mean of Dep. Variable in Pure Control Group	0.18	0.18	0.25	0.25
adj. R^2	0.280	0.320	0.089	0.148

Notes: Standard deviations clustered at school level are reported in parentheses. Strata fixed effects are included.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A1. Attrition Between Baseline and Endline

	(1)
Dep. Variable	Attrition (1 = Yes)
Training Only Dummy	-0.015
	(0.011)
Ordeal Only Dummy	0.005
	(0.013)
Ordeal + Training Dummy	0.006
	(0.013)
Free Only Dummy	0.006
	(0.014)
Free + Training Dummy	0.011
	(0.013)
Baseline Glasses (0/1)	0.004
	(0.010)
Baseline Severity of Myopia	-0.001
	(0.002)
Believes Wearing Glasses Will Harm Vision	0.001
	(0.009)
Awareness of Myopia (0/1)	-0.011
	(0.007)
Family Member Wearing Glass (0/1)	-0.006
	(0.007)
Number of Classmates Wearing Glasses	-0.001
	(0.002)
Gender (0/1)	-0.008
	(0.007)
Grade5 (0/1)	0.004
	(0.008)
Boarding (0/1)	0.034***
	(0.011)
Parent Migration (0/1)	0.011
	(0.011)
Father Education	0.003
	(0.008)
Mother Education	0.003
	(0.008)
Household asset	0.000
	(0.000)
N	3177
adj. R^2	0.004

Notes: Standard deviations clustered at school level are reported in parentheses. Strata fixed effects are included.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.2 Baseline Awareness and Misinformation

	(1) Understand what is myopia (1= Yes)	(2) Awareness of one's myopia status (1= Yes)	(3) Believe wearing glasses harm one's vision (1= Yes)	(4) Believe eye exercise treats myopia (1 =Yes)
Training Only Group	-0.051	-0.027	0.030	-0.030
	(0.045)	(0.047)	(0.044)	(0.040)
Ordeal Only Group	0.032	0.027	-0.046	-0.047
	(0.051)	(0.041)	(0.048)	(0.047)
Ordeal + Training Group	0.008	-0.017	-0.025	-0.057
	(0.047)	(0.042)	(0.042)	(0.041)
Free Only Group	0.035	0.008	0.004	-0.006
	(0.048)	(0.041)	(0.043)	(0.044)
Free + Training Group	0.018	-0.011	0.038	-0.051
	(0.042)	(0.043)	(0.037)	(0.040)
N	3177	3177	3177	3177
Mean of Dep. Var.		0.465	0.400	0.551
in Pure Control Group	0.429			
Mean of Dep. Var. in Full Sample	0.437	0.461	0.402	0.519
adj. R^2	0.002	-0.000	0.002	0.000

Notes: Standard deviations clustered at school level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Absolute versus Relative Performance Pay: Evidence from an Experiment targeting Child Malnutrition in West Bengal*

Prakarsh Singh†, Amherst College
Sandip Mitra, Indian Statistical Institute

November 15, 2013

Abstract

We carry out a randomized controlled experiment in West Bengal, India to test three separate performance pay treatments in the public health sector. Performance is judged on improvements in child malnutrition. First, we exogenously change wages of government employed child care workers through a basic level of absolute incentives. The second treatment introduces high absolute incentives. Finally, we also test for the impact of basic relative incentives on child health. All treatments include supplying mothers with recipe books. Overall, the results suggest that high absolute incentives reduce severe malnutrition by about 6.3% with controls and 4.9% without controls over three

*We are grateful to Hongye Guo, Alvaro Morales, Darius Onul and Arup Ratan Poddar for excellent research assistance and Chris Kingston, Jun Ishii and Geof Woglom for their comments. We also thank the mothers, workers and children who took part in this experiment and the enumerators who painstakingly carried out the interviews. Thanks also go to the Social Welfare Department at Maheshtala for their co-operation. Finally, we would like to thank the seminar participants at ISI Kolkata, Five College Seminar, University of the Basque Country and NOVA Business School.

†Corresponding author and PI: psingh@amherst.edu.

months. Large increases in weight are observed in boys as opposed to girls and same religion mother-worker pairs tend to be more productive in improving weight. This suggests that taste-based preferences may play a role even when workers are incentivized.

Word count: 7868

Keywords: performance pay; child malnutrition; absolute and relative incentives

JEL Classification: M52; I12; I38; J38

1 Introduction

Performance pay is often taken to mean piece rates where a payout is linked to only individual productivity or output. However, relative performance pay schemes (especially tournaments) are also widespread in organizations. In this study, we carry out a randomized experiment to compare absolute with relative performance pay in a public sector organization engaged in service delivery. The incentives are provided to child care workers to lower child malnutrition. We also test for the gradient of the absolute performance pay scheme in this experiment.

There are 1.31 million child day care centers in India under the Integrated Child Development Services (ICDS) government-sponsored program.[1] Each center is staffed by a child care worker who is responsible for supervising children between ages 3-6 years during day-time, providing them mid-day meals, teaching mothers about child nutrition, imparting non-formal pre-school education and facilitating health check-ups by doctors.[2] They are typically on fixed wages. The centers are widespread across India and admission is free. Nevertheless, quality of service delivery has often come under scrutiny (see

[1] Between 2007-2012, the value of the funds allocated by the central government to the Integrated Child Development Scheme was $7.4 billion (Lok Sabha, 2012).

[2] Centers also act as food distribution centers for pregnant women, adolescent girls and children under the age of three years.

for example, Chaudhury et al. (2006) on health worker absenteeism and Das and Hammer (2005) on quality of doctors in India). A recent household survey in 100 Indian districts reveals that 96 percent of the villages are served by these centers, although only 50 percent provide food on the day of survey and 19 percent of the mothers report that the center worker provides nutrition counselling (HUNGaMA report, 2011). A study on Bihar's day care centers found that 71 percent of the funds received by centers for the mid-day meal component were not spent on the beneficiaries and the performance of the programme should be monitored using outcome data rather than inputs (IGC report, 2013). Moreover, despite a high rate of economic growth in the past two decades, stunting in India is still higher than Sub-Saharan Africa and Jayachandran and Pande (2013) find a steep increase in stunting for higher birth order children in India.

One way of tackling the supply-side inefficiency can be through performance-based incentives. Performance pay has been shown to have a positive effect on productivity of workers in the public sector under certain conditions (Singh, 2013; Basinga et al. 2011) even though there could be crowding out of intrinsic motivation if incentives are too low (Gneezy and Rustichini, 2000). In Singh (2013), absolute performance pay to workers improves worker effort but child malnutrition only decreases when performance incentives are combined with information to the demand-side (recipe books given to mothers of children in day-care centers). In particular, the results from this experiment conducted in Chandigarh provided evidence for the complementarity between supply-side incentives and demand-side information in affecting child health.[3] However, it was not able to address how performance is af-

[3] In an experiment covering 145 child day-care centers, Singh (2013) implements three separate treatments. First, he engineers an exogenous change in compensation for child-care workers from fixed wages to performance pay. Second, he only provides mothers with information without incentivizing the workers. Third, he combines the first two treatments. Combining incentives to workers and information to mothers reduces weight-for-age malnutrition by 4.2% in three months. This complementarity is shown to be driven by more effective mother-worker communication and the mother feeding more calorific food

fected by a change in slope of absolute performance pay. This may be useful to know for policy-makers to have a better idea of the cost-effectiveness of such schemes. Secondly, there was no comparison of absolute pay with respect to relative pay, which may be more effective for public sector workers. Both these questions are addressed in the present experiment conducted in a different region of India – West Bengal in Maheshtala Municipality on the outskirts of Kolkata. As of December 2011, the percentage of malnourished children under 5 in West Bengal, in terms of weight-for-age, was 38.7%, below the national figure of 42.5% (Rajya Sabha, 2011). Exogenous treatments were assigned through cluster randomization contemporaneously along with a pure control group. All centers were based entirely within one geographic block and there was no endogenous selection into the treatment.

There has been very little experimental research on performance incentives in health programs of low and middle-income developing countries according to an excellent literature review by Miller and Babiarz (2013). An exception in providing incentives on the demand-side is Banerjee, Kothari and Duflo (2010) who incentivize immunization coverage in rural Rajasthan for mothers and find large increases in uptake. Hasnain, Manning and Pierskalla (2012) review the experimental evidence on performance pay in the public sector and conclude that it is extremely scarce, especially in developing countries.[4] The only two studies they cite in this sub-field are Basinga et al. (2011) in Rwanda and Singh (2013) in Chandigarh, India where positive effects are found on health. Another study by Miller et al. (2012) shows that financial incentives for anaemia reduction are modestly successful in China. Recently, De Walque et al. (2013) and Gertler and Vermeersch (2012) provide evidence to show positive effects of performance pay on HIV-testing and

at home.

[4]Christianson et al. (2008) make the same point about developed countries stating that control groups are often not comparable. One of the first experimental papers on performance incentives for nursing homes in US found that incentives lead to better resident health outcomes and shorter stays (Norton 1992). Figlio and Kenny (2007) find that merit pay for teachers matters and more so for parents who are more involved.

on child health outcomes in Rwanda. It is rare to implement performance incentives for service delivery in the public sector and also have a control group. Moreover, comparing relative and absolute incentive schemes as well as changing the gradient of the incentive pay have never been studied in the public health domain. Finally, there may be multitasking or gaming of performance incentives as shown by Figlio and Winicky (2005) but we do not find any evidence for gaming.

We carry out a randomized controlled experiment in West Bengal, India to test three separate performance pay treatments in the public health service delivery sector. Performance is judged on improvements in child malnutrition. First, we exogenously change wages of government employed child caregivers through a basic level of absolute incentives. The second treatment introduces high absolute incentives. Finally, we also test for the impact of basic relative incentives on child health. All treatments include supplying mothers with recipe books. Overall, the results suggest that high absolute incentives reduce severe malnutrition by about 6.3% over three months. There are no significant effects on health of the basic absolute or basic relative incentives during this period and the point estimates are close to zero. In order to check for common trends prior to the randomization we also have a placebo check that shows no differential trends between the different groups.

Latest research also focuses on social distance and interaction of incentives. Kingdon and Rawal (2010) show that a student's achievement in a subject in which the teacher shares the child's gender, caste and religion, is on average nearly a quarter of a standard deviation higher than the same child's achievement in a subject taught by a teacher who does not share the child's gender, caste or religion. However, Berg et al. (2013) find experimental evidence to suggest that incentive pay does not improve the transmission of knowledge with respect to households of similar characteristics as the agent but it does improve transmission with respect to socially distant households. Our results reveal that incentive pay interacts with improving health of boys more than girls. Similarly, a child is more likely to show significant gains in weight if the high-powered incentivized worker and the child's mother follow the same religion as compared to them following different religions.[5] This may indicate an interaction of taste-based preferences with incentives in improving child health, which could have implications about how an unequal distribution may result in response to performance incentives even as public sector efficiency is enhanced.

We delineate the theoretical framework in Section 2, the setting in Section 3, and treatment, methodology and empirical specification in Sections 4, 5 and 6 respectively. Section 7 reports the results and Section 8 provides policy implications and concludes.

2 Theoretical Framework

Theoretically, under risk neutrality, there is no difference between absolute and relative incentives for the policy maker to induce the same high effort for agents (Lazear and Rosen, 1981). With risk aversion and common shocks, however, relative incentives can dominate absolute incentives (Green and Stokey, 1983). Relative pay filters the common shock making agents face lower risk as compared to absolute pay. Under the relative incentive scheme, workers' pay depends on the ratio of individual productivity to average productivity among all co-workers in a field. Another practical advantage is that of budget predictability under the relative scheme. Recent experimental evidence by Gangadharan et al. (2013) suggests that an endogenous-prize tournament leads to a Pareto improvement in participants' payoffs and also increases collective output. Under the absolute incentive scheme – piece rates – individual pay only depends on individual productivity. Bandiera, Barankay and Rasul (2004) find that moving from relative to absolute incentives increases the productivity of a farm worker by 50%. In the public

[5] The caste of workers and mothers was not collected in this study.

sector, there has been no such empirical study comparing relative and absolute incentives. This may matter because of several reasons. First, shocks to output may be salient. Second, output may be less easily observable. Third, output may not directly respond to an increase in effort (may depend on demand-side also). Fourth, public sector workers may be more risk averse (Buurman et al., 2012). Fifth, workers may be differently motivated (Besley and Ghatak, 2005) and may exert more effort in response to non-financial awards as shown by Ashraf, Bandiera and Jack (2012). We motivate the treatments by proving that the costs associated with the absolute scheme are always greater than the relative scheme when a common shock is possible and public sector agents are risk-averse.

Assume one risk neutral principal (the government) and n risk averse agents (center workers), each responsible for producing output. In our scenario, the output could mean healthier (and less malnourished) children. Each agent exerts an effort and receives a wage from the principal. The agents' efforts are not visible to the principal, and therefore the principal can only contract based on the agents' output levels. For simplicity, we assume that the agents' effort levels are discrete, namely 0, 1, and 2. Agents produce two levels of outputs: 0 or 1. The probability of attaining the high output depends on the agents' effort level: the higher the agent's effort, the higher her chance for producing 1 as opposed to 0.

We also assume a common shock in production. This could be thought of as a shock that causes output to shrink for all workers and is outside their control. In our setting, this could represent the onset of a disease (for example, malaria especially during the rainy season). With probability θ, all the agents produce 0 regardless of their effort level. The agents' utility function is given as follows:

$$U_A = V(w) - \frac{1}{2}cx^2$$

.

Here, w is the wage this agent receives, c is the multiplier of the disutility caused by the agents' effort and x is the agents' effort level. Agents are assumed to be risk averse, thus V is an increasing and concave in w.

The principal's utility function is as follows:

$$U_p = M\sum_{i=1}^{n} p_i - \sum_{i=1}^{n} w_i$$

p_i is the output level for agent i, n is the number of the agents, M is the multiplier for the agents' outputs.[6] Moreover, it is always worthwhile for the principal to engage the agents in the highest level of effort ($x = 2$). Thus, the products are sufficiently valuable to the principal, i.e. M is sufficiently large.

2.1 Cost for an Absolute Performance Pay Scheme

Suppose the wage for the agent is a for output $= 0$, and b for output $= 1$. Clearly $b > a$. The expected utility for agents with each of the different effort levels is:

$$E(U_A(0)) = V(a)$$

$$E(U_A(1)) = \frac{1+\theta}{2}V(a) + \frac{1-\theta}{2}V(b) - \frac{1}{2}c$$

$$E(U_A(2)) = \theta V(a) + (1-\theta)V(b) - 2c$$

To implement the highest effort level, the principal needs to make sure that the agent's utility with highest level of effort ($x = 2$) is the greatest in comparison to the other two effort levels. This boils down to the following condition:[7]

[6] M is assumed to be greater than one.

[7] See Appendix for details.

$$V(b) - V(a) > \frac{3c}{1-\theta}$$

Thus, the difference between the utilities generated by the high wage and the low wage should be large enough. The principal could achieve this by lowering a. However, practically, there is a lower bound on a. If an agent produces 0, the principal needs to make sure that the agent can survive on her wage. This condition is particularly relevant in the public sector, where workers are difficult to fire and where unionization resists cutting down on wages of unproductive workers. We can suppose that $V(a) > U$. It follows that $V(b) > \frac{3c}{1-\theta} + U$. The total cost for the principal to implement a high effort for all agents under the absolute performance pay is as follows:

$$C_{absolute} = n \times (\theta V^{-1}(U) + (1-\theta)V^{-1}(\frac{3c}{1-\theta} + U)).$$

2.2 Cost for a Relative Performance Pay Scheme

When there are n agents, a complete relative performance pay scheme should have $n+1$ contingencies, which boils down to $2 \times (n-1) + 2 = 2n$ number of payments. For example, in a three agents case, a complete tournament scheme would need to specify payments for output combinations (1, 1, 1), (1, 1, 0), (1, 0, 0), (0, 0, 0), which contains six kinds of pay. Denoting the pay structure $(H_{0,}L_0; H_{1,}L_{1;}...; H_{n,}L_n)$, H corresponds to pay on high output, and L corresponds to low output pay. The subscript n on L_n and H_n corresponds to the number of agents who produce low output.

When all agents produce the same output – be it 1 or 0 – the relative rank for each agent is the same. For example, if there is a common shock, output will be low for all agents and all of them will receive L_n. On the other hand, if everyone exerts high effort and there is no common shock, the output will be high and all agents will receive H_0. As the rank structure is exactly the same in both cases, all workers should receive the same wage under a relative scheme: $H_0 = L_n$.

We only need to compute the cost for the principal to make effort level portfolio (2, 2, 2,..., 2) a Nash Equilibrium. agent choose 0 or 1 over 2, provided everyone else chooses 2. From our previous assumptions,

$$H_0 = L_n = T$$

Assuming that $(n-1)$ agents have already chosen effort level of 2, the expected utilities of the n^{th} agent to choose 0, 1, 2 are respectively:

$$E(U_A(0)) = \theta V(T) + (1-\theta)V(L_1)$$

$$E(U_A(1)) = \frac{1+\theta}{2}V(T) + \frac{1-\theta}{2}V(L_1) - \frac{1}{2}c$$

$$E(U_A(2)) = V(T) - 2c$$

The inequality that arises out of solving the above problem along with the minimum utility assumption leads to $V(T) \geqslant U + 3c$. The minimized total cost for a principal using tournament is

$$C_{relative} = n \times V^{-1}(U + 3c).$$

2.3 Comparison between absolute and relative schemes

The difference between the two cost functions is as follows:

$$C_{absolute} - C_{relative} = n(\theta V^{-1}(U) + (1-\theta)V^{-1}(\frac{3c}{1-\theta} + U) - V^{-1}(U + 3c)).$$

Notice the difference is actually 0 when $\theta = 0$. To see how this difference varies with θ, we take its derivative with respect to θ. We can prove that when $V(w)$ is concave:[8]

$$C_{absolute} > C_{relative} \text{ for } \theta > 0.$$

The cost for an absolute scheme is higher than that for a relative scheme whenever common shock is possible. This result is in contrast to the stylized one in the current literature, which states that the absolute scheme should dominate the relative schemes when the probability of common shock is low enough or when it is not present (Green and Stokey, 1983). The "public sector assumption" that assumes that agents should obtain a minimum level of utility even when they perform poorly leads to the principal optimally choosing a more cost effective relative scheme for extracting high effort from its agents. How these schemes fare in terms of heterogeneous effects of social interaction between mother and worker, depends on whether the communication is able to help in filtering the common shock, in which case the relative treatment would lose some of its advantage over the absolute treatment. However, communication may also signal greater transfer of a common shock through contagion. In this scenario, relative treatment would increase its superiority in the face of a negative shock.

3 Setting

There are 1.31 million day care centers across India that offer child care and nutritional counseling services. These are run by the government under the umbrella of Integrated Child Development Services (ICDS) through the Social Welfare Department. Each center is usually staffed by one government worker and an assistant. Workers can affect health of the child through two primary channels: first, providing mid-day meals to children and second, advising mothers on a nutritious diet. We study child care workers employed by the West Bengal government in Maheshtala Municipality in 24 South

[8] See Appendix for proof.

Pariganas District. Maheshtala is located in the Kolkata Metropolitan Region. According to the 2011 Indian census, the population of Maheshtala was 449,423. The sex ratio of Maheshtala city was 945 females per 1000 males. Average literacy rate of Maheshtala was 82.63 percent of which male and female literacy was 86.08 and 78.98 percent. Children constituted 9.67 % of the total population of Maheshtala. We were able to carry out this study in Maheshtala as the Social Welfare Department was keen to implement an experiment to tackle malnutrition. Moreover, Maheshtala happens to be understudied and has malnutrition comparable to slum areas in the Chandigarh experiment analyzed in Singh (2013). Here, the worker in a day care center has a fixed monthly salary of Rs. 4350 in Kolkata, which increases to Rs. 4413 after 10 years of service. All workers in this Municipality have similar tasks and operate under the ICDS scheme. Education, knowledge and experience of worker along with quality of infrastructure in the centers are controlled for in regressions below.

4 Treatment

There were three treatments that were implemented in the research project. All three treatments entailed free distribution of recipe books to all mothers apart from performance pay to workers. This was done because the combined treatment of incentives and information in Singh (2013) had been shown to increase weight in children and only incentives to workers or only recipes to mothers were individually ineffective. The recipe book is described later in the section. To understand each of the performance pay treatments, consider the categories established by WHO (2007), as these are also employed by the government workers. A child is classified as malnourished if she is more than two standard deviations away from her WHO standard weight-for-age mean.[9] She is moderately malnourished if she is more than two standard deviations

[9] The standard also differs for males and females.

but less than three standard deviations from the weight-for-age mean. A child is severely malnourished if she is more than 3 standard deviations away from the weight-for-age mean.

The first treatment, titled Basic Absolute (BA treatment), was the replication of the combined treatment from Singh (2013), which had reduced malnutrition by 4.2 percent over three months. This entailed Rs. 100 per child if the child's malnutrition grade improved from severe to moderate or moderate to normal and a corresponding Rs. 100 deduction from the total for a drop in grade from normal to moderate or moderate to severe. In other words, if:

N for each worker = # children who jump at least one grade - # children who drop at least one grade

The total payment promised for each worker was Rs. 100 * N after three months.[10] From the earlier experiment, it was not clear how the slope of performance pay was related to effort. In the second treatment, called High Absolute (HA treatment), the payout promised to the workers after three months was Rs. 200 * N, where N was defined in the same way. However, if more children suffered declines as opposed to improvements, the workers were not asked to make payments. As the schemes were implemented in the presence of senior staff and under the signed approval of the local Director of Social Welfare Department, the promises could be seen as credible commitments to the workers. For example, if the number of children who jump from severe to moderate = 4 and the number of children who drop from normal to moderate = 2, the total payment would be Rs. (4-2) * 100 = Rs. 200 in the BA treatment and (4-2) * 200 = Rs. 400 in the HA treatment.

The average payout in Chandigarh for the basic absolute incentive with recipe book had been Rs. 291 per worker but the monthly worker salary was Rs. 2000 per month in 2010. Since the workers' salary was almost twice

[10] Approximately, 1 US$ = Rs. 60 in July, 2013.

in Maheshtala as compared to Chandigarh, the high absolute treatment can also be considered a test of the previous experiment's treatment if the slope is considered to be a proportion of the salary. Thus, the first two treatments can also throw light on the relative effectiveness of a constant slope of Rs. 100 versus the same slope as a proportion of total income.

The third treatment, or the basic relative (BR treatment), was allocated a pot of money containing on average Rs. 291 per worker to keep it consistent with the ex-ante expected payout in the basic absolute treatment.

T for each worker = sum of all positive N in that treatment group
Payout = 55 workers * 291 * (N/T) = 16000 (N/T)

The total amount for 55 workers would be approximately Rs. 16000. However, the payout to each worker would depend upon her performance relative to others in the group. For example, if the worker's N = 10 and the sum of all positive N = 100 in her treatment group of 55 workers, the worker would get 10 percent of the total amount = Rs. 1600. If each worker performs equally, they each get approximately Rs. 300 in the basic relative treatment.[11] The relative treatment is also similar to Bandiera et al. (2004), where the workers' pay depended on the ratio of individual productivity to average productivity among all co-workers in a field. None of the schemes could reduce workers' income, and all were accompanied by information provided directly to mothers.

Workers in each treatment group participated in three separate workshops at the end of the baseline round (one for each treatment), where they were handed goal cards. The goal cards listed target weight for each child in their center after three months. Goal was the threshold for achieving moderate malnutrition status for the presently severely malnourished child (after accounting for the increase in age at endline) and achieving normal malnutrition status if the child is currently moderately malnourished. If the child

[11] These wards were numbered 1, 10, 14, 18, 22, 25, 29, 30.

was currently in the normal range, a maximum threshold was provided below which the child would become malnourished and penalty imposed. The workers were told about their respective treatments with the help of illustrative examples and all doubts were clarified.

The recipe book that was distributed in all treatments had ten economical and nutritious recipes for 3-6 year old children. These were government approved recipes and the recipes were chosen with the help of the Food and Nutrition Board, Kolkata. Each recipe could be made within a budget of Rs. 5 at home, contained local ingredients and listed the step-by-step method of preparation and nutritive values per 100 gms. The individual nutritive values were of calories, protein, iron and carotene. The book was translated and printed in Bengali, the local language. Some of the recipes were as follows: Puffed Rice Bengal Gram Mix, Rice Food Mix, Suji Porridge, Dalia Porridge, Chidwa Pulao and Chidwa Laddoo. These were rich in protein and calories to counter child malnutrition. Most used either lentils for increasing protein and rice, wheat or jaggery for increasing calorie count.

5 Methodology

After receiving approval and preliminary data from the Social Welfare Department, Government of West Bengal on the number of registered students and average malnutrition in each center, manual randomization at the cluster level was conducted at the Department by a lottery. Each cluster was made up of several centers. All centers with fewer than 20 registered students were dropped from the sample. Out of 35 clusters in the Municipality, one of clusters (number 28) did not have any centers with 20 or more students registered at the time. In total, 34 clusters were selected for the study covering 209 centers. The senior department officials (supervisors) asked to be involved in the selection procedure, so they were invited to participate in the lottery.

Four boxes were placed in the Department's head office, each corresponding to an undisclosed treatment or control group. The supervisors placed a folded slip (representing the cluster number and total centers in that cluster) inside each box sequentially. When the total number of centers accumulated for a treatment exceeded 50, the assistant was asked to shut the box. This was done to have at least 50 centers in each group for adequate sample size according to power calculations. This would also avoid the exceptional case where the wards with the highest or fewest number of centers get assigned to one treatment by chance. Cluster was chosen as the level to randomize to remove the possibility of contaminating spillover effects between workers within a cluster. Table 1 shows the total clusters and centers allocated to each group and Figure 1 plots the assignment on a map provided by the Department.

<Table 1 and Figure 1 about here>

The baseline was carried out during March-May, 2012 and the endline three months later between August-September, 2012. A window of three months was chosen for the experiment because it is the average time between two medical check-ups by the local Health Department. The duration was verified to be sufficient for a grade improvement to occur by doctors at the local office of the Health Department, Government of India and was comparable to the earlier experiment. At baseline and endline, a team of enumerators (supervised by an assistant and project manager) weighed all children present in the center on a digital weighing machine, interviewed their mothers and the center workers. The recipe books were distributed to all the mothers (except in the control group) after their interviews were taken at the centers. Previous weights of children (on average two months prior to baseline) were also recorded at baseline from the weight record registers of the workers.

6 Empirical Specification

The main regression specification for finding the average effect of the treatments on weight of a child is as follows:

$$\begin{aligned} w_{ijt} &= \alpha(post)_t + \beta(BA)_j + \gamma(HA)_j + \rho(BR)_j + \eta(post * BA)_{jt} + \\ &\quad \theta(post * HA)_{jt} + \omega(post * BR)_{jt} + X_{ijt} + \varepsilon_{ijt} \end{aligned}$$

w_{ijt} is the weight of a child i in cluster j at time t. The variable *post* is a dummy that is 0 for baseline and 1 for endline. The variables BA, HA and BR are 1 if the child is in the treatment basic absolute, high absolute or basic relative respectively and 0 otherwise. The omitted category is the control group. X_{ijt} are individual and center specific controls specified in the following section. The error term is clustered at the cluster level. The variable *post* accounts for the natural increase in weight in three months, all seasonal effects on weight, regional shocks to food prices and any management changes or unobservables that would impact all groups in the same way. β, γ and ρ are the baseline differences between the individual treatments and the control. η, θ and ω give us the difference-in-differences estimates for the effect of each of the three treatments. This interpretation rests on the identification assumption that there are no time varying and group-specific effects that are correlated with the treatments (common trend assumption). As the clusters were randomly assigned into one of the four groups, we should not expect there to be common trends amongst the groups.

Although usually it is not required to check this assumption with randomization, we carry out a placebo check to corroborate that pre-trends are similar across all groups. For the placebo check, we define $post = 1$ for baseline and 0 for the weight recorded in registers prior to baseline (on average about two months before). Running the above regression with this new definition should allow us to test if there are changes in the difference-in-difference estimates from what we had obtained earlier. We should not observe any significant difference-in-difference estimates with the placebo regression for common pre-trends assumption to hold.

7 Results

The summary statistics at baseline in Table 2 reveal that mother's age is on average 27 years and is similar across the four groups.[12] 73% of the mothers in the control group can read and this is also similar across groups, which is comparable to the 2011 Indian Census finding of 78% literacy for women in Maheshtala. The normalized differences show that differences between groups are not significant as long as the normalized difference is less than 0.25 (Imbens and Rubin, 2007). The literacy rate for women is particularly high in comparison to only 45% in Chandigarh. The weights and ages of children (around 13.5 kilograms and 4.2 years) as well as the malnutrition status are similar across all groups. The malnutrition rate is 33% at baseline with close to 9% being severely malnourished and the rest being moderately malnourished. The monthly income of a household (approximately $70) of four members (two adults and two children) is very similar to the data collected in Chandigarh, showing that families on average live below the poverty line in this sample. This is also statistically insignificant between the groups with HA group being the most similar to the control group along all variables. Ownership of mobile phones is high and water filters is low similar to Chandigarh. The least similar group, overall, is the BA group where mobile phone ownership and presence of water tap at home is significantly higher than the control. Anganwadi infrastructure as measured by presence of access to drinking water in the center or toilet in the center are also similar with close to 40% having a toilet and three quarters having drinking water. Thus,

[12] Mother's age was around 28 years for the experiment in Chandigarh (Singh, 2013).

on average the variables appear to be well matched across groups with the HA group being the most similar to the control, followed by the BR group and then the HA group.

<Table 2 about here>

We might be worried about selective weighing of children in these groups despite checking by an independent supervisor and enumerator. Table 3 shows attrition rates. These are around 26% and similar across the four groups. Attrition rates tend to be higher in these centers as most students use the centers as a temporary pre-school before they gain admission to a government school. This is because the pre-school educational quality is poor with no official syllabus or exams. As many of the fathers are daily wage laborers who are prone to migrating where a higher wage is offered, their family also keeps moving with them. Attrition is not different by the main outcome variables depending upon treatment. This is illustrated by Appendix Table A2 where all coefficients of the variable interacting treatments with main outcome variables (weight, malnourished, grade, severe malnourished) are insignificant. Stand-alone coefficients on health outcomes and individual treatments are also insignificant (not shown).

<Table 3 about here>

The regression results from specification (1) are reported in Table 4 for a panel of children who were weighed twice during the study. The dependent variables in the first four columns are weight, dummy for malnourished status, weight-for-age grade (ordered 0 to 2 from normal to severely malnourished) according to the WHO, and dummy for severely malnourished status. The results reveal that the weight increase in the control group over three months was on average 283 grams and significant (similar to Chandigarh). The baseline levels of weight and other measures are similar in the four groups. There appears to be no significant impact of the basic absolute (HA) and basic relative (BR) treatments. Moreover, BR appears to do slightly but insignificantly better than BA. The high absolute treatment shows a significant effect of an increase of 191 grams over and top of the 283 grams. Although malnutrition decrease is not significant at the 10% level, there appears to be a 4.65% reduction just under the 10% significance level. Relative risk of death from infection by malnutrition is twice as high for severely malnourished as for moderately malnourished children and nine times higher than normal weight children (Caulfield et al., 2004). Moreover, more than half of all deaths in young children through diarrhea, pneumonia, malaria and measles are attributable to undernutrition (Caulfield et al., 2004). Ordered grades decrease (as 0 is normal) on average in the HA treatment. The weight increase is driven by the movement towards moderate status of severely malnourished children as can be shown by column (4). Severe malnutrition declines by 4.88% and is significant at the 5% level. The next four columns have the same dependent variables but also include control variables. The control variables used in columns (5)-(8) are household demographics (age of child, gender of child, total number of siblings), household assets (monthly income, food expenditure, number of rooms, proportion of goods owned in the kitchen, proportion of non-kitchen goods), parent-specific controls (mother's age, mother's religion, whether the mother is a housewife, literacy of mother and father, whether the mother scored a high or low quiz score), worker-specific controls (whether the worker is experienced or not, whether the worker is highly educated or not, whether the worker scored a high or low quiz score), and center-specific controls (dummy variables for the center's facilities: electricity, fan, helper, chart, blackboard, drinking water, and toilet). The preceding result on severe malnutrition is now more significant and the decline in severe malnutrition is now estimated to be 6.3% for

HA treatment.[13] This is a big decrease given the time period and may be able to sustain itself after three months because of immunity and resistance acquired by moderately malnourished children. In the Chandigarh experiment, there was a similar reduction in overall malnutrition over three months (driven by more malnourished children), and malnutrition did not go back to its original level after a year when the incentive treatment was no longer in place. Next, we consider the placebo results wherein we look for differences in the rates of growth in children between the groups pre-treatment. This is shown in Appendix Table A1. We observe that even on average two months prior to treatment, the weight increase was not different across the different treatment and control groups and they were on same trajectories. There may be measurement error here as the weights were recorded by the center workers in their registers but there is no reason to expect a systematic upward or downward bias in any of the treatment groups. Figure 2 shows the differential trend in the high absolute treatment immediately after the treatment (located at two months) in comparison to the control group. The pre-baseline and baseline values are very similar in the high absolute and control groups. It is interesting to note that basic relative treatment appears to do better than the basic absolute treatment in the graph but the difference is not statistically significant and neither is the difference at baseline between these groups and the control group significant.

<Table 4 and Figure 2 about here>

As high absolute treatment seems to be effective in improving weights of children, we can conclude that slope of the incentive treatment as a proportion of salary matters in this context. In Appendix Table A3, we reproduce the main regression using Moulton clustering correction (as opposed to the standard clustering) as suggested by Angrist and Pischke (2008) in situations where the main regressors (treatments) are fixed within a cluster and

[13] The results are robust to including an intermediate set of controls. For details, see Table A4 in the Appendix.

heteroskedasticity is not a huge problem. While all the coefficients stay the same, the standard errors increase across the board, as expected, leading to a few changes in the degree of significance of some coefficients. Thus, the coefficient Post*HA becomes insignificant in the weight regression, but it is still significant at 5% and 10% in the grade and severely malnourished regressions, respectively. To better understand the channels, we would like to test if mothers report changes in diet at home in this treatment as this would lend more credence to the change in outcomes. In Table 5, we explore the demand-side channel of mother-reported diet given to child. The dietary variables considered here are lentils or pulses, fish, meat, green vegetables and sweets or desserts. Note that baseline levels of consumption are similar across all groups. Lentils or pulses intake (at least twice a week) shows a significant increase in consumption and so does meat. Lentil intake is consistent with mothers being asked to prepare recipes by the HA treatment workers. Recall that several recipes (six out of ten) contained lentils as their main ingredient. This is rich in protein and would help lower protein deficiency. There is no change in other types of dietary intakes for the HA group mothers. The mothers in the BA and BR group show no change in their reported food provision, which is again consistent with their results on weight change in Table 4.

<Table 5 about here>

There could be exogenous factors that may cause certain mothers to be more intensely targeted with information by incentivized workers. As there were a large number of inter-religious interactions in this setting, we explore the role of religious homogeneity in the weight change. First, we classify a Hindu-Hindu or Muslim-Muslim mother-worker interaction as same religion. Then we run the health outcomes' specification for same religion or different religion pairs. In Table 6, we find that quality of communication may be better with same religion mother-worker pairs in the control group as there is a significant increase in weight of 348 grams for these pairs but not for different religion pairs, consistent with Kingdon and Rawal (2010). However, we also show that the increase in weight is high and significant for same religion pairs as opposed to different religion pairs in the HA treatment. This may highlight a concern with provision of incentives in areas that have ethnic diversity. Incentives may crowd-out the worse outcomes generated by social distance (Berg et al., 2012), but we find that they can also accentuate outcomes, with incentivized workers only working to increase weights of children who are in the same social group. This may happen due to taste-based preferences or a lower cost of communicating with mothers effectively. Nevertheless, more effective social interaction can help in filtering the common shock making absolute performance pay less costly in inducing high effort. Another channel which appears to be important but cannot be tested is the provision of midday meals in the center that may have increased for all children (reduction in leakage).

<Table 6 about here>

Next, we explore heterogeneous effects of the treatments by sex of child. In the Chandigarh experiment, there were no gender-differential effects of the incentive treatment. In Table 7, we run a triple difference regression to check for significant differences in average treatment effects for boys and girls. Surprisingly, we find that boys show a much greater increase in the high absolute treatment relative to girls (345 grams higher and this is significant at the 10% level). This reinforces the taste-based preferences channel that we observed earlier with same religious pairs as there may be an underlying gender bias that becomes more salient with the introduction of high absolute incentives. It may also be that boys have greater appetite or ask for more food in the centers and thus are able to get more food from the incentivized worker. Workers may expect that boys will show more weight gain in response to food in the center or that mothers will be less likely to cut down on food for boys leading to complementarity. However, this expectation may be incorrect, as we do not observe differential food intakes reported at home for boys and girls suggesting that gender bias is getting triggered at the center and not at home.

<Table 7 about here>

Finally, we check for selective targeting of children who are close to the target threshold that may occur at the expense of those who are further away. The variable 'closetotarget' is defined to be 1 if the child is malnourished at baseline and her deficit weight (difference between the target weight and the actual weight) is less than the mean deficit weight and 0 otherwise. Target weight is defined as the threshold at which the child will improve her grade (i.e. would go from severely malnourished to moderately malnourished or from moderately malnourished to not malnourished). If there was selective targeting of those children in the incentive treatments who are close to the target threshold, the disincentive for a drop in grade may not have been as effective as anticipated. However, in Table 8, we observe that children close to the threshold do not appear to be selectively targeted in any of the treatments. Children who are closer to the target weight do not increase their weight at a statistically different pace than the ones who are further away from their target, which provides evidence against gaming of positive incentives. This seems to reiterate the importance of having disincentives for worse outcomes along with incentives for better outcomes.[14]

<Table 8 about here>

[14] One may also be concerned about kids being given water by workers before the measurement of weights at endline. However, 'center has access to drinking water' is one of the controls used in the main regression, and this access is associated with an insignificant and negative increase in weight on average.

8 Policy implications and Conclusion

There has been no research in the public health sector in developing countries focusing on the elasticity of outcomes with respect to gradient of performance pay and also comparing absolute with relative treatments. We carry out a randomized controlled experiment to test three separate performance pay treatments in the public health sector. First, we exogenously change wages of government employed child care workers to a component with basic absolute incentives to lower child malnutrition. The second treatment introduces high absolute incentives. Finally, we also test for the impact of basic relative incentives on child health. All treatments also include supplying mothers with recipe books. Overall, the results suggest that high absolute incentive works to reduce severe malnutrition by about 6.3% with controls and 4.9% without controls over three months. This is a strong result and is in line with an earlier experiment carried out in Chandigarh in Singh (2013) that found a benefit-cost ratio of around 20. As the government expands access to centers across the country, results from this paper suggest that worker pay should not remain fixed even though the gains may be distributed unequally. Compensation needs to have a component of performance pay that increases as the fixed income goes up. The component should be based on weight-for-age grade that is easily observable and well understood. Additionally, the worker should not have perverse incentives to reduce weights of children who are not malnourished. Apart from incentivizing workers, it is also imperative to create awareness on the demand-side. One way of spreading this information is through simple and concise recipe books that are customized by regional tastes, availability of ingredients and social norms.

Possible channels that may be generating this high incentive effect on weight include using recipe books to remind mothers to increase protein intake at home and also food distribution by worker in the center. Large increases in weight are observed in boys as opposed to girls, and same religion mother-worker pairs tend to be more productive in improving weight in the high absolute treatment than mother-worker pairs with different religions. This suggests that taste-based preferences may play a role even when workers are incentivized. In the future, we would also like to test the high relative treatment which we were unable to test here due to a small sample size. There could also be research on changing the slope of performance incentives along an observable dimension like gender to target both efficiency and equity.

The value of the funds released for ICDS by the central government in 2011-2012 was, for West Bengal, equal to $193.6 million. If the state-allocated funds are included, this figure goes up to $221.5 million. A large proportion of these funds go towards fixed wages of child care workers. Yet, little is known about the most effective way of organizing labor contracts in this important area of public sector service. We need to have a better handle at understanding what nudges work to motivate child care workers and reduce malnutrition rates, which have remained stagnant in India despite economic growth. Even without external validity, this has potentially life-saving implications on a large scale. Nevertheless, different settings within India can pose very different challenges, but one advantage of working within the same public organization can be easier replicability. We hope that further experimentation in this area can inform policy-makers how to make public health service delivery more efficient.

References

[1] Angrist, J. D. & J. S. Pischke (2008). Mostly harmless econometrics: An empiricist's companion. Princeton University Press.

[2] Ashraf, N., O. Bandiera, and K. Jack (2012) No Margin, no Mission? A Field Experiment on Incentives for Pro-Social Tasks, CEPR Discussion Paper No. 8834.

[3] Bandiera, O., I. Barankay, and I. Rasul (2004) Relative and Absolute Incentives: Evidence on Worker Productivity, CEPR Discussion Paper No. 4431.

[4] Banerjee, A., D. Kothari, and E. Duflo (2010) Improving Immunisation Coverage in Rural India: Clustered Randomised Controlled Evaluation of Immunisation Campaigns with and without Incentives, *BMJ*, 340, c2220.

[5] Basinga, P., P. Gertler, A. Binagwaho, A. Soucat, J. Sturdy, and C. Vermeersch (2011) Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation, *Lancet*, 377(9775), 1421 – 1428.

[6] Berg, E., M. Ghatak, R. Manjula, D. Rajasekhar, S. Roy (2013) Motivating Knowledge Agents: Can Incentive Pay Overcome Social Distance?, CEPR Discussion Paper No. DP9477.

[7] Besley, T. and M. Ghatak (2005) Competition and incentives with motivated agents. *American Economic Review,* 95, 616-636.

[8] Buurman, M., J. Delfgaauw, and R. Dur, S. Van den Bossche (2012) Public sector employees: Risk averse and altruistic?, *Journal of Economic Behavior & Organization*, 83(3), 279-291.

[9] Chaudhury, N., J. Hammer, M. Kremer, K. Muralidharan, and F. Rogers (2006) Missing in Action: Teacher and Health Worker Absence in Developing Countries. *The Journal of Economic Perspectives*, 20(1), 91-116.

[10] Christianson, J., S. Leatherman, and K. Sutherland (2008) Lessons From Evaluations of Purchaser Pay-for-Performance Programs—A Review of the Evidence, *Medical Care Research and Review*, 65(6), 5S-35S.

[11] Census of India (2011) Office of the Registrar General & Census Commissioner, India (ORGI).

[12] Das, J. and J. Hammer (2005) Which doctor? Combining vignettes and item response to measure clinical competence, *Journal of Development Economics*, Elsevier, vol. 78(2), pages 348-383, December.

[13] De Walque, D., P. Gertler, S. Bautista-Arredondo, A. Kwan, C. Vermeersch, J. de Dieu Bizimana, A. Binagwaho, J. Condo (2013) Using Provider Performance Rewards to Increase HIV testing and Counselling services in Rwanda, World Bank Policy Research Working Paper No. 6364.

[14] Figlio, David N. and Lawrence W. Kenny, 2007, "Individual teacher incentives and student performance," Journal of Public Economics, Elsevier, vol. 91(5-6), pages 901-914, June.

[15] Figlio, D. N., and J. Winicki, 2005, "Food for Thought: The Effects of School Accountability Plans on School Nutrition," Journal of Public Economics, 89, 381-94.

[16] Gangadharan, L., G. Musto, J. Rulliere (2013) Win for Being the Best or for Getting the Best: Theoretical and Experimental Evidence on Tournaments with Endogenous Prizes, 2013 Conference on Tournaments, Contests and Relative Performance Evaluation.

[17] Gertler, P. and C. Vermeersch (2012) Using Performance Incentives to Improve Health Outcomes (June 1, 2012). World Bank Policy Research Working Paper No. 6100.

[18] Gneezy, U. and A. Rustichini (2000) Pay Enough or Don't Pay at All, *The Quarterly Journal of Economics*, 115 (3), 791-810.

[19] Green, J., and N. Stokey (1983) A Comparison of Tournaments and Contracts, *Journal of Political Economy*, 91(3), 349- 364.

[20] Hasnain, Z., N. Manning, J. Pierskalla (2012) Performance-related Pay in the Public Sector : A Review of Theory and Evidence, World Bank Policy Research Working Paper No. WPS6043.

[21] HUNGaMA Survey Report (2011) The Hunger and Malnutrition (HUNGaMA) Survey Report 2011, The Naandi Foundation.

[22] IGC Report, 2013, Assessment of Bihar's ICDS Supplementary Nutrition Programme, IDinsight, available at: <http://www.theigc.org/events/assessment-bihar%E2%80%99s-icds-supplementary-nutrition-programme>

[23] Jayachandran, S. and R. Pande (2013) Why are Indian Children Shorter Than African Children?, Harvard Kennedy School Working Paper.

[24] Kingdon, G. and S. Rawal (2010) Akin to My Teacher: Does Caste, Religious and Gender Distance Between Student and Teacher Matter? Some Evidence from India. Department of Quantitative Social Science Working Paper 10-18, Institute of Education, University of London.

[25] Lazear, E. (2000) The Power of Incentives, *American Economic Review*, 90(2), 410-414.

[26] Lazear, E. and S. Rosen (1981) Rank-Order Tournaments as Optimum Labor Contracts, *Journal of Political Economy*, 89(5), 841-864.

[27] Lok Sabha (2012) Unstarred Question No. 1343, 21.03.2012. *available on IndiaStat.com*

[28] Miller, G. and K. Babiarz (2013) Pay-For-Performance Incentives in Low- and Middle-Income Country Health Programs, NBER Working Paper No. 18932.

[29] Miller, G., R. Luo, L. Zhang, S. Sylvia, Y. Shi, P. Foo, Q. Zhao, R. Martorell, A. Medina, S. Rozelle (2012) Effectiveness of provider incentives for anaemia reduction in rural China: a cluster randomised trial, *BMJ*, 345, e4809.

[30] Moulton B.R. (1986) Random Group Effects and the Precision of Regression Estimates, *Journal of Econometrics* 32, 385-397.

[31] Norton, E.C. (1992) Incentive regulation of nursing homes. *Journal of Health Economics,* 11, 105-128.

[32] Rajya Sabha (2011) Unstarred Question No. 2808, 19.12.2011. *available on IndiaStat.com*

[33] Singh, P. (2013), Performance Pay and Information: Reducing Child Undernutrition in India. Available at SSRN: http://ssrn.com/abstract=2270751 or http://dx.doi.org/10.2139/ssrn.2270751

[34] WHO (2007), The WHO Child Growth Reference Standards. Available at: <http://www.who.int/childgrowth/en>.

9 Tables and Figures

Table 1: Total clusters and centers in each group

	Total clusters	Total centers
Basic Absolute Treatment	8	50
High Absolute Treatment	8	53
Basic Relative Treatment	8	55
Control group	10	51

Table 2: Summary statistics at baseline

					Normalized Differences		
Variables	BA	HA	BR	C	BA-C	HA-C	BR-C
Mother's age	28.09	27.64	27.38	27.47	0.08	0.02	-0.01
	[5.11]	[5.43]	[5.23]	[5.67]			
Mother can read	0.73	0.67	0.75	0.73	0.00	-0.09	0.03
	[0.44]	[0.47]	[0.43]	[0.44]			
Monthly income	4556.27	4059.73	4400.02	3867.94	0.19	0.05	0.19
	[2608.34]	[2567.92]	[2823.25]	[2513.68]			
Number of rooms	1.43	1.39	1.43	1.40	0.02	-0.01	0.02
	[1.08]	[0.88]	[0.90]	[0.87]			
Mobile phone	0.84	0.66	0.83	0.66	0.30	0.00	0.28
	[0.37]	[0.47]	[0.37]	[0.47]			
Water tap at home	0.35	0.14	0.18	0.14	0.35	0.00	0.08
	[0.48]	[0.35]	[0.39]	[0.35]			
Weight of child	13.65	13.45	13.62	13.49	0.05	-0.01	0.04
	[2.05]	[2.09]	[2.28]	[2.10]			
Age of child	4.23	4.21	4.23	4.18	0.04	0.03	0.04
	[0.83]	[0.84]	[0.81]	[0.85]			
Fraction female	0.51	0.52	0.50	0.50	0.01	0.03	0.00
	[0.50]	[0.50]	[0.50]	[0.50]			
Total Siblings	1.29	1.15	1.21	1.43	-0.07	-0.15	-0.12
	[1.35]	[1.29]	[1.26]	[1.30]			
Toilet in AWC	0.39	0.30	0.49	0.39	0.00	-0.13	0.14
	[0.48]	[0.46]	[0.50]	[0.49]			
Drinking water in AWC	0.68	0.67	0.80	0.74	-0.09	-0.11	0.10
	[0.47]	[0.47]	[0.40]	[0.44]			

Notes: Standard deviations in parenthesis. Normalized differences are calculated using the formula as in Imbens and Wooldridge (2009) for a scale-free measure of the difference in distributions. A rule of thumb is that when normalized difference exceeds 0.25 in absolute value, linear regression methods tend to be sensitive to the specification (Imbens and Rubin (2007)).

Table 3: Attrition rates in children across groups

	BA	HA	BR	C
Baseline children	1333	1555	1369	1264
Endline children	971	1127	1031	933
Attrition (%)	0.27	0.28	0.25	0.26

Table 4: Main Results with and without Control Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Weight	Malnourished	Grade	Severe Malnourished	Weight	Malnourished	Grade	Severe Malnourished
Post	0.283***	-0.00291	0.00517	0.00807	0.271**	0.0234	0.0317	0.00830
	(0.0611)	(0.0206)	(0.0249)	(0.00835)	(0.114)	(0.0363)	(0.0416)	(0.0121)
Basic Absolute (BA)	0.165	-0.0123	-0.0413	-0.0289	0.0954	-0.0117	-0.0401	-0.0284
	(0.234)	(0.0355)	(0.0521)	(0.0186)	(0.155)	(0.0304)	(0.0453)	(0.0178)
High Absolute (HA)	-0.0411	0.0195	0.0290	0.00946	0.176	-0.0174	-0.0177	-0.000215
	(0.168)	(0.0247)	(0.0409)	(0.0191)	(0.126)	(0.0227)	(0.0373)	(0.0182)
Basic Relative (BR)	0.126	-0.0231	-0.0404	-0.0172	0.0193	0.00793	-0.0160	-0.0239
	(0.173)	(0.0318)	(0.0491)	(0.0202)	(0.145)	(0.0372)	(0.0512)	(0.0187)
Post*BA	-0.154	-0.00809	-0.00205	0.00604	0.0398	0.00229	0.00103	-0.00126
	(0.153)	(0.0258)	(0.0346)	(0.0149)	(0.176)	(0.0430)	(0.0551)	(0.0194)
Post*HA	0.191**	-0.0465	-0.0953**	-0.0488**	0.253*	-0.0557	-0.119**	-0.0630***
	(0.0899)	(0.0301)	(0.0444)	(0.0189)	(0.130)	(0.0405)	(0.0502)	(0.0175)
Post*BR	0.107	0.00682	-0.00385	-0.0107	0.225	-0.0439	-0.0530	-0.00906
	(0.0842)	(0.0239)	(0.0300)	(0.0136)	(0.141)	(0.0437)	(0.0518)	(0.0156)
Constant	13.49***	0.331***	0.416***	0.0848***	8.688***	0.0624	-0.0232	-0.0856
	(0.127)	(0.0172)	(0.0301)	(0.0142)	(0.341)	(0.0695)	(0.110)	(0.0694)
p-value Post*BA = Post*HA	0.0335	0.1617	0.0414	0.0134	0.1567	0.0611	0.011	0.0043
p-value Post*BA = Post*BR	0.0956	0.4513	0.9515	0.3124	0.2751	0.1918	0.273	0.6883
Control Variables					x	x	x	x
Observations	9376	9376	9376	9376	5342	5342	5342	5342

Notes: The control variables used in this regression are household demographics (age of child, gender of child, total number of siblings), household assets (monthly income, food expenditure, number of rooms, proportion of goods owned in the kitchen, proportion of non-kitchen goods), parent-specific controls (mother's age, mother's religion, whether the mother is a housewife, literacy of mother and father, whether the mother scored a high or low quiz score), worker-specific controls (whether the worker is experienced or not, whether the worker is highly educated or not, whether the worker scored a high or low quiz score), and center-specific controls (dummy variables for the center's facilities: electricity, fan, helper, chart, blackboard, drinking water, and toilet).
*** p<0.01, ** p<0.05, * p<0.1.

Table 5: Diet Results with Control Variables

	(1)	(2)	(3)	(4)	(5)
	Pulses	Fish	Meat	Green	Sweet
Post	-0.0523	0.0271	-0.0308	-0.0466	-0.123**
	(0.0435)	(0.0470)	(0.0499)	(0.0362)	(0.0516)
Basic Absolute (BA)	0.0617	0.122	0.0330	0.0900	-0.0252
	(0.0742)	(0.0882)	(0.0767)	(0.0700)	(0.0608)
High Absolute (HA)	-0.117	0.0373	-0.0446	-0.0309	-0.0874
	(0.0931)	(0.0872)	(0.0815)	(0.0906)	(0.0648)
Basic Relative (BR)	0.0752	0.133	0.0183	0.0291	0.00818
	(0.0735)	(0.105)	(0.0701)	(0.0730)	(0.0793)
Post*BA	0.0797	-0.00203	0.101	0.0189	0.0607
	(0.0964)	(0.100)	(0.0989)	(0.0914)	(0.100)
Post*HA	0.217**	0.0169	0.127*	0.000472	-0.00706
	(0.0803)	(0.106)	(0.0716)	(0.0876)	(0.0706)
Post*BR	0.109	-0.102	0.0221	0.0342	-0.0183
	(0.0689)	(0.0790)	(0.0695)	(0.0583)	(0.111)
Constant	0.235	0.177	-0.156	0.345**	0.130
	(0.179)	(0.166)	(0.164)	(0.152)	(0.0972)
Control Variables	x	x	x	x	x
p-value Post*BA = Post*HA	0.2246	0.886	0.8039	0.8782	0.5036
p-value Post*BA = Post*BR	0.7784	0.3688	0.433	0.8791	0.5519
Observations	5421	5430	5446	5441	5328

Notes: The control variables used in this regression are household demographics (age of child, gender of child, total number of siblings), household assets (monthly income, food expenditure, number of rooms, proportion of goods owned in the kitchen, proportion of non-kitchen goods), parent-specific controls (mother's age, mother's religion, whether the mother is a housewife, literacy of mother and father, whether the mother scored a high or low quiz score), worker-specific controls (whether the worker is experienced or not, whether the worker is highly educated or not, whether the worker scored a high or low quiz score), and center-specific controls (dummy variables for the center's facilities: electricity, fan, helper, chart, blackboard, drinking water, and toilet).
*** p<0.01, ** p<0.05, * p<0.1.

Table 6: Religious Heterogeneity after including Control Variables

	Panel A: Same religion				Panel B: Different Religion			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Weight	Malnourished	Grade	Severe	Weight	Malnourished	Grade	Severe
Post	0.348***	-0.00674	0.00703	0.0138	0.0987	0.0724	0.0819	0.00953
	(0.122)	(0.0322)	(0.0321)	(0.0151)	(0.125)	(0.0478)	(0.0540)	(0.0201)
Basic Absolute (BA)	0.0564	-0.00202	-0.00277	-0.000743	0.225	-0.0480	-0.114	-0.0663**
	(0.168)	(0.0292)	(0.0427)	(0.0192)	(0.223)	(0.0476)	(0.0749)	(0.0291)
High Absolute (HA)	0.183	-0.00306	0.0259	0.0290	0.0991	-0.0244	-0.0534	-0.0290
	(0.160)	(0.0306)	(0.0606)	(0.0362)	(0.203)	(0.0469)	(0.0824)	(0.0370)
Basic Relative (BR)	-0.0818	0.0547	0.0721	0.0175	0.165	-0.0399	-0.107	-0.0668**
	(0.162)	(0.0444)	(0.0570)	(0.0191)	(0.179)	(0.0455)	(0.0721)	(0.0291)
Post*BA	0.233	-0.0335	-0.0561	-0.0226	-0.0475	0.0157	0.0330	0.0172
	(0.154)	(0.0354)	(0.0404)	(0.0213)	(0.228)	(0.0568)	(0.0672)	(0.0236)
Post*HA	0.544***	-0.0659	-0.151*	-0.0847*	0.182	-0.0694	-0.122	-0.0523**
	(0.180)	(0.0498)	(0.0823)	(0.0420)	(0.185)	(0.0646)	(0.0786)	(0.0247)
Post*BR	0.290**	-0.0339	-0.0429	-0.00902	0.212	-0.0692	-0.0762	-0.00692
	(0.139)	(0.0415)	(0.0433)	(0.0178)	(0.168)	(0.0571)	(0.0706)	(0.0244)
Constant	8.009***	0.221**	0.143	-0.0778	9.316***	-0.203	-0.333*	-0.130
	(0.546)	(0.103)	(0.148)	(0.0806)	(0.490)	(0.127)	(0.188)	(0.0900)
Control Variables	x	x	x	x	x	x	x	x
Observations	2565	2565	2565	2565	2777	2777	2777	2777

Notes: The sample in Panel A is restricted to pairs of mothers and workers who are of the same religion (hindu/hindu or muslim/muslim). The sample in Panel B is restricted to pairs of mothers and workers who are of different religions(hindu/muslim or muslim/hindu). The control variables used in this regression are household demographics (age of child, gender of child, total number of siblings), household assets (monthly income, food expenditure, number of rooms, proportion of goods owned in the kitchen, proportion of non-kitchen goods), parent-specific controls (mother's age, mother's religion, whether the mother is a housewife, literacy of mother and father, whether the mother scored a high or low quiz score), worker-specific controls (whether the worker is experienced or not, whether the worker is highly educated or not, whether the worker scored a high or low quiz score), and center-specific controls (dummy variables for the center's facilities: electricity, fan, helper, chart, blackboard, drinking water, and toilet).

Table 7: Gender Difference in average treatment effects

	(1)	(2)	(3)	(4)
	Weight	Malnourished	Grade	Severe Malnourished
Post*BA	0.142	-0.0226	-0.0339	-0.0113
	(0.216)	(0.0428)	(0.0712)	(0.0391)
Post*HA	0.432***	-0.0989**	-0.191***	-0.0924**
	(0.145)	(0.0397)	(0.0620)	(0.0350)
Post*BR	0.219	-0.0548	-0.0682	-0.0133
	(0.164)	(0.0371)	(0.0548)	(0.0307)
Gender*BA*Post	-0.197	0.0475	0.0678	0.0203
	(0.163)	(0.0469)	(0.0887)	(0.0548)
Gender*HA*Post	-0.345*	0.0830*	0.140*	0.0567
	(0.193)	(0.0483)	(0.0780)	(0.0435)
Gender*BR*Post	0.0188	0.0202	0.0283	0.00811
	(0.176)	(0.0426)	(0.0726)	(0.0424)
Constant	8.784***	0.0442	-0.0635	-0.108
	(0.349)	(0.0699)	(0.109)	(0.0708)
Control Variables	x	x	x	x
Observations	5342	5342	5342	5342

Notes: Gender = 1 if girl and 0 if boy. The control variables used in this regression are household demographics (age of child, gender of child, total number of siblings), household assets (monthly income, food expenditure, number of rooms, proportion of goods owned in the kitchen, proportion of non-kitchen goods), parent-specific controls (mother's age, mother's religion, whether the mother is a housewife, literacy of mother and father, whether the mother scored a high or low quiz score), worker-specific controls (whether the worker is experienced or not, whether the worker is highly educated or not, whether the worker scored a high or low quiz score), and center-specific controls (dummy variables for the center's facilities: electricity, fan, helper, chart, blackboard, drinking water, and toilet).
*** p<0.01, ** p<0.05, * p<0.1.

Table 8: Selective targeting test

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Weight	Malnourished	Grade	Severe Malnourished	Weight	Malnourished	Grade	Severe Malnourished
Post*BA	-0.239	0.0630	0.0812	0.0182	-0.0206	0.0526	0.0892	0.0366
	(0.152)	(0.0546)	(0.0636)	(0.0446)	(0.158)	(0.0619)	(0.0799)	(0.0463)
Post*HA	0.289	-0.0227	-0.142	-0.119**	0.469**	-0.109*	-0.301***	-0.193***
	(0.181)	(0.0475)	(0.0886)	(0.0532)	(0.181)	(0.0623)	(0.0778)	(0.0410)
Post*BR	0.114	0.0526	0.0222	-0.0304	0.0483	-0.00445	-0.0204	-0.0159
	(0.192)	(0.0551)	(0.0834)	(0.0460)	(0.194)	(0.0760)	(0.102)	(0.0359)
Close To Target*BA*Post	0.213	-0.138	-0.120	0.0182	0.0700	-0.0862	-0.116	-0.0293
	(0.249)	(0.0950)	(0.114)	(0.0351)	(0.216)	(0.0967)	(0.110)	(0.0533)
Close To Target*HA*Post	0.263	-0.0845	-0.0830	0.00152	0.239	0.0207	0.0582	0.0375
	(0.198)	(0.0929)	(0.112)	(0.0276)	(0.212)	(0.0915)	(0.0914)	(0.0447)
Close To Target*BR*Post	0.0341	-0.0857	-0.101	-0.0150	0.0265	-0.0522	-0.0670	-0.0148
	(0.267)	(0.106)	(0.146)	(0.0523)	(0.299)	(0.126)	(0.162)	(0.0670)
Constant	11.67***	1.000***	1.227***	0.227***	5.987***	1.162***	1.104***	-0.0584
	(0.121)	(9.99e-10)	(0.0386)	(0.0386)	(0.415)	(0.0835)	(0.210)	(0.191)
Control Variables					x	x	x	x
Observations	3090	3089	3089	3089	1739	1739	1739	1739

Notes: The control variables used in columns (5)-(8) are household demographics (age of child, gender of child, total number of siblings), household assets (monthly income, food expenditure, number of rooms, proportion of goods owned in the kitchen, proportion of non-kitchen goods), parent-specific controls (mother's age, mother's religion, whether the mother is a housewife, literacy of mother and father, whether the mother scored a high or low quiz score), worker-specific controls (whether the worker is experienced or not, whether the worker is highly educated or not, whether the worker scored a high or low quiz score), and center-specific controls (dummy variables for the center's facilities: electricity, fan, helper, chart, blackboard, drinking water, and toilet). Other variables in this regression include treatments on their own, close to target on its own, post on its own and pair-wise interactions of close to target with post. The variable closetotarget is defined to be 1 if the child is malnourished at the baseline and its deficit weight (difference between the target weight and the actual weight) is less than the mean deficit weight and 0 if the child is malnourished at the baseline, but his or her deficit weight is more than or equal to the mean deficit weight. Target weight is defined as the threshold at which the child's grade will decrease by 1 (i.e. would go from severely malnourished to only moderately malnourished or from moderately malnourished to not malnourished).
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

MAHESHTALA BLOCK: TREATMENT AND CONTROL CLUSTERS

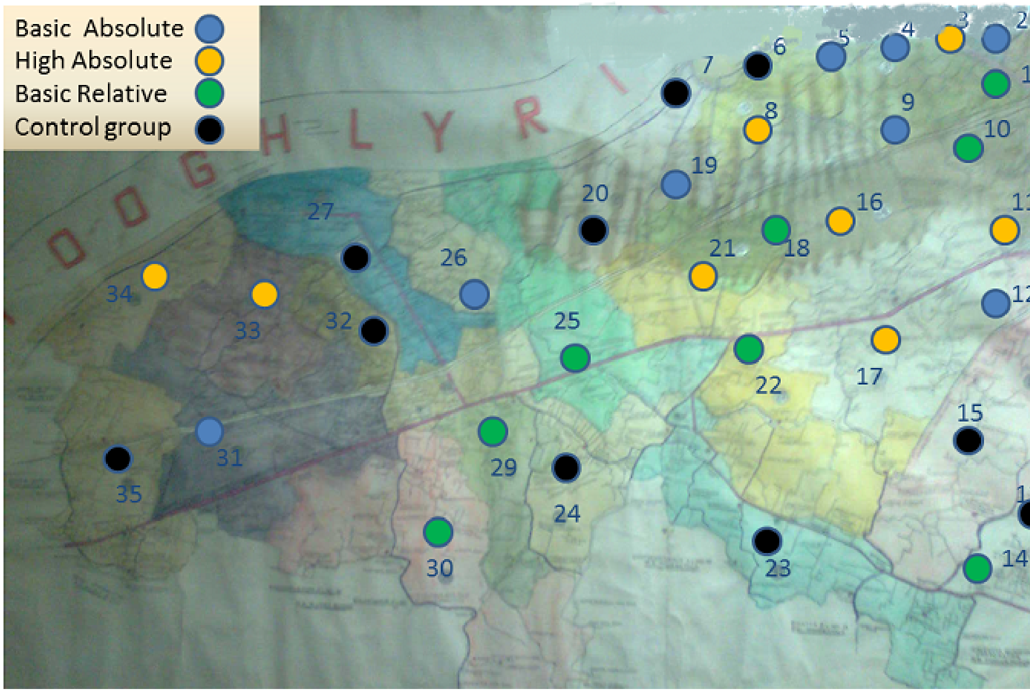


Figure 1: Map of Maheshtala Block and Clusters in the experiment

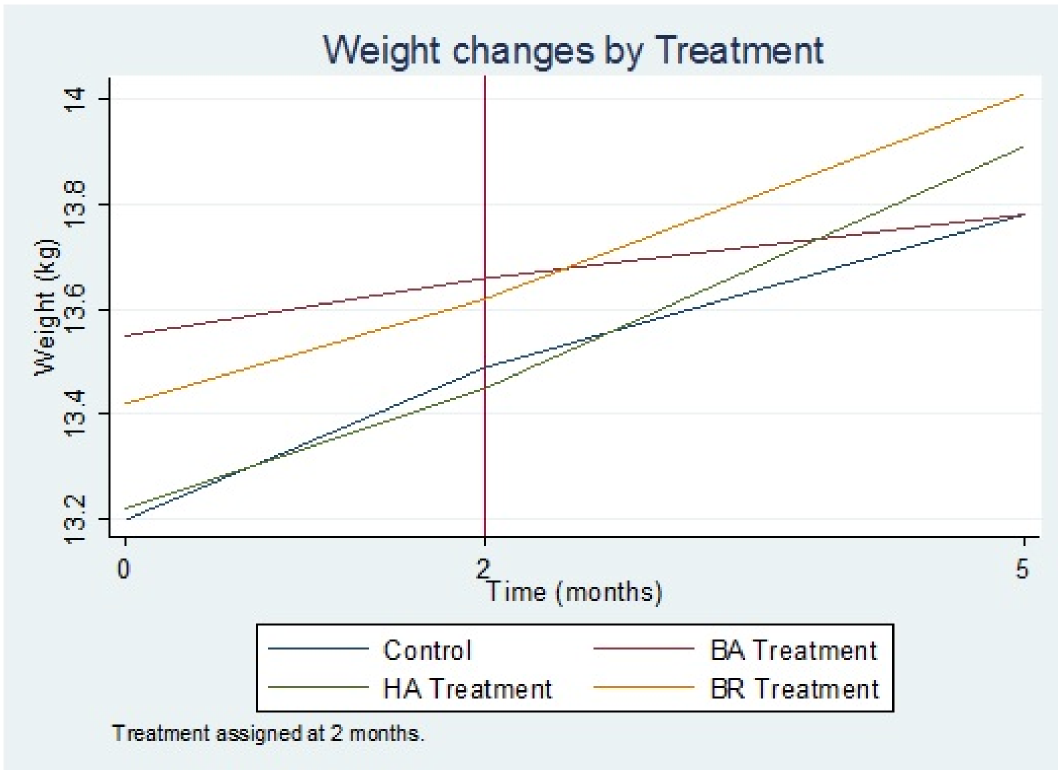


Figure 2: Weight changes by Treatment

10 Appendix

10.1 Tables

Table A1: Placebo test with and without Control Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Weight	Malnourished	Grade	Severe	Weight	Malnourished	Grade	Severe
Post	0.289**	0.0365	0.0703**	0.0338***	0.157**	0.0590**	0.0957***	0.0367***
	(0.111)	(0.0250)	(0.0302)	(0.00845)	(0.0621)	(0.0256)	(0.0312)	(0.0110)
Basic Absolute (BA)	0.353	-0.00926	-0.0223	-0.0131	0.0694	0.0339	0.0262	-0.00778
	(0.267)	(0.0432)	(0.0607)	(0.0191)	(0.181)	(0.0457)	(0.0596)	(0.0157)
High Absolute (HA)	0.0177	0.0225	0.0570	0.0346*	0.0253	0.0433	0.0662	0.0229
	(0.191)	(0.0402)	(0.0565)	(0.0185)	(0.153)	(0.0347)	(0.0516)	(0.0199)
Basic Relative (BR)	0.210	-0.0131	-0.0198	-0.00664	-0.161	0.0526	0.0612	0.00855
	(0.206)	(0.0354)	(0.0518)	(0.0199)	(0.161)	(0.0410)	(0.0532)	(0.0179)
Post*BA	-0.188	-0.00309	-0.0189	-0.0159	-0.0512	-0.0318	-0.0476	-0.0158
	(0.167)	(0.0298)	(0.0364)	(0.0100)	(0.0897)	(0.0294)	(0.0362)	(0.0137)
Post*HA	-0.0588	-0.00296	-0.0281	-0.0251	0.0466	-0.0396	-0.0580	-0.0184
	(0.164)	(0.0317)	(0.0440)	(0.0188)	(0.113)	(0.0300)	(0.0454)	(0.0233)
Post*BR	-0.0843	-0.00997	-0.0206	-0.0106	0.134	-0.0367	-0.0633	-0.0266
	(0.184)	(0.0356)	(0.0449)	(0.0138)	(0.111)	(0.0364)	(0.0437)	(0.0164)
Constant	13.20***	0.295***	0.346***	0.0510***	8.615***	-0.0329	-0.137	-0.104*
	(0.110)	(0.0280)	(0.0419)	(0.0151)	(0.417)	(0.109)	(0.124)	(0.0605)
N	9789	9789	9789	9789	5397	5397	5397	5397

Notes: The control variables used in this regression are household demographics (age of child, gender of child, total number of siblings), household assets (monthly income, food expenditure, number of rooms, proportion of goods owned in the kitchen, proportion of non-kitchen goods), parent-specific controls (mother's age, mother's religion, whether the mother is a housewife, literacy of mother and father, whether the mother scored a high or low quiz score), worker-specific controls (whether the worker is experienced or not, whether the worker is highly educated or not, whether the worker scored a high or low quiz score), and center-specific controls (dummy variables for the center's facilities: electricity, fan, helper, chart, blackboard, drinking water, and toilet).

Table A2: Attrition Table

	(1)	(2)	(3)	(4)
	Attrition	Attrition	Attrition	Attrition
Weight*BA	0.0190			
	(0.0155)			
Weight*HA	-0.00657			
	(0.0152)			
Weight*BR	-0.00680			
	(0.0169)			
Malnourished*BA		-0.0367		
		(0.0656)		
Malnourished*HA		0.0314		
		(0.0532)		
Malnourished*BR		0.0114		
		(0.0695)		
Grade*BA			-0.0301	
			(0.0555)	
Grade*HA			0.0114	
			(0.0460)	
Grade*BR			-0.00393	
			(0.0603)	
SevereMalnourished*BA				-0.0204
				(0.104)
SevereMalnourished*HA				-0.0297
				(0.101)
SevereMalnourished*BR				-0.0403
				(0.127)
Control Variables	x	x	x	x
Observations	3013	3013	3013	3013

Notes: The control variables used in this regression are household demographics (age of child, gender of child, total number of siblings), household assets (monthly income, food expenditure, number of rooms, proportion of goods owned in the kitchen, proportion of non-kitchen goods), parent-specific controls (mother's age, mother's religion, whether the mother is a housewife, literacy of mother and father, whether the mother scored a high or low quiz score), worker-specific controls (whether the worker is experienced or not, whether the worker is highly educated or not, whether the worker scored a high or low quiz score), and center-specific controls (dummy variables for the center's facilities: electricity, fan, helper, chart, blackboard, drinking water, and toilet). The attrition variable takes value 1 if the child's weight is unavailable in the second round, and 0 otherwise.
*** p<0.01, ** p<0.05, * p<0.1.

Table A3: Main Results with Control Variables and Moulton Correction

	(1)	(2)	(3)	(4)
	Weight	Malnourished	Grade	Severe Malnourished
Post	0.271**	0.0234	0.0317	0.00830
	(0.111)	(0.0270)	(0.0360)	(0.0153)
Basic Absolute (BA)	0.0954	-0.0117	-0.0401	-0.0284
	(0.197)	(0.0335)	(0.0500)	(0.0205)
High Absolute (HA)	0.176	-0.0174	-0.0177	-0.000215
	(0.193)	(0.0329)	(0.0491)	(0.0201)
Basic Relative (BR)	0.0193	0.00793	-0.0160	-0.0239
	(0.200)	(0.0341)	(0.0509)	(0.0208)
Post*BA	0.0398	0.00229	0.00103	-0.00126
	(0.206)	(0.0407)	(0.0576)	(0.0240)
Post*HA	0.253	-0.0557	-0.119**	-0.0630***
	(0.210)	(0.0408)	(0.0581)	(0.0242)
Post*BR	0.225	-0.0439	-0.0530	-0.00906
	(0.218)	(0.0428)	(0.0607)	(0.0253)
Constant	8.688***	0.0624	-0.0232	-0.0856
	(0.598)	(0.102)	(0.152)	(0.0622)
Control Variables	x	x	x	x
Observations	5342	5342	5342	5342

Notes: The control variables used in this regression are household demographics (age of child, gender of child, total number of siblings), household assets (monthly income, food expenditure, number of rooms, proportion of goods owned in the kitchen, proportion of non-kitchen goods), parent-specific controls (mother's age, mother's religion, whether the mother is a housewife, literacy of mother and father, whether the mother scored a high or low quiz score), worker-specific controls (whether the worker is experienced or not, whether the worker is highly educated or not, whether the worker scored a high or low quiz score), and center-specific controls (dummy variables for the center's facilities: electricity, fan, helper, chart, blackboard, drinking water, and toilet). Moulton correction for the standard errors is used in this table.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A4: Robustness Check for Main Results after excluding All Household Assets from Controls

	(1)	(2)	(3)	(4)
	Weight	Malnourished	Grade	Severe Malnourished
Post	0.281***	0.0131	0.0114	-0.00168
	(0.0792)	(0.0227)	(0.0265)	(0.00948)
Basic Absolute (BA)	0.281*	-0.0498	-0.0915**	-0.0417***
	(0.164)	(0.0317)	(0.0441)	(0.0152)
High Absolute (HA)	0.198	-0.0299	-0.0376	-0.00769
	(0.122)	(0.0219)	(0.0345)	(0.0162)
Basic Relative (BR)	0.155	-0.0323	-0.0608	-0.0286*
	(0.123)	(0.0275)	(0.0379)	(0.0150)
Post*BA	-0.00355	0.00528	0.0169	0.0116
	(0.146)	(0.0312)	(0.0413)	(0.0153)
Post*HA	0.263***	-0.0569**	-0.109***	-0.0518***
	(0.0844)	(0.0252)	(0.0342)	(0.0156)
Post*BR	0.162	-0.0277	-0.0268	0.000865
	(0.114)	(0.0291)	(0.0385)	(0.0140)
Constant	8.840***	0.0519	-0.0410	-0.0929*
	(0.293)	(0.0668)	(0.0944)	(0.0537)
Control Variables	x	x	x	x
p-value Post*BA = Post*HA	0.0567	0.0266	0.0039	0.0019
p-value Post*BA = Post*BR	0.2945	0.284	0.3301	0.5422
Observations	6564	6564	6564	6564

Notes: The control variables used in this regression are household demographics (age of child, gender of child, total number of siblings), parent-specific controls (mother's age, mother's religion, whether the mother is a housewife or not, literacy of mother and father, whether the mother scored a high or low quiz score), worker-specific controls (whether the worker is experienced or not, whether the worker is highly educated or not, whether the worker scored a high or low quiz score), and center-specific controls (dummy variables for the center's facilities: electricity, fan, helper, chart, blackboard, drinking water, and toilet).
*** p<0.01, ** p<0.05, * p<0.1.

10.2 Theory

In the absolute performance pay case:

$$\begin{aligned} E(U_A(2)) - E(U_A(0)) &= \theta V(a) + (1-\theta)V(b) - 2c - V(a) \\ &= (1-\theta)(V(b) - V(a)) - 2c > 0 \end{aligned}$$

i.e.

$$V(b) - V(a) > \frac{2c}{1-\theta}$$

$$\begin{aligned} E(U_A(2)) - E(U_A(1)) &= \theta V(a) + (1-\theta)V(b) - 2c - \frac{1+\theta}{2}V(a) - \frac{1-\theta}{2}V(b) + \frac{1}{2}c \\ &= \frac{1-\theta}{2}(V(b) - V(a)) - \frac{3}{2}c > 0 \end{aligned}$$

i.e.

$$V(b) - V(a) > \frac{3c}{1-\theta}$$

Notice that this condition dominates the first one.

In the relative performance pay case:

To make 2 a best response, it is necessary that:

$$\begin{aligned} E(U_A(2)) - E(U_A(0)) &= V(T) - 2c - \theta V(T) - (1-\theta)V(L_1) \\ &= (1-\theta)(V(T) - V(L_1)) - 2c > 0 \end{aligned}$$

which implies that

$$V(T) - V(L_1) > \frac{2c}{1-\theta}$$

$$\begin{aligned} E(U_A(2)) - E(U_A(1)) &= V(T) - 2c - \frac{1+\theta}{2}V(T) - \frac{1-\theta}{2}V(L_1) + \frac{1}{2}c \\ &= \frac{1-\theta}{2}(V(T) - V(L_1)) - \frac{3}{2}c > 0 \end{aligned}$$

which implies that V

$$(T) - V(L_1) > \frac{3c}{1-\theta}(*)$$

Notice this condition dominates the first one.

Similarly, we assume that in the worst case scenario, the agent should also attain the same minimum utility level U. This means that:

$$\theta V(T) + (1-\theta)V(L_1) \geqslant U \quad (**)$$

The cost is minimized when $(**)$ holds, i.e., when $V(L_1) = \frac{U-\theta V(T)}{1-\theta}$. Plug it in $(*)$, it follows that:

$$V(T) - \frac{U-\theta V(T)}{1-\theta} \geqslant \frac{3c}{1-\theta},$$

which implies that

$$(1-\theta)V(T) - U + \theta V(T) \geqslant 3c.$$

$$\begin{aligned} C_{absolute} - C_{relative} &= n \times (\theta V^{-1}(U) + (1-\theta)V^{-1}(\frac{3c}{1-\theta} + U)) - n \times V^{-1}(U+3c) \\ &= n(\theta V^{-1}(U) + (1-\theta)V^{-1}(\frac{3c}{1-\theta} + U) - V^{-1}(U+3c)) \end{aligned}$$

Notice the difference is actually 0 when $\theta = 0$. To see how this difference varies with θ, we take its derivative about θ. For convenience, suppose $V^{-1}(x) = h(x)$. Since $V(x)$ is concave and increasing, its inverse function $h(x)$ is convex and increasing.

$$\begin{aligned} (C_{absolute} - C_{relative})' &= n(h(U) - h(\frac{3c}{1-\theta} + U) + (1-\theta)h'(\frac{3c}{1-\theta} + U) \times \frac{3c}{(1-\theta)^2}) \\ &= n(h(U) - h(\frac{3c}{1-\theta} + U) + h'(\frac{3c}{1-\theta} + U) \times \frac{3c}{(1-\theta)}) \end{aligned}$$

Since $h(x)$ is a continuous function on its domain, by Mean Value Theorem, it follows that

$$h(\frac{3c}{1-\theta}+U)-h(U)=(\frac{3c}{1-\theta}+U-U)h'(U^*)=\frac{3c}{1-\theta}h'(U^*)$$

where

$$U^* \in (U, \frac{3c}{1-\theta}+U)$$

Plug it back in, it follows that:

$$\begin{aligned}(C_{absolute}-C_{relative})' &= h'(\frac{3c}{1-\theta}+U)\times\frac{3c}{(1-\theta)}-\frac{3c}{1-\theta}h'(U^*)\\ &= \frac{3c}{1-\theta}(h'(\frac{3c}{1-\theta}+U)-h'(U^*))\end{aligned}$$

Notice that

$$\frac{3c}{1-\theta}+U>U^*$$

Since $h(x)$ is convex, it follows that

$$h'(\frac{3c}{1-\theta}+U)>h'(U^*)$$

and therefore

$$(C_{absolute}-C_{relative})'>0$$

Now that $C_{absolute}-C_{relative}=0$ for $\theta=0$, then $C_{absolute}>C_{relative}$ for $\theta>0$. To put it another way, the cost for an absolute scheme is higher than that for a relative scheme whenever common shock is present. This result is similar but not the same as that in the literature, which states that the absolute scheme should dominate the relative schemes when the common shock level is low, or at least when common shock is not present. This result is caused by the "public sector assumption", which assumes that agents should obtain a minimum level of utility even when they perform poorly.

Do Health Care Providers Respond to Demand-Side Incentives? Evidence from Indonesia

Margaret Triyana*

November 2013

Abstract

This paper exploits the sub-district randomization of Indonesia's household Conditional Cash Transfer (CCT) program to analyze how the program affects the local health care market. The CCT program is associated with increased use of midwives, who are the main delivery attendants. The program is associated with a 10% increase in the number of midwives and a 10% increase in delivery fees charged by midwives in treated communities. Program participants experience higher quality of prenatal care, but this is driven by increased utilization among participants, instead of improvements in the quality of care provided by midwives.

JEL codes: I1, I3, O1

*Email: mtriyana@stanford.edu. I am especially grateful to Robert Lalonde, Dan Black, Kerwin Charles, Ioana Marinescu, and Rema Hanna for their guidance and numerous helpful discussions. I also thank Ben Olken, Alicia Menendez, Marianne Bertrand, Harold Alderman, Jeff Grogger, Lily Hoo, Yulia Herawati, Lina Marliani, Julia Tobias, Matthew Wai-poi, Jessica Pan, Alice Chen, Yi-Lin Tsai, Maria Rosales, and seminar participants at the University of Chicago and Harvard Kennedy School for valuable suggestions. Research support from the Hewlett Foundation/IIE Dissertation Fellowship and the Harvard Kennedy School Indonesia Program is gratefully acknowledged.

1 Introduction

Developing countries have used different strategies to improve health outcomes in low-resource settings, especially among the poor. Price is often cited as a barrier to health care access among the poor (Whitehead et al., 2001); therefore, household Conditional Cash Transfer (CCT) programs have been implemented in many developing countries as an anti-poverty strategy that seeks to increase human capital investments. CCT programs provide cash transfers to poor households on the condition that participating households meet the specified health and educational investments on their children. Although most CCT programs have been shown to improve health-seeking behavior, the program effects on health outcomes are mixed (Fizbein et al., 2009). Similarly, Indonesia's CCT program has been shown to improve its targeted prenatal indicators, but there are no significant effects on birth outcomes (Alatas et al., 2011). One potential explanation for this puzzle is the supply-side response to CCT programs. This paper adds to the impact evaluation and existing literature by estimating the program effects on the local health care market. In particular, I analyze whether the demand shock from the program is associated with changes in health care availability, price, and the quality of care.

The household CCT program is a demand-side intervention to ensure that expectant mothers and children from poor households receive basic health care services. Thus, the program's health requirements induce a demand shock in the health care market. Although CCT programs have no significant general equilibrium effects as measured by local prices and wages (Fizbein et al., 2009; Angelucci and De Giorgi, 2009), there may be price effects in the health care market because the program specifically targets investments in health and most health care services have an out-of-pocket component in many countries, including Indonesia. In the short run, when supply is fixed, health care price should increase. In turn, the price increase may dampen program effects because higher prices limit the affordability of health care. Although price increases alone would lower utilization (Kremer and Miguel, 2007; Cohen et al., 2010), participating households are required to use health care services in order to receive the cash transfer, even if price increased. In anticipation of the possible price increase, CCT programs are implemented in supply-ready communities. These communities have sufficient health care providers (and schools) so that the local health care (and educational) system is able to meet the additional demand without straining the existing capacity. In the long-run, supply can also increase to further mitigate the short-run price increases. Despite the supply response, the demand shock in the health care market would still be weakly associated with a price increase, even though only poor households are directly affected by the program. To analyze the supply response, this paper estimates changes in the availability of health care providers and price changes associated with the CCT program.

The general equilibrium effect on quality of care and health outcomes is ambiguous. On the one hand, increased demand may lower health care quality as providers see more patients, and obtaining low quality of care would do little to improve health outcomes. On the other hand, quality may increase because CCT programs educate program participants on maternal and child health, including prenatal care. Since prenatal care is an input in the production of birth outcomes, with this intervention, women should have better knowledge and seek higher quality prenatal care, which should translate into better birth outcomes (Barber and Gertler, 2010). However, better prenatal care also improves the detection of prenatal complications, which could subsequently lead to worse surviving birth outcomes (Alexander and Korenbrot, 1995). Because birth outcomes are affected by multiple factors in addition to prenatal care, this paper focuses on prenatal care quality, which is more readily observed. Program participants often do not obtain prenatal care, but are now required to do so from a trained health care provider, so the program should improve the quality of care received in the target population. To explore whether quality changes among participants are reflected in the local health care market, I compare prenatal quality changes experienced by participating households to changes in the self-reported prenatal quality provided by providers.

Following the impact evaluation, this paper exploits the sub-district randomization of Indonesia's CCT pilot program to estimate the intent to treat and treatment on the treated parameters. On the supply side, I focus on midwives since they are the main health care providers and birth attendants in Indonesia[1]. For households,

[1] 62% of births are attended by midwives (Source: Indonesia: Demographic and Health Surveys 2007).

the program effects are measured using a sample of near poor and poor households, which is the population most likely to be affected by the program. Households in treated sub-districts are enrolled in the program based on a proxy means test. Since household participation is endogenous, the treatment on the treated effect is estimated using instrumental variable, with the sub-district randomization as instrument for household participation. To analyze changes in health care utilization, this paper estimates changes in the use of midwives for delivery assistance. Since delivery assistance is also the most costly health requirement, the local general equilibrium effect is measured by changes in the delivery fees charged by midwives. Quality of care is captured by the quality of prenatal care received by patients and the prenatal quality provided by midwives.

Among near poor and poor households that live in treated sub-districts, the CCT program is associated with a 15% increase in the use of midwives for delivery assistance, and a 25% increase in delivery fees paid to midwives. For households that receive the cash transfers, program participation is associated with a 45% increase in the use of midwives for childbirth. Consequently, participating households experience an increase in delivery fees paid to midwives, which amounts to 30% of the cash transfer. In terms of the general equilibrium effects in the local health care market, the demand shock from the CCT program is associated with a 10% increase in the number of midwives. In spite of the supply response, the program is associated with a 10% increase in the delivery fees charged by midwives. Participating households experience a 0.15 standard deviation increase in prenatal care quality received, but midwives report no change in the quality of care provided. The quality improvement is therefore driven by increased utilization in the target population, instead of quality improvements in the local health care market. These results suggest the importance of provider availability and quality of care to ultimately improve health outcomes in CCT programs.

The remainder of the paper is organized as follows. Section 2 presents the institutional background and the pilot program in Indonesia. Section 3 describes the data and estimation strategy. Section 4 presents the results. Section 5 provides a brief discussion of the results in relation to previous findings in the literature, and section 6 concludes.

2 Background

Indonesia's health workforce is made up of nurses or paramedics, midwives, and doctors. Doctors and midwives are trained delivery attendants, while nurses and paramedics are only allowed to assist midwives and doctors in deliveries. Although midwives are the primary birth attendant in Indonesia, 36% of births are attended by traditional birth attendants, who receive no medical training and are not trained to identify or manage delivery complications[2]. The use of traditional attendants is associated with a higher probability of infant death in the first month and 11 months[3]. In spite of the low quality of traditional attendants, they continue to assist childbirth because they charge much lower fees relative to midwives. In addition, women sometimes choose to use traditional birth attendants accompanied by a trained attendant because of family or cultural tradition. Given these serious concerns regarding traditional attendants, the main objectives of the CCT program include reducing maternal and infant mortality by requiring participants to obtain prenatal care and use doctors or midwives for childbirth assistance.

In the public health care system, each sub-district has at least one community health center, headed by a doctor, and staffed by several nurses and midwives, who are salaried and receive a fee-for-service payment for each service rendered[4]. The majority of midwives are employed by the government, and such midwives live in the village in the catchment area to serve one or several villages. In addition to prenatal care and delivery assistance, village midwives provide contraceptives, vaccinations, curative examinations, and provide medications. Village midwives are supervised by the sub-district clinic and submit a report of their activities to the clinic monthly.

Government midwives are allowed to hold private practice outside of their public hours to supplement their income (Heywood and Harahap, 2009). 90% of midwives have their own private practice, and about 60% of their income comes from private practice (Ensor et al., 2009). Dual practice, which is private practice undertaken by health care workers employed in the public sector, has increased the supply of health care services in Indonesia, but there is limited oversight, and there are concerns that

[2]Source: Indonesia: Demographic and Health Surveys 2007

[3]Source: Indonesia: Demographic and Health Surveys 2007

[4]http://www.searo.who.int/en/Section313/Section1520_6822.htm

dual practice reduces the incentive to deliver services to the poor (Barber et al., 2007). Starting from 2005, *Askeskin*, the government's health insurance scheme for the poor, reimburses midwives for services rendered to low-income households. More recently, two types of health interventions were launched to further improve health care access for the poor.

The household CCT program, *Program Keluarga Harapan*[5] (PKH), was piloted along with a community CCT program, *PNPM Generasi*[6] in 2007. The government used geographic targeting to pilot the two programs in 5 provinces: West Java and East Java on the main island of Java, and the following off-Java provinces: North Sulawesi, Gorontalo, and East Nusa Tenggara (NTT). The household CCT program also includes sub-districts in Jakarta, the capital city. The pilot program ensured that there was no overlap between the two programs. Randomization was done at the sub-district level because many facilities, including secondary schools and health centers, are provided at the sub-district level. In addition, the cluster design takes into account the possibility of local externalities resulting from the sub-district treatment (Miguel and Kremer, 2004; Olken, 2007). Both programs target the same maternal and child health indicators and educational indicators.

The household CCT program, PKH, was piloted in sub-districts that were considered supply-ready to ensure that local facilities would be able to meet the additional demand. The program sets a lower threshold for sub-districts outside of the main island of Java because health and education services are more limited off-Java. 588 sub-districts were identified for PKH pilot, and the sample was stratified by urban classification. 329 sub-districts were randomized into treatment and 259 sub-districts were in the control group. Within treated sub-districts, PKH targeted households classified as extremely poor by Statistics Indonesia (*Badan Pusat Statistik*, BPS). Statistics Indonesia used a proxy-means test to all poor households to identify program beneficiaries. Extremely poor households with expectant or lactating women, children under 5, and school-aged children are eligible for the household CCT program. Expectant women are required to obtain at least 4 prenatal visits, prenatal

[5] Hopeful Family Program

[6] *Program Nasional Pemberdayaan Masyarakat: Generasi Sehat dan Cerdas* (National Program for Community Empowerment: a Healthy and Bright Generation)

iron supplement, delivery assistance from a doctor or midwife, and 2 postnatal visits. Children under 5 are required to be weighed monthly, obtain complete vaccinations and vitamin A twice a year. School-aged children are required to be enrolled (in either primary or junior high school to attain the 9 years of mandatory education) and have 85% attendance rate.

Like other household CCT programs, PKH delivers a quarterly cash transfer to mothers, which is done through the nearest post office. The amount that each household receives depends on household composition, ranging from a minimum transfer of *Rp.* 600,000 (USD 60) to a maximum of *Rp.* 2,200,000 (USD 220). The transfer amounts to 15% to 20% of estimated total consumption of poor households. The total cost to meet all the program requirements amounts to 50% of the transfer[7]. Each household receives the transfer every quarter so long as they meet the program requirements. Verification for both programs is conducted by trained facilitators who collect monthly attendance sheets from schools in the villages, and patient and service lists from health care providers. The PKH district office checks household compliance before initiating the following payments. Non-compliant households will first receive a warning letter delivered by the facilitator. A second breach will result in a 10% loss of benefit, and a third breach will result in expulsion.

Dual practice affects providers' response to the demand shock associated with the CCT program. The CCT program allows households to choose between private and public practice for their health care services, and government health care providers are required to report both their public and private patients to the sub-district health center. In public practice, midwives are often required to follow pricing guidelines from the district's Health Department[8], so the program is expected to have no effect on public fees. However, in private practice, midwives are able to set their private fees because they can respond to demand shocks and also induce private demand. With dual practice, the program effect on private fees is theoretically ambiguous[9], and increased demand in public or private practice may have different effects on private

[7]Source: Author's calculations based on average household expenditure on prenatal care, delivery, postnatal care, vaccinations, and education at baseline.

[8]Source: Various district and province level health regulations (*Peraturan Daerah*).

[9]Theoretical framework for dual practice midwives is based on Bir and Eggleston (2003), available upon request.

price. If increased demand from program beneficiaries only increases public demand, this may result in non-participants moving to private practice, thereby increasing private demand. Since the program may affect both public and private demand, how dual practice providers set private fees in response to a demand shock is an empirical question that this paper addresses.

3 Data and Estimation

3.1 Data

The data comes from a series of surveys conducted for the impact evaluation of the household CCT program. Prices and quality are collected directly in the household survey and the midwife survey. Two waves of the survey were carried out in control and treated sub-districts as part of the evaluation series. The baseline round was conducted in 2007 prior to program implementation and a follow-up survey was conducted in 2009. The surveys include household, village, midwife, and facility surveys. The sample covers the sub-districts that were included in the initial randomization[10].

The household CCT impact evaluation survey follows a panel of 14,326 households in 2,723 villages. Households in the sample are poor or near poor households, because they are most likely to be affected by the CCT program: 95% of the households in the sample ever received an unconditional cash transfer (*Bantuan Langsung Tunai*, BLT), and 94% received subsidized rice (*Raskin*) (Alatas et al., 2011). 98% of households were re-interviewed at follow-up, with 13,602 married women and 5,616 children under the age of 3[11]. The household survey contains information on household size, education, age, household asset ownership, and consumption expenditure. The indicator for program participation is equal to one when households report re-

[10] About 10% of the initial control sub-districts managed to gain access to the program, they are included in the sample, and their treatment status is based on their initial randomization status.

[11] In every hamlet in the sample, a list of households with the following conditions was collected: (i) pregnant or lactating mothers or women who were pregnant in the last 24 months and (ii) households with children aged 6-15 years. Two households were randomly selected from group (i), and three households from group (ii) were selected for the survey.

ceiving cash transfers from the household CCT program[12].

The household survey includes a survey of ever married women, which contains women's pregnancy history in the 24 months prior to the survey. The survey includes detailed information on each pregnancy, such as prenatal visits, birth weight, and delivery assistance. Parity is constructed based on women's pregnancy history. The survey also contains information on the fees paid for prenatal care, delivery, and postnatal care. To replicate the impact evaluation and estimate the program effect on birth outcomes, the main health outcomes are infant death and birth weight. Infant death includes stillbirths and deaths up to 11 months. Birth weight is based on women's recall and reported in grams, conditional on being weighed at birth. Increased use of trained delivery attendants should lower the probability of maternal death from delivery complications, so maternal death is included as an additional health outcome, even though the number of maternal deaths is extremely small in the sample.

The women's survey is also used to estimate the following: demand for health care services, fees paid for delivery assistance, and the quality of prenatal care received. To estimate changes in the demand for each type of delivery attendant, the outcomes of interest are the probability of using doctors, midwives, and traditional birth attendants. Changes in delivery fees paid to each type of provider are also estimated, since participating households should increase delivery fees paid to doctors and/or midwives as a result of the requirement to use of trained delivery attendants. Changes in the quality of care are estimated using a prenatal quality index. The quality index is constructed using principal component analysis of all items in a complete prenatal check-up, the number of tetanus toxoid vaccinations received, an indicator for receiving any information on pregnancy complications, and an indicator for receiving any iron pills[13]. A complete prenatal check-up includes the measurement of the mother's weight, height, blood pressure, fundal height, fetal

[12] Households that report receiving payments in contaminated sub-districts are coded as program beneficiaries. The impact evaluation compares the survey response to administrative data and finds that 4% of households that received the cash transfer did not report it in the survey (Alatas et al., 2011).

[13] Since receiving iron pills is a program requirement, I exclude this indicator for robustness, and the results are qualitatively similar.

heartbeat, blood test, and external and internal pelvic examinations.

To estimate supply changes in treated sub-districts, this paper uses information from the village, health center, and midwife surveys. Villages and sub-district health centers were contacted in both waves to form a panel, and this information is used to estimate the program effect on the number of providers at the village and sub-district clinic levels. Midwives in these sub-districts were also sampled, and 65% of the 1,407 midwives at baseline were re-interviewed in the follow-up survey. The midwife survey is used to estimate the program effect on delivery fees, dual practice and hours worked.

To estimate the program effects on the local health care price and quality, the midwife survey is used to estimate changes in midwife income, list prices, transaction price, and prenatal quality. Changes in midwife income capture changes in her salary, and the quantity and price of all the services she provides. Changes in the list price, given by the price that midwives charge in their public and private practice for a normal delivery[14], capture price changes in public and private practice. The transaction price is given by the actual fees received from each of the last three deliveries they assisted[15]. Transaction price is used as an additional outcome because of the possibility of price discrimination, when midwives receive fees above or below the list price. The analyzed sample is restricted to deliveries that do not use the government's insurance scheme for the poor, *Askeskin*, because there is a separate price and reimbursement scheme for such patients[16]. The midwives' prenatal quality index is constructed using principal component analysis based on self-reported prenatal items[17].

3.2 Summary Statistics

Households in treated and control areas share similar characteristics at baseline, and baseline differences, with district fixed effects included, are not jointly significant

[14] 89% of deliveries in the sample are normal deliveries.

[15] 95% of midwives reported all 3 deliveries.

[16] For the purposes of this paper, prices paid by *Askeskin* patients are not directly affected by the program, and the program did not change the share of *Askeskin* deliveries.

[17] The quality of prenatal care provided is only asked in the follow-up survey, so the estimate is based on the cross-sectional survey.

(Table 1). Baseline pregnancy outcomes between control and treated sub-districts are similar (Panel A). Baseline infant mortality in the sample is about 13 per 1,000, lower than the national average of 26 per 1,000[18]. Maternal mortality in the sample is about 140 per 1,000,000, also lower than the WHO estimate of 220 per 100,000. Reported birth weight in control and treated sub-districts are 3,200 grams (7 lbs.). Delivery assistance at baseline is similar between control and treated sub-districts (Panel B). In control sub-districts, 62% of births were attended by either a doctor or midwife, with 56% of deliveries attended by midwives at baseline. Traditional attendants were present at 43% of births at baseline, and 34% of deliveries were only attended by untrained traditional attendants. Since it is possible for women to use more than one type of provider, the probabilities of using each type of provider are not mutually exclusive. Baseline delivery fees are similar at baseline (Panel C), households spend about about *Rp.* 300,000 (USD 30). Households spend about *Rp.* 100,000 (USD 10) on doctors, *Rp.* 170,000 (USD 17) on midwives, and *Rp.* 40,000 (USD 4) on traditional birth attendants[19].

Both control and treated sub-districts have similar number of doctors, midwives, nurses, and traditional attendants at baseline. Since midwives are the primary care providers, there are about 9 midwives per sub-district clinic, or one per village. Table 2 describes midwife characteristics at baseline. Midwives provide about 14 different services at baseline, including prenatal care, vaccinations, delivery assistance, family planning, and general curative care. Almost 90% of midwives in the sample hold dual practice, and on average, almost half of their *Rp.* 3,000,000 (USD 300) monthly income comes from private practice. Midwives charge *Rp.* 130,000 (USD 13) for normal delivery in their public practice, and the private fee is approximately double the public list price. On average, the fees received from the last three deliveries are similar to the private price at baseline. The high number of hours spent in private practice is consistent with the substantial share of private income.

[18]http://data.worldbank.org/indicator/SP.DYN.IMRT.IN

[19]Conditional on using doctors, households spend about *Rp.* 1,000,000 (USD 100) for delivery assistance. Conditional on using midwives, households spend about *Rp.* 300,000 (USD 30). Conditional on using traditional birth attendants, households spend about *Rp.* 140,000 (USD 14).

3.3 Estimation Strategy

At the household level, the ITT and treatment on the treated (TOT) effects are estimated using the women's survey. The ITT captures the average program effect among near poor and poor households residing in treated sub-districts. This parameter is estimated by the following reduced form equation:

$$y_{isdt} = \delta CCT_{sdt} + \gamma \overline{y_{sd1}} + \alpha_d + \epsilon_{isdt}$$

where y_{isdt} is the delivery fees[20] paid by woman i who resides in sub-district s, in district d, at time t. CCT_{sdt} takes the value one if the sub-district is randomized into treatment. $\overline{y_{sd1}}$ is the baseline value for the sub-district. α_d is a district fixed effect that captures non time-varying district characteristics. Because of the sub-district randomization, the error term, ϵ_{isdt}, is not correlated with individual treatment status, so δ captures the ITT effect of the program on delivery fees paid by near poor and poor households. All standard errors are clustered at the sub-district level. All prices are expressed in 2007 *Rupiah* (1 USD ~ *Rp.* 10,000). Other outcomes of interest include changes in the probability of using each type of delivery attendant, prenatal care quality, and birth outcomes.

The treatment on the treated effect captures the program effects among program beneficiaries. Because individual participation is endogenous, the estimation uses an Instrumental Variable (IV) strategy. The instrument for program participation is the sub-district randomization[21]. There were no refusals from eligible households, so there is no selection into initial compliance among program beneficiaries. The following baseline characteristics are included: mother's education, father's education,

[20] Although the use of log prices allows for a clearer interpretation of the price changes, using log prices results in the loss of observations with zero prices. Consistent with increased use of trained delivery attendants, the program is associated with a lower probability of paying zero fees to trained delivery attendants and midwives, so in order to take these changes into account, level regressions are used.

[21] For any given program beneficiary, individual participation status is correlated with the sub-district assignment, since only households in treated sub-districts who were offered the program could enroll, and the random sub-district assignment is uncorrelated with the unobserved characteristics of program participants.

mother's age, log per capita expenditure, and indicators for asset ownership[22].

On the supply side, the intent-to-treat (ITT) is estimated using midwives' survey. This paper empirically estimates the effect of an exogenous demand shock on local health care price as measured by changes in delivery fees charged by midwives. Since midwives treat both poor and non-poor households, the fees reported by midwives represent the local price. At the midwife level, a similar reduced form equation is used:

$$y_{msdt} = \beta CCT_{sdt} + \mu \overline{y_{sd1}} + \alpha_d + \nu_{msdt}$$

where y_{msdt} is the delivery fees charged by midwife m at sub-district s, in district d, at time t in public and private practice. $\overline{y_{sd1}}$, and α_d are described in the previous equation. ν_{msdt}, the error term, is not correlated with the treatment status because of the sub-district randomization, so β is the ITT parameter for providers, which captures the program effect on midwife fees in sub-districts that are randomized into treatment.

4 Results

4.1 Changes in Utilization, Fees, and Quality

The CCT program requires participating households to obtain prenatal care and delivery assistance from trained attendants to ultimately improve birth outcomes, lower maternal mortality and infant mortality. Table 3 presents the program effects on birth outcomes. For infant mortality, this paper replicates the impact evaluation (Alatas et al., 2011) and finds similar results[23]. In addition, the program also has

[22] The results are similar when individual characteristics are excluded.

[23] The results on birth outcomes may be related to the potential of a perverse incentive on fertility associated with the program. The CCT program may affect fertility by increasing total fertility or reducing birth spacing so as to qualify for the cash transfer. The data does not allow for the first analysis because the women are still in their reproductive years, but birth spacing can be analyzed. Shorter birth spacing (less than 15 months) is associated with adverse birth outcomes, including low birth weight and infant mortality (Conde-Agudelo et al., 2006). The CCT program affects the timing of birth by increasing spacing among program participants. To account for the right censored data, using the Cox survival model, the CCT program is also associated with lower hazard of pregnancy in treated sub-districts. To the extent that longer birth spacing is associated

no significant effect on maternal mortality[24]. For live births, the program has no significant effect on birth weight[25], which is similar to earlier CCT results on child growth indicators in Brazil, Ecuador, Honduras, and Nicaragua (Fizbein et al., 2009).

To estimate demand changes, table 4 presents changes in the utilization trained attendants. Panel A presents the reduced form estimates and panel B presents the IV estimates. Column 1 reproduces the impact evaluation estimates by Alatas et al. (2011) and column 2 replicates it. In the impact evaluation, the initial sub-district randomization is used to estimate the Local Average Treatment Effect (LATE) to capture the program effect on areas that were initially randomized into control that later managed to receive treatment. In the impact evaluation, individual baseline values are included as a regressor, and missing baseline values are imputed based on nearest-neighbor propensity score matching. In the replication, I use the impact evaluation's IV strategy to estimate LATE, but baseline sub-district average was used as a regressor to avoid imputations of individual baseline values. The point estimates of the average placement effects are similar to earlier results. However, the IV effects among program participants are larger than the point estimates in the impact evaluation. These differences are likely to stem from the potentially noisy individual baseline values.

Table 4 also presents changes in the probability of using each type of delivery attendant: doctors, midwives, and traditional attendants (cols. 3-6). In treated communities, the household CCT program is associated with a 50% increase in the probability of using a doctor. Among program participants, the utilization rate increases four-fold relative to their baseline utilization rate. In treated sub-districts, the program is associated with a 15% increase in the probability of using a midwife. Among program participants, the program increases midwife use by 45%, thereby

with better birth outcomes, the results on birth outcomes suggest that changes in birth outcomes are not driven by shorter birth spacing.

[24]Estimated program effects using a rare event logit developed by King and Zeng (2001) are qualitatively similar, and the estimates are also not statistically significant.

[25]Although birth weight is a predictor of infants' long-term health outcomes, the birth weight sample is a selected sample because birth weight is observed conditional on being weighed at birth, and the program increases the probability of being weighed at birth. Low birth weight (birth weight less than 2,500 grams) is 2 percentage points higher in treated sub-districts, but this is not a preferred outcome because of measurement error in birth weight and increased probability of being weighed at birth.

increasing midwife utilization rate to the level of non-participants. Lastly, the CCT program succeeds in lowering the use of traditional attendants by 17% in treated communities and by 30% among program participants. Therefore, the program has successfully increased the use of doctors and midwives and decreased the use of traditional attendants, which is consistent with program objective to increase demand for higher quality health care providers among poor households[26].

To analyze the supply response, table 5 presents changes in provider availability. The household CCT program has no significant effect on the number of doctors, but the program increases the number of midwives affiliated with sub-district clinics by 10% (col. 2). Villages also consistently report a 10% increase in midwife availability (col. 4) and no significant change in the number of doctors or untrained traditional attendants (cols. 5-6). We expect the number of midwives to respond more easily than doctors since nurses could receive one year of additional training to become midwives[27]. In terms of changes at the midwife level (cols. 6-8), the program is associated with a 0.5 reduction in the number of services provided[28]. While this is a potential concern for the availability of health care services, there are no statistically significant changes in dual practice or total hours worked[29]. Overall, the program is associated with increased number of midwives, which should mitigate the price increase in treated sub-districts.

In spite of the increased number of midwives, among near poor and poor households, the program is associated with higher delivery fees paid to midwives, which is consistent with increased utilization. Table 6 presents changes in delivery fees, followed by fees paid to trained delivery attendants, doctors, midwives, and traditional attendants. On average, the program is associated with a 10% increase in total expenditure on childbirth, which is driven by increased fees paid to doctors or midwives. Near poor and poor households residing in treated sub-districts spend 30% more on delivery fees paid to doctors and 25% more on midwives. These results are consis-

[26] On the supply side, sub-district clinics and midwives in treated areas report a small but not statistically significant increase in the number of deliveries.

[27] Anecdotal evidence suggests that this is the case in West Java.

[28] The reduction can be attributed to lower probabilities of providing family planning consultation and implant contraception.

[29] Changes in hours spent in private and public service are also not statistically significant.

tent with increased utilization of doctors and midwives for delivery assistance. In addition, because the program is associated with lower use of traditional attendants, households in treated sub-districts now spend 30% less on traditional attendants for childbirth. Program participants experience a 60% increase in total expenditure on childbirth, driven by a 150% increase in delivery fees paid to midwives. In addition, program participation is associated with a 50% decrease in delivery fees paid to traditional attendants. Program beneficiaries now pay approximately the same fees that non-participants pay to midwives, suggesting that the poor can now afford higher quality delivery assistance. These estimates imply that program beneficiaries who receive the minimum transfer amount spend about 30% of the transfer on delivery fees.

To analyze price changes in the local health care market, table 7 presents changes in delivery fees as reported by midwives. The program is associated with 10% higher monthly income for midwives (col. 1), which may come from increased salary or income from public and private practice. As expected, the CCT program has no statistically significant effect on public fees for normal delivery (col. 2). On the other hand, midwives are able to respond to the demand shock in private practice (cols. 3-4). The program leads to a 10% increase in private fees for normal delivery and a 5% increase in delivery fees received from the last three deliveries[30]. These results provide evidence that the program is associated with a price increase in the local health care market[31]. Although these price increases are small and do not affect the affordability of care for program participants, this price increase may have implications on the affordability of care among poor households that are not eligible for the program.

The discrepancy between the midwife and household reports is related to the

[30]Since dual practice midwives may be able to increase price more easily, restricting the sample to dual practice midwives yields similar results. As an alternative specification, the sample is also restricted to panel midwives to include midwife fixed effects, the results are similar to the estimation using the full sample. Among panel midwives, the program is associated with increased midwife income and private fees. Using difference-in-differences, the program is associated with a 5% increase in delivery fees in private practice, while the effect is not significant on public practice.

[31]The program is also associated with higher private fees for the following services: general check-up, prenatal care, BCG, measles, and tetanus vaccinations. The program has no significant effects on other public fees.

population that midwives serve and the changes in utilization among poor households. Midwives treat both program participants and non-participants, so the price increase reported by midwives represents the estimated program effect in the local health care market. However, the larger fee increase reported by households is driven by increased utilization among poor households. The household CCT program lowers the probability of paying zero midwife fees among poor households. On the other hand, midwives report no statistically significant change in the probability of receiving zero fees. These results suggests that the fee increase reported by poor households is driven by their increased use of midwife services, and not because midwives increase prices for the poor to capture some of the transfers received by households.

Price and utilization changes associated with the CCT program may affect the quality of health care provided as measured by prenatal quality. Table 8 presents quality changes as reported by households (cols. 1-2) and midwives (cols. 3-4). Among program participants, prenatal quality improvements accompany the increased utilization of midwives for prenatal care[32]. Using the prenatal quality index, the program is associated with an average increase of 0.08 standard deviation, and a 0.17 standard deviation increase among program beneficiaries[33]. On the supply side, using midwives' self-reported quality in their public and private practice, there is no evidence of any significant quality change (cols. 3 and 4). These results suggest that the quality improvements experienced by households result from increased use of prenatal care, instead of improvements in the average quality of care in the local market. This provides suggestive evidence that low quality of care partly contributes to the lack of improvements in birth outcomes in spite of higher utilization.

To further explore the role of midwife quality on program effectiveness, table 9

[32] The program has increased the probability of obtaining prenatal care from midwives by 7% on average and by 13% among participants, which is consistent with the impact evaluation that finds a 13% increase in obtaining at least four prenatal visits (Alatas et al., 2011). Midwives in treated sub-districts also report an increase in the number of prenatal visits.

[33] Although the program has no statistically significant effect on the probability of receiving a complete prenatal check, the program is associated with a 30% higher probability of receiving the alternative prenatal check indicator, which excludes external and internal examinations. Among program beneficiaries, the program is associated with an 80% increase in quality, which increases the average prenatal quality received to the level of non-participants.

analyzes changes in midwife characteristics and the possibility of midwives selecting into treated areas. The first possibility is the migration of experienced midwives to treated areas, followed by changes in the qualification and experience of midwives, and the number of services provided. A midwife is coded as a migrant if she has spent less than 2 years in the sub-district clinic, but has more than 3 years of experience[34]. The program has no statistically significant effect on the probability of midwife migration to treated areas[35] (col. 1). I then explore the program effect on midwives' level of education and years of experience as a measure of midwife quality. Before 1998, midwives could be certified with a 1-year diploma, but midwives are now required to complete a three-year diploma program, *Akademi Bidan*[36] (Midwife Academy) (Heywood et al., 2010). The CCT program is associated with an increase in the number of educated midwives (col. 2), but the program is also associated with lower levels of experience (cols. 3 and 4). In addition, the CCT program is associated with fewer available services (col. 5), which is consistent with less experienced midwives providing a more limited selection of services, and it also raises the possibility of the program putting a strain on the capacity of the available midwives. These results suggest that the program does little to improve provider characteristics associated with better quality of health care service.

4.2 Heterogeneous Treatment Effects

The program increases delivery fees among near poor and poor households, which is a concern if price increases are passed on to the poor and the increase becomes a barrier to health care access. In addition, it is also common for health care providers in developing countries to price discriminate by charging higher prices to the wealthy

[34]Recent graduates are typically assigned to sub-districts and sign a three-year initial contract, while more experienced midwives have higher mobility.

[35]To estimate migration from control to treated areas, I estimate the change in the number of midwives in control sub-districts that are located near a treated sub-district. The distance indicator for takes the value one when a sub-district is within 3.5 miles of a treated sub-district, which corresponds to the median distance. There is also no evidence of control areas losing midwives, which suggests that midwives do not move to treated sub-districts in response to the CCT program.

[36]Prior to 1989, midwife training only required one year of education after junior high school. In 1989, the government started a midwife program, *Program Pendidikan Bidan* (Midwife Education Program), to provide midwives with basic nursing qualifications. The current workforce consists of midwives trained in all three programs, with an increasing share of midwives with a 3-year diploma.

and lower prices to the poor (Gertler and Solon, 2000; Killingsworth et al., 1999). Since program beneficiaries now have more resources to pay for health care services, health care providers may increase the price charged to the poor. To address the possibility of heterogeneous treatment effects, I estimate changes among households in the bottom quintile relative to households from the highest expenditure quintile (table 10[37]). In terms of access to midwives (col.1), there is no evidence of differences between households in the lower quintile and those in the top quintile. Although there is evidence of an average price increase in the treated communities, there is no evidence of heterogeneous price changes by baseline household expenditure level, so the price increase does not appear to be disproportionately passed on to the poorest households (col. 2).

Although the program is associated with improved prenatal quality received by the poor, for the program to serve the poor and improve equality of health care access, it is important to analyze whether the poor benefit from the quality improvements. I find that households in the bottom quintile experience larger quality improvements relative to those in the top quintile (col. 3). The CCT program is associated with a 0.4 standard deviation increase for households in the bottom three quintiles, which is the population most likely affected by the CCT program. Although the quality improvements benefit the poorest households, the program has no differential effect on birth weight (col. 4), which further suggests that the reported improvements in quality on the demand side come from increased access. These results suggest the need to further analyze the link between prenatal care and birth outcomes.

To further explore the interaction between prenatal quality and health outcomes, an interaction term between high baseline quality and treatment status is included (table 11). The indicator for high quality takes the value one if households in the sub-district report receiving above-median prenatal quality at baseline. Although there is no statistically significant effect on infant mortality, the program improves some birth outcomes in higher quality areas. The program continues to be associated with a higher incidence of reported low birth weight, but in high quality areas, the

[37] One concern with the use of household expenditure as a wealth measure is that it may be a poor measure of wealth. To address this concern, the same analysis is run using a wealth index, which is created based on asset ownership using principal component analysis. The results are similar using this alternative wealth measure.

program reduces the reported incidence of low birth weight. The program is also associated with higher average birth weight in high quality areas. Using midwives' self-reported quality in the follow-up survey, the cross-sectional results are not statistically significant, but are qualitatively similar for low birth weight and infant mortality. These results further suggest that prenatal quality is key to improving birth outcomes.

CCT programs require supply-readiness, and areas in Java are more supply-ready than areas off-Java. Consequently, the impact evaluation finds stronger results on prenatal and post-natal visits in Java (Alatas et al., 2011). In terms of price changes, the program is associated with a larger price increase off Java as reported by both households and midwives. On the demand side, households in Java report a larger increase in prenatal quality received, where the average baseline quality is higher. On the supply side, midwives in Java and off-Java report no change in the quality of care provided. These results further underscore the need for supply-readiness for CCT programs to be effective.

5 Discussion

Unlike earlier results that find no substantial price or wage increases in the local economy, Indonesia's CCT program is associated with a small price increase in the local health care market, which is directly affected by the program. The price increase is driven by changes in the private health care market, which suggests that dual practice contributes to providers' ability to respond to the demand shock. One concern with increased private fees is that higher fees would limit the affordability of health care services for the poor. Although private practice seems to be responsible for the price increase, dual practice increases provider availability, since midwives may not enter the profession at all if they were not able to enter private practice to supplement their income (Gruen et al., 2002). The CCT program increases utilization among the near poor and poor households, and the increase in fees paid to providers is driven by increased utilization, suggesting that the price increase does not appear to limit the effectiveness of the CCT program in meeting the targeted indicators.

To explore whether the results are unique to the household CCT program, I compare the household CCT to the community CCT program that was piloted simultaneously in the same provinces in Indonesia. The community CCT program is also a demand-driven program, but it is simpler and cheaper to administer compared to the household CCT program. The community CCT program allows communities to target both demand and supply constraints, so we expect the community CCT program to generate both demand and supply shocks. However, individual participation in the community CCT program is voluntary, so participants are likely to have a higher propensity to use health care services. In spite of the differences in mechanism, the two programs are comparable since both programs target the same indicators. The community CCT impact evaluation finds a small and non-significant increase in the use of trained delivery attendants, but the community CCT is associated with a 5% increase in delivery fees with no statistically significant change in prenatal care quality[38] (Olken et al., 2010). These results provide further evidence that any demand shock in the health care market leads to local price increases.

Although Indonesia's CCT program increases the use of trained delivery attendants and prenatal quality among near poor and poor households, these improvements do not translate to better birth outcomes. These results can be partly explained by Indonesia's provider availability compared to other countries that have implemented CCT programs, such as Mexico. Mexico's CCT program has been shown to improve health outcomes (Gertler, 2004), and the improvements in birth outcomes can be explained by improved prenatal care quality (Barber and Gertler, 2010). The CCT program requires supply readiness, so that the local health care system would be able to meet the additional demand without increasing price or compromising the quality of care. However, in spite of the selection criteria, Indonesia's provider availability was more limited compared to Mexico's when the CCT program was launched in 1997. In 2000, Mexico had 11 nurses and midwives per 10,000 and 17 physicians per 10,000[39]. In contrast, Indonesia had 8 nurses and midwives per 10,000 and 1 physicians per 10,000 in 2007[40]. These statistics suggest that provider availability needs to be improved as Indonesia's CCT program expands.

[38]Replicating the estimation, I also find that the community CCT program is associated with a price increase.

[39]WHO Health Systems Statistics, 2005

[40]WHO Health Systems Statistics, 2005

To further explore the role of quality of care in producing health outcome, I compare Indonesia's household CCT program to the community CCT program. The community CCT impact evaluation finds no statistically significant change in prenatal care quality (Olken et al., 2010), and consequently, no significant change in low birth weight (Triyana, 2013). However, consistent with results from the household CCT program, reported incidence of low birth weight falls in community CCT areas with higher baseline prenatal quality. These results are consistent with results from Mexico's CCT program that finds that the reduction in low birth weight is driven by improvements in prenatal care quality (Barber and Gertler, 2009, 2010). These results highlight the importance of prenatal care quality in improving birth outcomes.

6 Conclusion

Indonesia's CCT program generates a demand shock in the health care market as measured by increased use of trained providers for delivery assistance. Although supply readiness mitigates the price increase in the health care market, the demand shock generated by the program still generates a health care price increase. As CCT programs become more widely implemented in developing countries, one key consideration before program implementation is the availability of high quality health care providers to meet additional demand so as to mitigate price increases. The price increase in the health care market could limit access among the poor who are not eligible for the program, since affordability is one barrier to health care access among the poor.

The increase in delivery fees paid by near poor and poor households is accompanied by quality improvements as measured by prenatal care quality received, but these changes do not lead to improved birth outcomes. Improvements in birth outcomes are linked to higher baseline quality of prenatal care, which underscores the importance of quality of care. Although CCT programs require sufficient provider availability, a more important consideration is the quality of health care services. For CCT programs to successfully improve human capital, they need to address both the quantity and quality of health care supply in order to translate behavioral changes to improved health outcomes.

References

Alatas, V., A. Banerjee, R. Hanna, B. A. Olken, and J. Tobias (2012). Targeting the poor: Evidence from a field experiment in indonesia. *American Economic Review 102*(4), 1206–40. NBER Working Paper No. w15980.

Alatas, V., N. Cahyadi, E. Ekasari, S. Harmoun, B. Hidayat, E. Janz, J. Jellema, H. Tuhiman, and M. Wai-Poi (2011). Main findings from the impact evaluation of indonesia's pilot household conditional cash transfer program.

Alexander, G. and C. Korenbrot (1995). The role of prenatal care in preventing low birth weight. *The future of children*, 103–120.

Almond, D., K. Y. Chay, and D. S. Lee (2005). The costs of low birth weight. *Quarterly Journal of Economics 120*(3), 1031–1083.

Angelucci, M. and G. De Giorgi (2009). Indirect effects of an aid program: How do cash transfers affect ineligibles' consumption? *The American Economic Review 99*(1), 486–508.

Attanasio, O., E. Battistin, E. Fitzsimons, and M. Vera-Hernandez (2005). How effective are conditional cash transfers? evidence from colombia.

Barber, S., P. Gertler, and P. Harimurti (2007). Differences in access to high-quality outpatient care in indonesia. *Health Affairs 26*(3), w352–w366.

Barber, S. L. and P. J. Gertler (2008). The impact of mexico's conditional cash transfer programme, oportunidades, on birthweight. *Tropical Medicine and International Health 13*(11), 1405–1414.

Barber, S. L. and P. J. Gertler (2009). Empowering women to obtain high quality care: Evidence from an evaluation of mexico's conditional cash transfer programme. *Health Policy and Planning 24*(1), 18–25.

Barber, S. L. and P. J. Gertler (2010). Empowering women: how mexico's conditional cash transfer programme raised prenatal care quality and birth weight. *Journal of Development Effectiveness 2*(1), 51–73.

Barker, D. J. (1990). The fetal and infant origins of adult disease. *BMJ: British Medical Journal 301*(6761), 1111.

Behrman, J. R. and M. R. Rosenzweig (2004). Returns to birthweight. *The Review of Economics and Statistics 86*(2), 586–601.

Bhutta, Z., G. Darmstadt, B. Hasan, and R. Haws (2005). Community-based interventions for improving perinatal and neonatal health outcomes in developing countries: a review of the evidence. *Pediatrics 115*(Supplement 2), 519–617.

Bir, A. and K. Eggleston (2003). Physician dual practice: access enhancement or demand inducement. *Tufts university, department of economics, working paper 11.*

Blanc, A. and T. Wardlaw (2005). Monitoring low birth weight: an evaluation of international estimates and an updated estimation procedure. *Bulletin of the World Health Organization 83*(3), 178–185d.

Boerma, J., K. Weinstein, S. Rutstein, and A. Sommerfelt (1996). Data on birth weight in developing countries: can surveys help? *Bulletin of the World Health Organization 74*(2), 209.

Christian, P., K. West, S. Khatry, S. Leclerq, E. Pradhan, J. Katz, S. Shrestha, and A. Sommer (2003). Effects of maternal micronutrient supplementation on fetal loss and infant mortality: a cluster-randomized trial in nepal. *The American journal of clinical nutrition 78*(6), 1194–1202.

Cohen, J., P. Dupas, B. I. G. Economy, and Development (2010). Free distribution or cost-sharing? evidence from a randomized malaria prevention experiment*. *Quarterly Journal of Economics 125*(1), 1.

Conde-Agudelo, A., A. Rosas-Bermúdez, and A. Kafury-Goeta (2006). Birth spacing and risk of adverse perinatal outcomes. *JAMA: the journal of the American Medical Association 295*(15), 1809–1823.

Cunha, J., G. De Giorgi, and S. Jayachandran (2011). The price effects of cash versus in-kind transfers. Technical report, National Bureau of Economic Research.

Currie, J. (2009). Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. *Journal of Economic Literature 47*(1), 87–122.

Currie, J. and R. Hyson (1999). Is the impact of health shocks cushioned by socioeconomic status? the case of low birthweight. Technical report, National Bureau of Economic Research.

Das, J., Q. Do, and B. Özler (2005). Reassessing conditional cash transfer programs. *The World Bank Research Observer 20*(1), 57–80.

Eggleston, K. and A. Bir (2006). Physician dual practice. *Health Policy 78*(2), 157–166.

Engle, P., L. Fernald, H. Alderman, J. Behrman, C. O'Gara, A. Yousafzai, M. De Mello, M. Hidrobo, N. Ulkuer, I. Ertem, et al. (2011). Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries. *The Lancet*.

Ensor, T., Z. Quayyum, M. Nadjib, and P. Sucahya (2009). Level and determinants of incentives for village midwives in indonesia. *Health policy and planning 24*(1), 26–35.

Fizbein, A., N. Schady, F. Ferreira, M. Grosh, N. Kelleher, P. Olinto, and E. Skoufias (2009). *Conditional cash transfers: reducing present and future poverty.* World Bank.

Gertler, P. and J. Molyneaux (1996). Pricing public health services: results from a social experiment in indonesia. *RAND mimeograph*.

Gertler, P. and O. Solon (2000). Who benefits from social health insurance in developing countries? *Draft. UC Berkeley*.

Gertler, P. J. (2004). Do conditional cash transfers improve child health? evidence from progresa's control randomized experiment. *The American Economic Review 94*(2), 336–341.

Gruen, R., R. Anwar, T. Begum, J. Killingsworth, and C. Normand (2002). Dual job holding practitioners in bangladesh: an exploration. *Social science & medicine 54*(2), 267–279.

Handa, S. and B. Davis (2006). The experience of conditional cash transfers in latin america and the caribbean. *Development Policy Review 24*(5), 513–536.

Heywood, P. and N. Harahap (2009). Human resources for health at the district level in indonesia: the smoke and mirrors of decentralization. *Human Resources for Health 7*(1), 6.

Heywood, P., N. Harahap, E. Mimin Ratminah, et al. (2010). Current situation of midwives in indonesia: Evidence from 3 districts in west java province. *BMC Research Notes 3*(1), 287.

Jan, S., Y. Bian, M. Jumpa, Q. Meng, N. Nyazema, P. Prakongsai, and A. Mills (2005). Dual job holding by public sector health professionals in highly resource-constrained settings: problem or solution? *Bulletin of the World Health Organization 83*(10), 771–776.

Killingsworth, J., N. Hossain, Y. Hedrick-Wong, S. Thomas, A. Rahman, and T. Begum (1999). Unofficial fees in bangladesh: price, equity and institutional issues. *Health Policy and Planning 14*(2), 152–163.

King, G. and L. Zeng (2001). Logistic regression in rare events data. *Political analysis 9*(2), 137–163.

Kremer, M. and E. Miguel (2007). The illusion of sustainability. *The Quarterly Journal of Economics 122*(3), 1007–1065.

Macours, K., N. Schady, and R. Vakis (2008). *Cash transfers, behavioral changes, and cognitive development in early childhood: evidence from a randomized experiment.* World Bank.

Macq, J., P. Ferrinho, V. De Brouwere, and W. Van Lerberghe (2001). Managing health services in developing countries: between ethics of the civil servant and the need for moonlighting: managing and moonlighting. *Human Resources for Health Development Journal 5*, 17–24.

Makowiecka, K., E. Achadi, Y. Izati, and C. Ronsmans (2008). Midwifery provision in two districts in indonesia: how well are rural areas served? *Health Policy and Planning 23*(1), 67–75.

Miguel, E. and M. Kremer (2004). Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica 72*(1), 159–217.

Morris, S., P. Olinto, R. Flores, E. Nilson, and A. Figueiro (2004). Conditional cash transfers are associated with a small reduction in the rate of weight gain of preschool children in northeast brazil. *The Journal of nutrition 134*(9), 2336–2341.

Mosley, W. and L. Chen (1984). An analytical framework for the study of child survival in developing countries. *Population and development review 10*, 25–45.

Mundial, B. (2004). World development report 2004: Making services work for poor people. *Banque mondiale, Washington, DC*.

Olken, B. A. (2007). Monitoring corruption: Evidence from a field experiment in indonesia. *Journal of Political Economy 115*(2), 200–249.

Olken, B. A., J. Onishi, and S. Wong (2010). Indonesia's *PNPM Generasi* program: Final impact evaluation report. Olken Onishi Wong.

Onishi, J. (2010). *A multilevel analysis of village and local health systems factors affecting maternal health service delivery and use in rural Indonesia*. Ph. D. thesis, THE JOHNS HOPKINS UNIVERSITY.

Paxson, C. and N. Schady (2010). Does money matter? the effects of cash transfers on child development in rural ecuador. *Economic Development and Cultural Change 59*(1), 187–229.

Rawlings, L. and G. Rubio (2005). Evaluating the impact of conditional cash transfer programs. *The World Bank Research Observer 20*(1), 29–55.

Rivera, J., D. Sotres-Alvarez, J. Habicht, T. Shamah, and S. Villalpando (2004). Impact of the mexican program for education, health, and nutrition (progresa) on rates of growth and anemia in infants and young children. *JAMA: the journal of the American Medical Association 291*(21), 2563–2570.

Rokx, C., J. Giles, E. Satriawan, P. Marzoeki, and P. Harimurti (2010). *New insights into the provision of health services in Indonesia: a health workforce study*. World Bank Publications.

Schubert, B. and R. Slater (2006). Social cash transfers in low-income african countries: Conditional or unconditional? *Development Policy Review 24*(5), 571–578.

Todd, P. E. (2007). Evaluating social programs with endogenous program placement and selection of the treated. *Handbook of development economics 4*, 3847–3894.

Triyana, M. (2013). *The Effects of Household and Community-based Interventions: Evidence from Indonesia*. Ph. D. thesis.

Wai-Poi, M. (2011). *Three Essays on Development Economics: Household Welfare*. Ph. D. thesis, COLUMBIA UNIVERSITY.

Whitehead, M., G. Dahlgren, T. Evans, et al. (2001). Equity and health sector reforms: can low-income countries escape the medical poverty trap? *Lancet 358*(9284), 833–836.

Tables and Figures

Table 1: Baseline Household Characteristics

		Control	Treatment	Adjusted Difference
		(1)	(2)	(3)
Panel A. Baseline delivery outcomes				
Infant mortality		0.011	0.013	0.002
		(0.106)	(0.115)	(0.003)
		2,814	2,773	
Maternal mortality		139.7	125.7	18.6
(per 100,000)		(3,735)	(3,543)	(96.0)
		2,148	3,183	
Birth weight (grams)		3,180.7	3,167.1	-13.25
		(552.4)	(565.5)	(19.21)
		1,719	1,714	
Panel B. Baseline delivery assistance				
Birth at a healthcare facility		0.440	0.435	-0.011
		(0.497)	(0.496)	(0.018)
		2,301	2,264	
Birth Attendant:	Trained Attendant	0.623	0.639	0.013
		(0.485)	(0.480)	(0.015)
		2,301	2,264	

Notes: Baseline differences in Panels A, B, and C are not jointly significant, F-test p-value 0.704. Column (3) presents the difference between households residing in treatment and control sub-districts with district fixed effects included. All prices in 2007 *Rupiah* (1 USD ~ *Rp.* 10,000). Trained delivery attendants include doctors and midwives. Infant mortality includes stillbirths and infant deaths up to 11 months. Maternal mortality includes all pregnancy-related deaths reported by households. Birth weight in grams as reported by mothers.
Robust standard errors in parentheses, clustered at sub-district level, * p<0.1, **p<0.05, ***p<0.01.

Table 1: Baseline Household Characteristics

	Control	Treatment	Adjusted Difference
	(1)	(2)	(3)
Doctor	0.088	0.079	-0.012
	(0.283)	(0.269)	(0.008)
	2,301	2,264	
Midwives	0.562	0.588	0.024
	(0.496)	(0.492)	(0.015)
	2,301	2,264	
Traditional Attendant	0.431	0.419	-0.008
	(0.495)	(0.494)	(0.018)
	2,301	2,264	
Panel C. Baseline delivery fees			
Delivery fees	298,739	284,611	-17,216
	(766,112)	(573,472)	(17,914)
	2,283	2,238	
Delivery fees paid to:			
Trained professional	256,255	243,838	-15,902
	(773,842)	(584,289)	(18,123)
	2,283	2,238	
Doctor	102,722	82,113	-22,163
	(746,292)	(541,178)	(17,339)
	2,283	2,238	

Notes: Baseline differences in Panels A, B, and C are not jointly significant, F-test p-value 0.704. Column (3) presents the difference between households residing in treatment and control sub-districts with district fixed effects included. All prices in 2007 *Rupiah* (1 USD $\sim$ *Rp.* 10,000). Trained delivery attendants include doctors and midwives. Infant mortality includes stillbirths and infant deaths up to 11 months. Maternal mortality includes all pregnancy-related deaths reported by households. Birth weight in grams as reported by mothers.

Robust standard errors in parentheses, clustered at sub-district level, * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 1: Baseline Household Characteristics

	Control	Treatment	Adjusted Difference
	(1)	(2)	(3)
Midwife	176,276	178,389	-325
	(363,480)	(336,313)	(9,668)
	2,283	2,238	
Traditional Attendant	40,711	39,192	-1,057
	(90,292)	(81,617)	(2,683)
	2,194	2,171	

Notes: Baseline differences in Panels A, B, and C are not jointly significant, F-test p-value 0.704. Column (3) presents the difference between households residing in treatment and control sub-districts with district fixed effects included. All prices in 2007 *Rupiah* (1 USD $\sim$ *Rp.* 10,000). Trained delivery attendants include doctors and midwives. Infant mortality includes stillbirths and infant deaths up to 11 months. Maternal mortality includes all pregnancy-related deaths reported by households. Birth weight in grams as reported by mothers.
Robust standard errors in parentheses, clustered at sub-district level, * p<0.1, **p<0.05, ***p<0.01.

Table 2: Baseline Provider Characteristics

		Control	Treatment	Adjusted Difference
		(1)	(2)	(3)
Number of midwives:	Sub-district clinic	9.067	8.129	0.140
		(5.844)	(5.320)	(0.470)
		180	180	
	Village	1.200	1.232	0.111
		(1.539)	(1.548)	(0.079)
		1,337	1,358	
Midwife practice:	Share with private practice	0.894	0.857	-0.044**
		(0.308)	(0.350)	(0.019)
		696	696	
	Total income	3,036,459	3,079,905	226,836*
		(2,304,431)	(2,477,733)	(123,713)
		702	702	
	Share Private income	0.472	0.468	-0.006
		(0.282)	(0.281)	(0.018)
		698	698	
Number of services		13.920	14.039	-0.123
		(3.563)	(3.221)	(0.167)
		702	702	
Price for normal : delivery	Public	128,684	149,758	-18,554
		(146,333)	(201,070)	(13,194)
		603	603	
	Private	295,539	296,371	-2,887
		(141,156)	(139,032)	(5,504)

Notes: Col. 3 presents the difference between treatment and control groups with district fixed effects included. All prices and income in 2007 *Rupiah* (1 USD $\sim$ *Rp.* 10,000). Baseline differences are not jointly significant, F-test p-value 0.377. Robust standard errors in parentheses, clustered at the sub-district level, * p<0.1, **p<0.05, ***p<0.01.

Table 2: Baseline Provider Characteristics

		Control	Treatment	Adjusted Difference
		(1)	(2)	(3)
		665	665	
Hours worked in the last 3 days	: Public	16.776	16.104	0.751
		(9.997)	(9.854)	(0.691)
		702	702	
	Private	11.677	11.943	-0.923
		(11.562)	(11.690)	(0.988)
		702	702	

Notes: Col. 3 presents the difference between treatment and control groups with district fixed effects included. All prices and income in 2007 *Rupiah* (1 USD ~ *Rp.* 10,000). Baseline differences are not jointly significant, F-test p-value 0.377. Robust standard errors in parentheses, clustered at the sub-district level, * p<0.1, **p<0.05, ***p<0.01.

Table 3: Changes in Health Outcomes

	Infant mortality		Maternal mortality	Birth weight
	Alatas et al. (2011)	Replication by Triyana (2013)		
	(1)	(2)	(3)	(4)
Panel A. Reduced Form Effect of CCT Program in Sub-District				
Treatment	0.002	0.003	-0.0004	-0.570
	(0.00)	(0.003)	(0.0008)	(26.04)
Observations	-	8,302	7,896	4,987
R-Squared	-	0.083	0.014	0.089
Baseline Dependent	0.0102	0.011	0.001	3,180.76
Variable		(0.106)	(0.037)	(552.39)
Panel B. IV Effect of Program Participation				
Ever Received CCT	0.005	0.006	-0.0008	-1.307
	(0.00)	(0.007)	(0.0017)	(59.15)
Observations	-	8,302	7,896	4,987
R-Squared	-	0.084	0.014	0.089
Baseline Mean among	-	0.031	0.004	3200.00
Program Participants		(0.173)	(0.063)	(628.78)

Notes: District fixed effects included in all specifications. Infant mortality includes still-births and deaths up to 12 months. Maternal mortality includes all pregnancy-related deaths reported by households. Birth weight in grams, reported by mothers. Sub-district randomization is used as instrument for individual program participation. Individual characteristics include education, husband's education, log per capita expenditure, age, and indicators for home ownership, and land ownership. Robust standard errors in parentheses, clustered at sub-district level, * p<0.1, **p<0.05, ***p<0.01.

Table 4: Changes in the Utilization of Delivery Attendants

	P(Trained Delivery Attendant)		P(Use of Doctor)	P(Use of Midwife)	P(Use of Traditional Attendant)
	Alatas et al. (2011)	Replication Triyana (2013)			
	(1)	(2)	(3)	(4)	(5)
Panel A. Reduced Form Effect of CCT Program in Sub-Districts					
Treatment	0.020	0.037*	0.039***	0.082***	-0.081***
	(0.03)	(0.020)	(0.012)	(0.017)	(0.015)
Observations	-	6,628	6,628	6,628	6,628
R-Squared		0.306	0.120	0.243	0.302
Baseline Mean	0.602	0.623	0.088	0.562	0.431
		(0.485)	(0.283)	(0.496)	(0.495)
Panel B. IV Effect of Program Participation					
Ever Received CCT	0.037	0.092*	0.083***	0.172***	-0.171***
	(0.03)	(0.049)	(0.025)	(0.035)	(0.033)
Observations	-	6,628	6,628	6,628	6,628
R-Squared		0.304	0.117	0.237	0.298
Baseline Mean among	-	0.394	0.025	0.375	0.600
Program Participants		(0.490)	(0.157)	(0.485)	(0.491)

Notes: District fixed effects included in all specifications. Sub-district randomization is used as instrument for individual program participation. Individual characteristics include education, husband's education, log per capita expenditure, age, and indicators for home ownership, and land ownership. Trained healthcare professionals include doctors and midwives. In column 1, Alatas et al. (2011) estimate LATE using all sub-districts in the sample, instrumenting for sub-districts' actual treatment status using the intial randomization. They control for baseline values at the individual level, and missing values are imputed based on propensity score matching. Column 2 uses the same IV strategy to estimate LATE, but sub-district baseline values are included to avoid individual baseline imputations. The probabilities of using midwives, doctors, and traditional attendants are not mutually exclusive because women may use more than one type of delivery attendant. Robust standard errors in parentheses, clustered at sub-district level, * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 5: Changes in the Availability of Health Care Services

	Number of Providers					Midwife Availability		
	Clinic Report (Sub-District Level)		Community Report (Village Level)			Midwife Report (Midwife Level)		
	Doctors	Midwives	Doctors	Midwives	Traditional Attendants	Any private practice	Number of Services	Total Hours
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	-0.029	0.748***	-0.004	0.115***	0.047	0.022	-0.456**	-0.681
	(0.075)	(0.190)	(0.057)	(0.022)	(0.030)	(0.015)	(0.208)	(0.663)
Observations	713	713	5,426	5,434	5,429	2,785	2,790	2,790
R-squared	0.717	0.835	0.204	0.123	0.149	0.454	0.205	0.203
Mean Dependent	1.637	9.339	0.348	1.200	0.575	0.894	13.92	28.453
Variable	(1.009)	(4.941)	(0.831)	(1.539)	(0.971)	(0.308)	(3.563)	(15.541)

Notes: District fixed effects included in all specifications. Treatment indicator equals one when clinics, villages, or midwives are located in sub-districts randomized into treatment. Robust standard errors in parentheses, clustered at sub-district level, * p<0.1, **p<0.05, ***p<0.01.

Table 6: Household Report: Changes in Fees Paid for Childbirth

	Delivery Fees	Fees Paid to Trained Attendant	Fees Paid to Doctors	Fees Paid to Midwives	Fees Paid to Traditional Attendants
	(1)	(2)	(3)	(4)	(5)
Panel A. Reduced Form Effect of CCT Program in Sub-District					
Treatment	37,203*	44,484**	32,720*	45,917***	-12,285***
	(20,436)	(20,968)	(18,850)	(13,138)	(2,725)
Observations	6,544	6,544	6,546	6,544	6,247
R-squared	0.166	0.167	0.075	0.182	0.147
Baseline Mean	298,739	256,255	102,721	176,276	40,711
	(766,112)	(773,842)	(746,473)	(363,480)	(90,292)
Panel B. IV Effect of Program Participation					
Ever Received CCT	78,544*	93,919**	69,089*	96,949***	-26,311***
	(43,305)	(44,544)	(39,820)	(28,722)	(5,955)
Observations	6,544	6,544	6,546	6,544	6,247
R-squared	0.163	0.164	0.073	0.174	0.142
Baseline Mean among	114,032	62,359	0	62,359	49,407
Program Participants	(135,320)	(130,049)	(0)	(130,049)	(78,151)

Notes: District fixed effects included in all specifications. Sub-district randomization is used as instrument for individual program participation. Individual characteristics include education, husband's education, log per capita expenditure, age, and indicators for home ownership and land ownership. All prices in 2007 *Rupiah* (1 USD ~ *Rp.* 10,000). Trained attendant includes doctors and midwives. Robust standard errors in parentheses, clustered at sub-district level, * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 7: Midwife Report: Changes in Delivery Fees

	Total Income	Fees for Normal Childbirth: Public fees	Fees for Normal Childbirth: Private fees	Fees Received from Last 3 Deliveries
	(1)	(2)	(3)	(4)
Treatment	335,314***	-3,161	27,062***	14,925**
	(115,473)	(10,554)	(4,488)	(5,806)
Observations	2,790	2,181	2,556	5,884
R-squared	0.277	0.228	0.661	0.269
Mean Dependent	3,036,459	128,684	295,539	335,238
Variable	(2,304,431)	(146,333)	(141,156)	(203,184)

Notes: District fixed effects included in all specifications. All prices in 2007 *Rupiah* (1 USD $\sim$ *Rp.* 10,000). Treatment indicator equals one when midwives are located in sub-districts randomized into treatment. Robust standard errors in parentheses, clustered at sub-district level, * p<0.1, **p<0.05, ***p<0.01.

Table 8: Changes in Prenatal Care Quality

	Household Report		Midwife Report	
	Reduced Form Effect	Effect of Program Participation	Public Practice	Private Practice
	OLS	IV	OLS	OLS
	(1)	(2)	(3)	(4)
Treatment	0.080**	0.167**	-0.009	-0.021
	(0.034)	(0.070)	(0.050)	(0.052)
Observations	8,302	8,302	1,396	1,396
Baseline Mean	0.000	-0.317	0.000	0.000
	(1.000)	(1.122)	(1.000)	(1.000)

Notes: District fixed effects included in all specifications. Complete prenatal check includes the measurement of mother's weight, height, blood pressure, fundal height, fetal heartbeat, blood test, external pelvic examination, and internal examination. The prenatal quality index is constructed using principal component analysis of the following items: all items for a complete prenatal check-up, the number of tetanus toxoid vaccinations received, information on pregnancy complications, and an indicator for receiving any iron pills. Sub-district randomization is used as instrument for individual program participation. Individual characteristics include education, husband's education, log per capita expenditure, age, and indicators for home ownership and land ownership. Midwives' prenatal quality is only available in the 2009 follow-up survey. Robust standard errors in parentheses, clustered at sub-district level, * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 9: Changes in Midwife Characteristics

	Migration of experienced midwives	Higher education (3-year diploma)	Less Than 3 years of experience	Years of experience	Number of Services
	(1)	(2)	(3)	(4)	(5)
Treatment	0.012	0.139***	0.053**	-0.967*	-0.498*
	(0.023)	(0.036)	(0.025)	(0.559)	(0.272)
Observations	942	955	958	958	958
R-squared	0.151	0.252	0.112	0.190	0.254
Mean Dependent	0.002	0.302	0.085	11.88	13.79
Variable	(0.043)	(0.459)	(0.279)	(7.386)	(2.617)

Notes: District fixed effects included in all specifications. Sample restricted to non-panel midwives only. Migration of experienced midwives is an indicator that takes the value one when midwives with at least 3 years of experience have spent less than 2 years at the sub-district clinic. Robust standard errors in parentheses, clustered at sub-district level, * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 10: Heterogeneous Changes Baseline per Capita Expenditure Level

	P(Use of Midwife)	Fees Paid to Midwives	Quality Index	Birth Weight
	(1)	(2)	(3)	(4)
CCT	0.0896*	49,091	-0.136	-61.15
	(0.0479)	(65,247)	(0.109)	(77.46)
Quintile 1	-0.0008	-18,855	-0.093**	-44.41*
(Poorest)	(0.0189)	(12,336)	(0.045)	(26.70)
Quintile 2	-0.0338	-44,591***	-0.047	21.24
	(0.0214)	(12,355)	(0.043)	(29.45)
Quintile 3	-0.0056	-26,677***	0.085*	-28.92
	(0.0205)	(10,029)	(0.045)	(30.15)
Quintile 4	0.0051	-25,477**	0.046	-13.36
	(0.0204)	(11,619)	(0.044)	(28.91)
Quintile 1	-0.0015	1,066	0.309**	59.42
x CCT	(0.0517)	(66,809)	(0.125)	(82.52)
Quintile 2	0.0118	67,335	0.254**	93.55
x CCT	(0.0764)	(84,414)	(0.129)	(140.9)
Quintile 3	-0.0225	-35,283	0.210*	162.0
x CCT	(0.0686)	(71,959)	(0.119)	(105.8)
Quintile 4	0.0209	34,636	0.089	75.63
x CCT	(0.0694)	(81,102)	(0.136)	(110.3)
Observations	6,628	6,546	8,302	4,987
R-squared	0.223	0.150	0.161	0.078

Notes: Standard errors in parentheses * $p<0.1$, ** $p<0.05$, *** $p<0.01$. District FE included, all standard errors clustered at sub-district level. Quality index constructed using principal component analysis based on list of prenatal care items. Sample restricted to women with midwife assistance for delivery. TT estimates use the sub-district randomization as instrument. Individual characteristics include education, husband's education, log per capita expenditure, age, and indicators for home and land ownership.

Table 11: Heterogeneous Changes in Health Outcomes by Baseline Quality

	Infant mortality	Birth weight (grams)	Low birth weight (<2,500 gr.)
	(1)	(2)	(3)
Panel A. Reduced Form Effect of CCT Program in Sub-District			
Treatment	-0.0001	-60.86	0.065***
	(0.0051)	(37.77)	(0.018)
High baseline	-0.0032	-8.335	0.008
quality	(0.0029)	(11.83)	-0.006
Treatment x High	0.0104	118.9**	-0.088***
baseline quality	(0.0081)	(50.60)	(0.024)
Observations	8,302	4,987	4,987
R-Squared	0.044	0.087	0.062
Baseline Dependent	0.011	3,180.76	0.077
Variable	(0.106)	(552.39)	(0.266)
Panel B. IV Effect of Program Participation			
Ever Received CCT	-0.0003	-134.9	0.145***
	(0.0099)	(82.04)	(0.039)
High baseline	-0.0031	-5.598	0.0055
quality	(0.0028)	(11.65)	(0.0061)
Treatment x High	0.0104	115.2**	-0.0852***
baseline quality	(0.0075)	(48.94)	(0.0235)
Observations	8,302	4,987	4,987
R-Squared	0.044	0.084	0.049
Baseline Mean among	0.031	3200.00	0.061
Program Participants	(0.173)	(628.78)	(0.241)

Notes: District FE included, all standard errors clustered at sub-district level. High quality is an indicator that takes the value one for sub-districts with above median prenatal quality at baseline. The prenatal quality index is constructed using principal component analysis, which includes the measurement of mother's weight, height, blood pressure, fundal height, fetal heartbeat, blood test, internal examination, the number of tetanus toxoid vaccinations received, information on pregnancy complications, and an indicator for receiving any iron pills. Sub-district randomization is used as instrument for individual program participation. Infant mortality includes stillbirths and deaths up to 12 months. Individual characteristics include education, husband's education, log per capita expenditure, age, and indicators for home and land ownership. Standard errors in parentheses * p<0.1, **p<0.05, ***p<0.01.

Table 12: Program Effects in Java and off-Java

Panel A. Household Report

	Fees paid to Midwife		Quality	
	Java	off-Java	Java	off-Java
	(1)	(2)	(3)	(4)
Treatment	54,252***	67,390***	0.080**	0.106
	(18,268)	(20,874)	(0.037)	(0.073)
Observations	4,444	2,102	5,659	2,643
Baseline Mean	212,865	85,719	0.078	-0.194
	(403,506)	(211,521)	(0.941)	(1.110)

Panel B. Midwife Report

	Private fees		Private Quality	
	Java	off-Java	Java	off-Java
	(1)	(2)	(3)	(4)
Treatment	18,809***	54,823***	-0.027	-0.002
	(4,235)	(11,313)	(0.039)	(0.165)
Observations	1,975	588	1,029	367
Baseline Mean	343,578	146,156	0.258	-0.729
	(91,857)	(156,241)	(0.664)	(1.361)

Notes: Reduced form estimates. District FE included, all standard errors clustered at sub-district level. The prenatal quality index is constructed using principal component analysis, which includes the measurement of mother's weight, height, blood pressure, fundal height, fetal heartbeat, blood test, internal examination, the number of tetanus toxoid vaccinations received, information on pregnancy complications, and an indicator for receiving any iron pills. Individual characteristics include education, husband's education, log per capita expenditure, age, and indicators for home and land ownership. Standard errors in parentheses * $p<0.1$, ** $p<0.05$, *** $p<0.01$.