Pre-Analysis plan of an experimental evaluation of midwives retention incentives in Nigeria

Marcus Holmlund
Pedro Rosa Dias
Marcos Vera-Hernández

6 February 2015

1. Introduction

This document outlines our plan for analysing the relative effectiveness of alternative retention incentives for Midwives in Nigeria. We begin by providing a summarised description of the intervention. We then describe our plan of analysis, specifying the main regression models (including respective covariates) used to estimate the main impact of incentives. These include the analysis of treatment effect heterogeneity and, crucially, the understanding of the individual behavioural responses that determine the effectiveness of retention incentives. At the time this document is written, the baseline data is available but the endline data is not yet available.

2. Intervention and context

This document describes our analysis of the effects of retention incentives, provided to midwives employed by the Subsidy Reinvestment and Empowerment Programme (SURE-P) Maternal and Child Health (henceforth, SURE-P). SURE-P is a nationwide programme, implemented by the Federal Ministry of Health. It has the overall objective of increasing the use of maternal and neonatal healthcare, which is believed to be key for also cutting maternal and child mortality rates. SURE-P includes both supply-side and demand-components.

Its supply side aims at improving the availability and quality of maternal and child healthcare. Thus, it provides SURE-P Primary Health Centres (PHCs) with basic infrastructure upgrades, key maternal and child health commodities, and health professionals: qualified midwives and community health extension workers (CHEWs). These components constitute the *basic improvement package*, and all SURE-P PHC facilities benefit from it. In addition, in some SURE-P PHCs, midwives receive retention incentives, described in detail in Section 2.1. This pre-analysis plan only describes our analysis of the effect of the midwives retention incentives.

Still on the supply side, in addition to the basic improvement package, SURE-P plans to include a community-based stock monitoring scheme. This is described in Section 2.2 below but has not been yet rolled out.

Also in addition to the basic improvement package, but on the demand side, SURE-P seeks to boost the demand antenatal care and institutional deliveries by providing pregnant women a conditional cash transfer scheme. To date this has only been piloted in a small number of facilities (as described in 2.2, below)

SURE-P runs from 2012-2015 in 1,000 public primary healthcare centres (PHCs) spread across Nigeria's 36 states and the Federal Capital Territory. PHCs are grouped into clusters of four, and each cluster is linked to a referral facility (hospital).

2.1 Midwife retention incentives

The objective of this intervention is to reduce attrition amongst midwives by providing them with monetary and non-monetary retention incentives. Attrition of health staff is a well-known challenge facing health systems (see for example Hatt et al. 2007, on the expansion of midwifery services in Indonesia). In Nigeria, high levels of attrition were experienced by a previous large maternal and child health programme – the Midwives Service Scheme.[1]

[1] More generally, the crucial importance of promoting health workers retention is well documented in Deserranno (2014) and references therein.

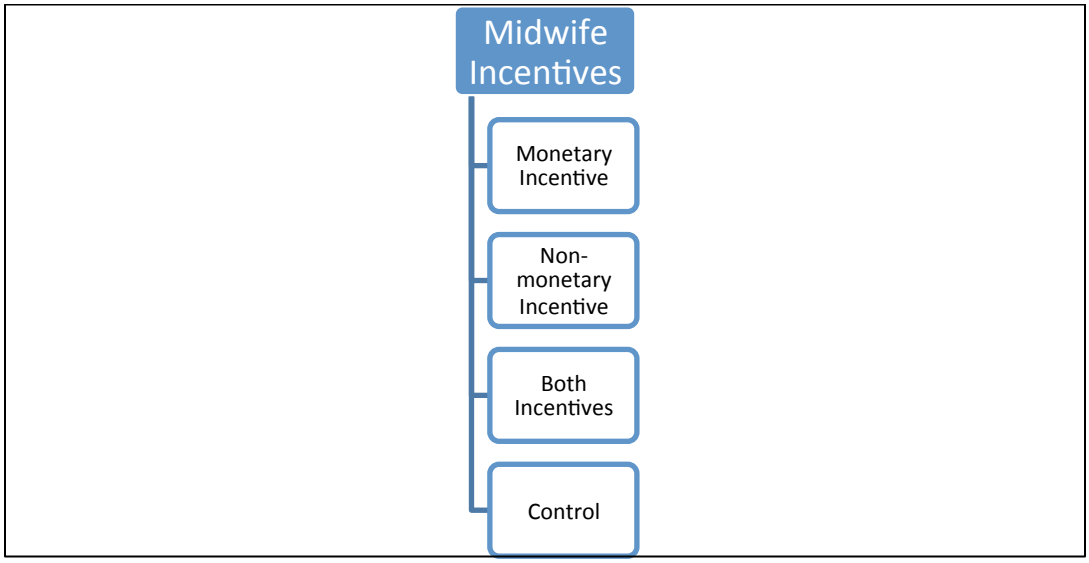
All midwives who are part of this intervention are employed by SURE-P in SURE-P PHCs. In order to promote retention, midwives that remain at their duty health facility for increments of three months are eligible to receive a package of retention incentives. The incentives are cumulative, i.e. a midwife cannot qualify for the second three months' incentive if she has not qualified for the first three months', and so on. Our analysis compares the relative effectiveness of three alternative packages of incentives:

a) Monetary incentives: quarterly payments of N30,000, over and above midwives' salaries. For example, suppose a midwife signs the retention incentive contract at the beginning of a given month (let us call it month 1). This midwife only receives the first quarterly payment of N30,000 if she remains working in SURE-P for three consecutive months (i.e. until the end of month 3). If she leaves before this date she does not receive the incentive. This makes it different to a monthly salary increase of N10,000 because the amount forgone by leaving earlier will be higher with this retention incentive than with the monthly salary increase. The same applies to incentives paid on month 6, 9 and 12.
b) Non-monetary incentives: quarterly provision of non-monetary incentives, consisting of a special SURE-P uniform, a customised wall calendar, a SURE-P MCH wall clock, and a second type of special uniform.
c) Both types of incentives: quarterly payments of N30,000 in tandem with the provision of non-monetary incentives described in b).

The implementation of SURE-P is divided in two phases. Phase 1 includes the first 500 PHCs to begin implementing the program, while Phase 2 expands this to a further 500 PHCs. The implementation of SURE-P in both Phase 1 and Phase 2 PHCs continues through the end of 2015. In Phase 1, the midwife incentives intervention began in December 2013, with midwives eligible to receive quarterly incentives based on ongoing attendance through the end of February 2014, May 2014, August 2014, and November 2014. The second year of the midwives incentives scheme, including midwives working in both Phase 1 and Phase 2 PHCs, began in January 2015 with midwives eligible to receive quarterly incentives based on ongoing attendance through March 2015, June 2015, September 2015, and December 2015.

In order to evaluate the relative effectiveness of this intervention, midwives are randomly assigned[2] to three treatment groups and one control group as shown in Figure 1.

***Figure 1**: Midwife incentives – treatment and control arms*



2.2 Additional components of SURE-P MCH

In addition to the provision of retention incentives SURE-P MCH includes other supply and demand side components beyond the basic improvement package. While this pre-analysis plan concerns the evaluation of the effect of incentives, our empirical strategy takes into account the implementation of two of these components, which will be evaluated in separate randomised control trials. It should be noted that these additional components ran (or will be run) at a much smaller scale than the midwife retention incentive.

2.2.1 Community-based stock monitoring component

Previous maternal health interventions in Nigeria were hampered by the high prevalence of drug stockouts at the PHC level. Thus, SURE-P plans to include a Community-based Stock Monitoring (CBSM) programme. The CBSM programme seeks to use community feedback to monitor drug stock-outs in PHCs, and then harness the power of Ward Development Committees (WDC), and the communities they represent, to advocate for a more consistent

[2] A detailed account of the randomisation procedure is given below.

supply of essential drugs. This program was piloted in a eight PHCs in four states in 2014; a scale-up is planned, but the date is uncertain at the time of writing this document. To isolate the effect of the midwife incentive from the impact of the CBSM programme, our empirical strategy randomises the CBSM across the arms of the retention incentives component, as shown in Section 3.

2.2.3 Conditional cash transfer component

Another potential obstacle to the success of SURE-P is the typically low demand for antenatal care and institutional deliveries in Nigeria. In order to promote demand for maternal healthcare, SURE-P includes a conditional cash transfer (CCT). Pregnant women in areas eligible for the CCT are given monetary transfers conditional on their attending antenatal care, delivering with a skilled birth attendant, and attending postnatal care (including child immunization). This payment is provided in 4 tranches: N1000 upon registration in a PHC, N1000 after completing the set of standard antenatal visits, N2000 upon institutional delivery and N1000 after zero-dose immunisation is given to the new-born baby. To date, the CCT component is active in 18 (purposefully selected) states. Within these, it has been implemented in 9 clusters (randomly chosen amongst the clusters of 9 purposefully selected states) and other 9 purposefully selected clusters, which are in other 9 different states. There are plans to expand this component to a further 18 randomly selected clusters.

2.2.3 Isolating the effects of midwife retention incentives from the CCT and CBSM interventions

In order to isolate the effect of midwife incentives from the impact of these other experimentally evaluated components of SURE-P, our empirical strategy randomises the CBSM and CCT components across the arms of the incentives component, as shown in Section 3. Note that the basic improvement package took place in all SURE-P PHCs, hence there is no variation in exposure to this package across either of the interventions discussed in section 2.2.

3. Methods

3.1 Randomization

The unit of randomization is the cluster, which is comprised of SURE-P PHCs that share the same referral hospital. There are four PHCs in each cluster. The random process of allocating clusters to treatment arms was carried out in July 2013 for Phase 1 facilities, and November 2014 for Phase 2 facilities.

For Phase 1 facilities, the randomization procedure was set up to isolate the likely effect of Quality Enhancement intervention which was going to run in some facilities of the clusters of our study area. To prevent contamination from this intervention, we formed three large strata: Stratum 1 included 24 clusters containing PHCs which benefited from the Quality Enhancement Program run by the Government of Nigeria, which is also subject to an experimental impact evaluation (this program was randomized at the PHC rather than at the cluster level). The remaining 101 clusters were randomly divided in a stratum 2 of 96 clusters and a stratum 3 of 5 clusters. Both in strata 1 and strata 2, clusters were ordered according to the median number of recently hired midwives in the cluster (as reported by administrative records)[3], and groups (randomization blocks) were formed with clusters whose ranks were 1-8, 9-16, 17-24, etc. Within each randomization block, clusters were randomized to (1) No Incentive & No CBSM, (2) Monetary Incentive only & No CBSM, (3) Non-Monetary Incentive Only & No CBSM, (4) Both Incentives & No CBSM, (5) No Incentive & CBSM, (6) Monetary Incentive only & CBSM, (7) Non-Monetary Incentive Only & CBSM, and (8) Both Incentives & CBSM. The clusters within the stratum containing 5 clusters were randomized into to (1) No Incentive & No CBSM, (2) Monetary Incentive only & No CBSM, (3) Non-Monetary Incentive Only & No CBSM, (4) Both Incentives & No CBSM, and (8) Both Incentives & CBSM. At the time of writing this document, the CBSM intervention had not been rolled out, so it might be that the distinction between CBSM and non-CBSM becomes irrelevant.

[3] Ties were broken by ordering according to the median number of midwives who had been hired in the past.

There were no planned interventions for Phase 2 facilities beyond the SURE-P basic improvement package (such as the Quality Enhancement Program or CBSM) at the time of randomization. Hence, we decided to stratify according to groups of States in accordance with the implementation of data collection, which was implemented by geopolitical zone.[4]. Data collection was implemented in three waves, with each wave covering two geopolitical zones: wave 1 (North East and North West) comprised 51 clusters, wave 2 (South West and North Central) comprised 39 clusters, and wave 3 (South East and South South) comprised 36 clusters.[5] The total of 126 clusters in Phase 2 of SURE-P were divided in four strata. Stratum 4 included 3 randomly selected clusters from wave 1 and 3 randomly selected clusters from wave 2. Stratum 5 included the remaining 48 clusters from wave 1. Stratum 6 included 36 clusters from wave 2. And Stratum 7 included the 36 clusters in wave 3.

Within Stratum 4, two clusters were randomly allocated to the Monetary Incentive only arm, two other clusters were allocated to the Non-Monetary only arm, 1 cluster to the Both Incentives arm, and 1 to the No Incentive arm. Within each stratum of strata 5, 6, and 7: clusters were ordered according to the average number of recently hired midwives in the cluster (as reported by administrative records), and groups (randomization blocks) were formed with clusters whose ranks were 1-4, 5-8, 9-12, etc. Within each randomization block, clusters were randomized to (1) No Incentive, (2) Monetary Incentive only, (3) Non-Monetary Incentive Only, (4) Both Incentives. For Phase 2 facilities, we do not consider the CBSM because it is not planned for this set of facilities.

Regarding the randomization of the CCT program, we were asked to randomly choose 9 clusters to receive the CCT program amongst the Phase 1 clusters in the states of Abia, Benue, Enugu, Ekiti, Delta, Kogi, Oyo, Sokoto and Taraba; our choice was restricted to one cluster per State, and that there was at least one cluster of each treatment arm of the retention incentive & CBSM intervention CCT program was equally distributed across the

[4] Nigeria is divided into six geopolitical zones: North West, North Central, North East, South East, South South, and South West.

[5] Team 1 collected data in Adamawa, Bauchi, Borno, Yobe, Gombe, Taraba, Kano, Kaduna, Katsina, Zamfara, Jigawa, Sokoto, Kebbi; Team 2 collected data in Ekiti, Ogun, Osun, Oyo, Lagos, Ondo, Niger, Kogi, Kwara, FCT, Nassarawa, Plateau, Benue; and Team 3 collected data in Edo, Delta, Bayelsa, Rivers, Akwa-Ibom, Cross-River, Anambra, Enugu, Abia, Ebonyi, Imo.

retention incentive & CBSM interventions. In October 2014, we repeated this process in for the following 9 states: Anambra, Bauchi, Bayelsa, Ebonyi, FCT, Kaduna, Niger, Ogun, Zamfara.[6]

In the analysis, we will control for the CCT implementation.

3.2 Data Collection

Data to be used in this analysis comes from three rounds of purposive surveys. The first of these took place between September and November 2013, and covered the first 500 PHCs selected for participation in SURE-P and their surrounding catchment areas. These 500 PHCs form Phase 1 of the program, and so we refer to this first survey round as the Phase 1 Baseline. The second survey round took place between December 2014 and February 2015. This round includes a follow-up survey in the first 500 PHCs and their catchment areas (Phase 1 First Follow-up) and extends the data collection to an additional 500 PHCs and their catchment area, which form Phase 2 of the program (Phase 2 Baseline). A third round of data collection (Phase 1 Second Follow-up and Phase 2 First Follow-up) is planned for late 2015.

3.2.1 First round of data collection: Phase 1 Baseline, 9 September – 27 November 2013

Baseline data was collected in the first 500 PHCs selected for participation in SURE-P and their surrounding catchment areas. The area covered includes Nigeria's 36 states and the Federal Capital Territory. Dedicated survey questionnaires were administered targeting the following groups: (1) all midwives working in SURE-P PHCs; (2) the officers-in-charge at each SURE-P PHC; (3) a representative of the local ward development committee; and (4) women that had given birth in the three months preceding the survey.

[6] In 2013, the SURE-P program had purposefully selected one cluster from each State for the CCT program. Obviously, these clusters were excluded before doing the randomization.

All interviews were conducted face-to-face by trained enumerators. Data was captured electronically using *Open Data Kit*. Sampling and content for each instrument is briefly described below.

A. Midwife survey

At the time of the survey, 1,285 midwife posts were filled, and each of these midwives were interviewed. The midwife survey solicited information through standard survey questions and through decision activities (or behavioral games).

The standard survey questions covered the following topics:

1. Basic personal information
2. Education and experience, including exposure to rural settings
3. Preferred job attributes
4. Work conditions and relationship with colleagues and family
5. Relationship with and support from the community
6. Maslach Burnout Inventory to measure emotional exhaustion, depersonalization, and personal accomplishment
7. Household income and assets
8. Non-experimental measures of intrinsic motivation and pro-sociality
9. Knowledge and practice of antenatal care
10. Future expectations related to work and colleague and community relations following receipt by each midwife of her incentives contract or, for the control group, a placebo letter and thank you message. The delivery of these contracts was integrated into the baseline data collection.

The following decision activities (behavioral games) were included in the survey. Midwives were incentivized to reveal their true preferences through the use of actual money in these activities, which was paid into their bank account following their survey.

1. Two versions of a basic altruism decision activity, where each midwife is asked to divide a fixed sum of real money between herself and the Nigerian Red Cross

Society. In the first "private" version, the midwife's allocation decision is not revealed to her colleagues. In the second "public" version, the midwife's choice is revealed to her colleagues working at the same PHC. These terms are explained to midwives prior to each decision being made. In both private and public versions, each midwife allocated NGN 2000 (approximately USD 12) in NGN 200 increments to either themselves or to the Nigerian Red Cross Society.

2. A time preference activity, designed along the lines of Andreoni and Sprenger (2012) and Giné et al (2014), where each midwife is asked to allocate NGN 6000 in NGN 500 increments between two time periods: "now" (payment into their bank account within 1 week) and "future" (payment into their bank account with a longer delay, either 30 days or 60 days). Money allocated to the future earns interest at a rate of 5%, 10%, 25%, 50%, and 100%. Each midwife makes a total of ten allocation decisions, one at each interest rate for both the near (30 days) and far (60 days) future. Once all ten allocation decisions have been made, the midwife randomly selects a number between one and ten to determine her actual payout, but this final step is only of interest insofar as it determines the actual amount due to the midwife.

3. A risk preferences activity following an approach close to that in Liu (2013). Each midwife is given an initial endowment of NGN 525 and asked to choose between two options ("A" and "B") with different risk-return profiles across eighteen decisions split into three series. Each option is a gamble, with risk held constant across option A and option B within each series, but with a greater maximum return (or loss) under option B. Within each series, the ratio of the expected return under option B to that of option A increases with each subsequent decision, and the objective is to identify the point at which the midwife switches from the relatively safe with low maximum return option A to the relatively risky with high maximum return option B. Once the midwife has selected either option A or option B for each decision, she randomly selects a number between one and eighteen which determines the gamble she will play for real money. The gamble itself involves randomly choosing a number between one and ten, with each number corresponding to a pre-

specified payout, but this is of interest only insofar as it determines the amount due to the midwife.

4. A social norms decision activity, carried out after the midwife has received her incentives contract (or, in the case of the control group, her placebo letter and thank you message) designed to elicit the extent to which midwives can anticipate the responses of their colleagues in the same treatment group regarding how socially appropriate or inappropriate it is for a hypothetical midwife (who has just received the same incentives contract or placebo letter) to quit her job two months, five months, eight months, eleven months, fourteen months, seventeen months, twenty months, or twenty-four months after joining the program. In line with the approach in Krupka and Weber (2013), midwives are told that, once all surveys have been completed, one of the time periods will be selected at random and that NGN 1000 will be paid into their bank account if her responses matches the modal response of all midwives who have the same retention bonus contract as her, i.e. if she was able to correctly anticipate the response of the majority of her peers.

B. Primary Healthcare Facility Officer-in-Charge survey

This survey targeted the officers-in-charge at the 500 initial SURE-P PHCs. A response rate of 95% (476 PHCs) was achieved. This survey consisted of standard survey questions covering the following topics:

1. Basic information about the facility
2. Facility administration, management, and human resources, including absenteeism and performance management
3. Maslach Burnout Inventory to measure emotional exhaustion, depersonalization, and personal accomplishment, administered to the respondent
4. Record keeping
5. Health services provision, including community outreach
6. User fees
7. Equipment availability

8. Drug storage and availability
9. Compliance with national protocols

C. Ward Development Committee survey

This survey targeted the chairman of each Ward Development Committee (WDC) whose geographic area of responsibility includes a SURE-P PHC. Wards are an administrative unit, and WDCs are charged with supervising and supporting development within their ward, including PHC services. The survey targeted each of the 500 WDCs linked to each SURE-P PHC, and a response rate of 95% (473) WDCs was achieved. The survey consisted of standard survey questions covering the following topics:

1. General information about the WDC
2. Community characteristics and access to basic services
3. Social capital and community empowerment
4. Positive and negative external shocks experienced by the community in the past two years
5. Direct observation of the community

D. Household survey targeting women that had given birth in the three months preceding the survey

This survey was administered to households including a woman that had given birth in the three months preceding the survey, as a proxy for the population eligible to receive SURE-P services. Whenever possible, much of the questionnaire was administered directly to this woman. 20 households were randomly sampled within the catchment area of each SURE-P PHC for a target sample size of 2,500 households. A response rate of 95% (2,384 households) was achieved. The survey consisted of standard survey questions covering the following topics:

1. Household roster and basic demographic information
2. Household income and consumption

3. General health services utilization and payment, including expectations on service availability and quality
4. Antenatal care service utilization and reproductive health, including child mortality
5. Labor and delivery
6. Postpartum care and breastfeeding
7. Postnatal depression, measured through the Edinburg Postnatal Depression Scale
8. Maternal attitudes and knowledge related to antenatal and infant care
9. Exposure to media and mobile phones
10. Interaction with village leaders and community relations
11. Housing characteristics, amenities, and assets
12. Adverse events experienced by the household

3.2.2 Second round of data collection: Phase 1 first follow-up and Phase 2 baseline, 8 December 2014 – 6 February 2015

A first round of follow-up data was collected in the 500 SURE-P Phase 1 PHCs and their surrounding catchment areas. A baseline, using the same instruments, was also carried out in the 500 SURE-P Phase 2 PHCs and their catchment areas. Interviews were conducted face-to-face by trained enumerators, and data was captured electronically using Survey CTO.

The same four target groups from the baseline were surveyed in both Phase 1 and Phase 2 areas. The Phase 1 baseline survey tools were updated with some modification of question wording, questionnaire format, and addition/deletion of a small number of questions.
While the same tools were used for the Primary Healthcare Facility, Ward Development Committee, and Household Survey across the full sample in both Phase 1 and Phase 2 areas, the midwife survey was adapted for four different groups of midwives:

1. *Midwives included in the Phase 1 Baseline who were still working at their PHC*

Questions were added to (i) verify the delivery of the appropriate incentives throughout the year depending on the treatment or control status of each midwife; (ii) assess each midwife's

perceptions of her control over factors affecting her personally and her environment; and (iii) assess each midwife's trust in various government and non-government entities.

2. *Midwives included in the Phase 1 Baseline who had since left their jobs*

Questions were added on the reasons why these midwife left their jobs during the year following the baseline survey. They were also asked the additional questions asked of Phase 1 Baseline midwives that were still working.

3. *Midwives working in Phase 1 PHCs that were not yet working at the time of the Phase 1 Baseline*

These midwives were administered the baseline survey, with additional questions on locus of control (including questions on trust).

4. *Midwives working in Phase 2 PHCs*

These midwives were administered the baseline survey, with additional questions on locus of control (including questions on trust).

3.2.2 Third round of data collection: Phase 1 second follow-up and Phase 2 first follow-up, end 2015

The final round of data collection will comprise a second follow-up survey in Phase 1 SURE-P PHCs and their catchment areas, and a first follow-up survey in Phase 2 SURE-P PHCs and their catchment areas. The round two instruments will be used, with anticipated minor modifications.

4. Analysis

4.1 Estimation of treatment effects: model specification

The main outcome variable of the study is the retention of midwives. We will define two main dependent variables: y_{ifc} and m_{ifc}. y_{ifc} takes value 1 if the midwife i, working at baseline in PHC f from cluster c is still working in the SURE-P PHC after G months of the retention

bonus contract being delivered to her. m_{ifc} is the number of months that midwife *i,* working at PHC *f* from cluster *c* worked in PHC *f* since the contract was delivered to her; this may be right censored.

For each G that we define below, we will estimate the following linear probability model[7] [8]:

$$Y_{ifc}(G) = \alpha + \beta_1 T\_M_c + \beta_2 T\_N_c + \beta_3 T\_BOTH_c + X_{ifc}\gamma_1 + W_{fc}\gamma_2 + Z_c\gamma_3 + u_{ic}, \quad (1)$$

where T_M_c takes value 1 if midwives in cluster *c* were provided with a retention contract that included only a monetary incentive, and 0 otherwise; T_N_c takes value 1 if midwives in cluster *c* were provided with a retention contract that included only a non-monetary bonus, and 0 otherwise, and T_BOTH_c takes value 1 if midwives in cluster c were provided with a retention contract that included both a monetary and non-monetary bonus, 0 otherwise. The vectors $X_{ifc,}$ W_{fc}, Z_c include covariates that are specific to, respectively, midwife *i*, PHC *f* and cluster *c*. The computation of standard errors and p-values will be clustered at the level at which the incentives were randomized (cluster).

Our rationale to include covariates is to improve the precision of the estimates. In most RCTs, including the value of the outcome variable at baseline in the regression will result in substantial improvements in precision. However, because of the nature of our main outcome variable, this strategy is not available to us. Hence, we choose to include a relatively rich set of covariates. Given the lack of research on midwife attrition, it is unclear what set of covariates will yield substantial gains in precision. Thus, we intend to use a reduced and a extended definition of the vectors $X_{ifc,}$ W_{fc}, Z_c. The extended version includes more covariates and will result in a smaller residual variance, although at the expense of reducing the degrees of freedom (this is given by the difference between the number of clusters and the number of covariates). The reduced and extended versions of the covariate vectors are defined below. If the percentage of missing values for a given covariate is less than 10%, we

[7] The parameters of the model depend on G, but we omit this for simplicity.
[8] In some cases it may be clearer to report the estimates of an equivalent regression model to (1), where the treatment arms are defined as having any monetary or any non-monetary incentive, and T_BOTH is replaced by the interaction term between monetary and any non-monetary incentives.

will replace this with its sample average[9]. If that percentage exceeds 10%, we will drop the covariate.[10]

We will define G to be 9 months. This is because the contract was delivered to midwives in Phase 2 facilities between December 2014 and February 2015, and the follow-up survey for those facilities is planned to take place between October 2015 and December 2015. Hence, we can only guarantee to follow-up Phase 2 midwives for 9 months. Depending on the specific timing of data collection, there might be a small set of facilities for which we cannot observe midwives for 9 months. For that set of facilities, we will use the largest possible G.

We will also estimate a discrete-choice duration model as in Jenkins (1995). This would allow us to learn about effects of the incentives on longer durations, which can be studied with the facilities from Phase 1 but not with the facilities from Phase 2. We will follow the same specification of covariates as in (1), which we will augment with a hazard function. Through interaction terms between the baseline hazard and the incentive dummies, we will also explore a specification in which the incentives only have an effect after some minimum length of time.

Midwife specific covariates: restricted and extended version of X_{ifc}

The restricted version of X_{ifc} comprises**:**

- Age at baseline
- Number of years that the midwife has been living in the community where she lived at baseline and the square of this number
- Tenure in the PHC
- Whether the midwife lived in a rural area at baseline
- Experience as midwife
- The score of the midwife's Maslach Burnout Inventory at baseline

[9] In the case of binary covariates missing values will be replaced with the value of their most frequent category.
[10] We will not perform this replacement when we report the balance across treatment arms.

In addition, the extended version of X_{ifc} also includes the following covariates:

- Minority religious group (i.e. the midwife's religion is practiced by a minority of people in the catchment area of her PHC).
- Whether the midwife lived during primary school in a place that was more rural than the place she lived at baseline (binary variable)

PHC specific covariates: W_{fc}

- Number of midwives working in the PHC *f* from cluster *c* at baseline
- Threatened or afraid (indicator variable)
- Equipment available and in working condition in the PHC, at baseline.
- Whether the PHC *f* was included in a primary PHC quality improvement program run by the Federal Ministry of Health (randomly assigned to PHCs in a subset of six states and stratified across the arms of this RCT)

Cluster specific covariates: restricted and extended version of Z_c

The restricted version of Z_c includes:

-The percentage of women who achieved primary education grade 6 or higher
-The cluster average asset ownership (question 18) composite index (*a la* Anderson 2008)[11]
-Whether cluster *c* was part of a CCT programme to promote antenatal care and skilled birth attendance (two variables: one binary variable for CCT-treatment clusters which were part of an experimental evaluation, and another binary variables for cluster in which the CCT was implemented non-experimentally)
- Whether the cluster *c* was part of a stock-out intervention (randomly assigned). This intervention has not been yet been rolled out, and it has not been established whether it will be rolled out within the time span of this study.
- Randomization strata (note that this will include whether the cluster entered into the study in Phase 1 or 2
-It is common to include randomization block dummy variables in the specification. However, we are concerned that, given our relatively large number of covariates, this would

[11] We chose this cut-off after tabulating education and maternal and child health care use at baseline.

reduce excessively the available number of degrees of freedom. For that reason, we follow a more parametric approach: we include the interaction between the randomization strata and a quadratic polynomial on the number of recently hired midwives (the variable used to create the randomization blocks)[12]. This seems to us a balanced solution. However, since this differs from the most common approach, for robustness we will also present estimation results including randomization block dummy variables.

In addition, the extended version of Z_c includes:
- State dummies

In addition to the variables above, we might add variables (at the appropriate level) that include information on very large community level shocks or other interventions that might affect the outcomes of interest and that we are not aware of at the time of writing this document.

At the time of writing this document, there is some uncertainty regarding the availability of funds for a third round of data collection (which will include a second follow-up for Phase 1 and the first follow-up for Phase 2). If this is not available, we will carry out the analysis using solely the data from Phase 1 facilities.

In addition, there is the possibility of delays in the payment and delivery of incentives, due to logistic and administrative constraints. If these occur, and are found to be large, we will present two sets of results: one set with only the first period of the experiment (period between baseline and first follow-up of Phase 1 facilities) and another one with all the available data.

Reports from the field suggest that the delivery of non-monetary incentives in two clusters in Ondo state was suspended three months into the treatment due to administrative error. If this is confirmed, we will present impact estimates with and without the Ondo state clusters.

[12] Note Strata 3 and 4 only had few clusters (5 and 6 respectively) so their randomisaiton within these strata did not use the number of recently hired midwives. Hence, there is no need to use the quadratic polynomial for these two strata

5. Hypotheses

Our research design allows us to test a number of hypotheses regarding the causal effect of incentives on the retention of midwives in the programme. These comprise three groups of hypotheses: hypotheses on outcomes; hypotheses on impact heterogeneity; and hypotheses on causal mechanisms.

5.1 Group A: Hypotheses on outcomes (retention rates, absenteeism, other aspects of performance)

A1: Retention incentives (monetary and non-monetary, provided independently or jointly) are likely to have positive average effects on midwives retention rates.

This hypothesis will be tested using regression model (1) as well as its discrete-time duration model versions. The dependent variables will be built as specified in section 4 (main analysis).

A2: Monetary and non-monetary retention incentives are likely to be complements: when provided jointly their positive effect on the retention rate is likely to be larger than the effects of the monetary and non-monetary incentives provided separately.

This hypothesis will be tested using regression model (1) as well as its discrete-time duration version. The dependent variables will be built as we specify in section 4 (main analysis)

A3: Monetary and non-monetary retention incentives (provided independently or jointly) are likely to reduce midwife burnout.

Incentives are likely to partly compensate midwives for the unpleasant and stressful aspects of their job, by making them feel supported and rewarding their effort. This is likely to result in a reduction of the prevalence of *burnout* amongst them.

We will test this hypothesis using regression model (1) and the midwives' Maslach Burnout Inventory score to measure emotional exhaustion, depersonalization, and personal accomplishment.

A4: Monetary and non-monetary retention incentives (provided independently or jointly) are likely to yield improvements in maternal and child healthcare use.

Retention bonuses have to potential to improve maternal and child health since:

(1) Retention incentives increase the number of midwives available in the health care facility (see hypothesis M6 below).

(2) Retention incentives lead midwives to consider that they are supported adequately for their work, and that their employer appreciates the work they do (see hypothesis M7 below). In turn, this leads midwives to exert more effort (see hypothesis A6)

(3) Retention incentives may reduce burnout, thereby leading midwives to exert effort (see hypothesis A3)

(4) Retention incentives will change the composition (average level of intrinsic and pro-social motivation) of midwives who stay working in the PHC. This is because the incentives' effect on retention will be heterogeneous according to midwives intrinsic and pro-social motivation (see Group B hypotheses). This might have an effect on midwives effort because more intrinsically or pro-socially motivated midwives :

   a. might be more likely to exert more effort
   b. might be more likely to to cooperate with other workers (midwives or not)
   c. might be more skilled or knowledgeable[13]

Note that because of (2), (3), and (4), different type of retention incentives might affect outcomes differently even if their effect of retention rates is the same.

[13] We will be able to assess this correlation using data collected on academic grades received during their midwifery course, as well as their responses to practical midwifery questions.

To test this hypothesis, we will estimate the following model:

$$H_{ifc} = \alpha' + \beta'_1 T\_M_c + \beta'_2 T\_N_c + \beta'_3 T\_BOTH_c + Q_{ifc}\gamma'_1 + W_{fc}\gamma'_2 + Z_c\gamma'_3 + v_{ic}, \quad (2)$$

which follows the same specification as (1) except that the sub-index *i* refers to a child or mother, H_{ifc} measures the health of a child or mother of household *i* living in the catchment area of PHC *f* in cluster *c*; Q_{ifc} comprises age, gender (except for women outcomes), a binary variable of whether the mother has completed primary education grade 6 or higher and an index (*a la* Anderson 2008*)* of asset ownership constructed using question 18, $\mathbf{Z_c}$ includes the average value of $\mathbf{H_{ifc}}$ across households of cluster *c* interviewed at baseline.

We will test this hypothesis using (2) and specify H_{ifc} as a composite index (*a la* Anderson 2008), which includes the following variables: dummy variables: a set of questions that relate to the quality of antenatal care (question 7.8 in the follow-up questionnaire: urine analysis, blood analysis…), skilled birth attendance, institutional delivery, whether the baby was bathed immediately after birth (reversed), whether the first thing that the baby drank was the mother's milk, baby only drank mother's milk after birth, whether mother's health was checked by a health care professional after birth, whether the baby's health was checked by a health care professional within the first week after birth, whether the baby has been vaccinated against Polio, whether the baby has been vaccinated against TB.[14]

A5: Monetary and non-monetary retention incentives (independently or jointly provided) are likely to yield improvements in maternal and child health.

This hypothesis follows the same rationale as A4, and it will be tested in the same way. We will use two different outcome variables:

[14] To provide the reader with information of what components are more likely to be driving any change in the index, we will also report the results of each component separately.

-A dummy variable that measures if the mother scores 10 or higher in the Edinburgh maternal depression index

-neonatal mortality, which will be measured following an extension of the sisterhood method that is normally used to measure maternal mortality. We designed this extension to overcome the statistical problem caused by the relative infrequency of neonatal mortality. Since this method has not been used to measure neonatal mortality before, and we are unsure about the quality of the data. To test this hypothesis, we plan to run a regression in which each observation is the woman that we interview or a sister of her who has recently given birth. In this case, we will not be able to use Q_{ifc} (2), but instead we will augment Z_c with the average in cluster *c* of the wealth index, and the proportion of women in cluster *c* achieving primary education grade 6 or above.

A6: Retention incentives are likely to increase midwives' effort

Reasons for this have been outlined under hypothesis A4. To test this hypothesis, we will regression (2) with the following dependent variable:

- A composite index (*a la* Anderson 2008) which will include a set of questions that relate to the quality of antenatal care (question 7.8 in the follow-up questionnaire: urine analysis, blood analysis…). This is a good indicator of the quality of the work that midwives carry out because attendance rate to antenatal care at baseline was very high. Hence, variations in this indicator are more likely to be driven by midwife's effort than by local women's demand for health care.

A7: Incentives may increase conflict between incentivised midwives and other health workers that so not receive incentives, such as nurses, community health extension workers (CHEWs) and doctors. It may also negatively affect their performance.

Recent evidence from behavioural experiments (see Fehr et al, 2009) indicates that individual performance can be heavily affected by workers' perception of fairness regarding the reward of effort at the workplace. Resentful demoralisation in face of perceived unfairness may lead individuals to underperform and / or increase workplace conflict. Thus, incentives to

midwives may reduce the performance of non-incentivised health workers (nurses, CHEWs) as well as lead to conflict between them and the incentivised midwives.

To test this hypothesis, we will use the questions 5.2.19 and 5.2.20 of the first follow-up midwife questionnaire which asks whether cooperation with nurses and CHEWs got easier or worse since we interviewed the midwife at baseline. We will use a model like (1) but we will augment the covariates with the question 12.2 of the baseline questionnaire (how do you rate your relationship with nurses). Our data contains very little information on the performance of workers other than midwives, but we will use the information on absenteeism by nurses and CHEWs, which is reported in the facility survey. When testing this, we will control for the average level of absenteeism in the facility at baseline.

A8: Incentives (monetary and non-monetary) are likely to shift midwives' locus of control towards an internal one (and away from an external one).

Measures of locus of control (LOC) elicit individuals' beliefs about the extent to which the events that affect them are under their control (Rotter 1966). Individuals with external LOC attribute such events to external factors (e.g., fate, luck, other people), thereby considering that these are beyond their control. Conversely, individuals with internal LOC generally believe that the events that affect them are largely the result of their own actions (i.e. are within their control. Although LOC is more malleable during childhood and adolescence recent research (see Powdthavee et al, 2013 and references therein) indicates that this can change substantially throughout the life cycle as a response to stimuli. Midwives incentives reward individual effort and are therefore likely to increase their perception that outcomes are within their control. We measure LOC as standard in the literature: midwives are presented with a series of pairs of opposing statements (e.g. a) many of the unhappy things in people's lives are partly due to bad luck; b) People's misfortunes result from the mistakes they make) and asked which of the two is closer to their feelings (question 17.2 – Phase 1 follow up Phase 2 baseline survey). We will test this hypotheses using regression model (1) with locus of control as dependent variable.

A.9 Incentives (monetary or non-monetary, independently or jointly provided) are likely to increase midwives' trust in institutions and other people.

In addition, incentive contracts are clear, transparent and respected by the Nigerian authorities; this can enhance midwives' their ability to trust other people and institutions. We measure trust in other people by asking midwives about the extent to which they trust other people (question 17.3.1 – Phase 1 follow up Phase 2 baseline survey). In addition, we measure midwives' trust in institutions by asking them about the extent to which they have confidence in a series of institutions (question 17.3.1 – Phase 1 follow up Phase 2 baseline survey). We will test this hypotheses using regression model (1).

A.10 Incentives (monetary and non-monetary, independently or jointly provided) are likely to improve midwives' levels of life satisfaction and happiness

Incentives reward midwives effort and are likely to make them feel that their work is appreciated. We hypothesise that this has the potential to improve their present happiness and also their life satisfaction. We follow the literature on subjective well-being and measure life satisfaction by asking midwives to rate their life satisfaction on a numerical scale (supported by visual representation) – question 1.c.1 Phase 1 follow up Phase 2 baseline survey. Analogously, we follow the literature and measure happiness by asking midwives to rate their happiness at that moment in time (question 1.c.2) and whether they are happier while working for SURE-P than before (question 1.c.3). We will test this hypotheses using regression model (1).

5.3.2 Group B: Hypotheses on impact heterogeneity

To test the hypotheses below, we will use extensions of model (1) and the duration model. In these extensions, we will include the heterogeneity variable in the regression, as well as its interaction with the treatment arms.

H1: Monetary retention incentives improve the retention of extrinsically motivated midwives more than the retention of intrinsically motivated midwives, which could even be reduced, due to monetary incentives.

To measure the extent of intrinsic/extrinsic motivation, we will use whether interaction with patients and applying knowledge acquired in midwifery school are important job attributes

vis a vis job stability, salary and future career opportunities.[15] We will also measure it using psychological intrinsic motivation scales but we note that these scales are subject to high social desirability bias, and indeed there seems to be little variation in the baseline data.

We also measure the midwives' degree of intrinsic motivation using a (non-incentivised) activity. We ask midwives to imagine that they were at the beginning of their midwifery studies and elicit their preference over a set of courses. These range from themes that lie at the core of midwifery, such as 'community midwifery', to subjects that are not directly related to it, such as 'principle of management'. Midwives with a higher preference for core midwifery subjects are considered more intrinsically motivated.

H2: The effect of Monetary retention incentives on midwives' retention decreases with midwives' pro-sociality.

We measure pro-social motivation in two ways:

First, we will use the altruism decision activity described in 3.2. In the private version of this activity, each midwife is allocated NGN 2000 (approximately USD 12) and asked to split this sum of money with the Nigerian Red Cross Society; she is reassured that the result of this allocation is treated as strictly confidential. Midwives who donate more money to the Nigerian Red Cross Society are considered more pro-socially motivated.

Second, we use midwives responses to questions on the pro-social work motivation scales (questions 13.1.1 to 13.1.13). Midwives are asked the degree to which they agree or disagree to a series of statements that describe the pro-social dimension of their job (e.g. " I do my best when I am working on a task that contributes to the well-being of others"). We hypothesise that the positive effect of non-monetary incentives on retention will be higher for midwives that score on pro-sociality according to these questions.

[15] At baseline midwives played a game where they are presented with labels describing attributes of your current job. These include stability, interaction with patients, respect from the community, opportunity to apply knowledge acquired in midwifery school, salaries and career opportunities. Midwives are then asked to identify the attributes that are most important to them.

H3: The effect of Non-monetary retention incentives increases with the midwives' pro-sociality.

We will measure pro-sociality as in H2.

H4: The effect of the simultaneous provision of monetary and non-monetary incentives exhibits heterogeneity, depending on midwives' levels of pro-social and intrinsic motivation.

We will measure intrinsic motivation as in H1 and pro-sociality as in H2

H5: The effect of incentives on retention depends on the midwife's own views on working harder in response to incentives

The hypothesis behind the image motivation hypothesis (see hypothesis M3): if an individual works harder in response to incentives, this has detrimental effect on other's people perceptions of that individual . In this hypothesis, however, we recognize that there might be an important cultural component, and that working harder in response to incentives might not be necessarily seen as negative.[16]

At baseline, we elicited from each midwife on what she thought (in terms of being generous, caring, etc.) of a midwife that worked harder in response to the introduction of a monetary incentive by the government (questions 2.16.1 to 2.16.4). Each question has three possible answers. We will score each question with a 1 if the answer portrays the midwife as being generous/caring and -1 if it portrays it as being selfish/uncaring, and 0 for the middle option. We will add the scores up to build an indicator that will be interacted with the treatment dummies in regression model (1), as well as in the duration model. To rely less on the cardinality of the above indicator, we will also test this hypothesis using binary variable that takes value 1 if there are more answers that portray the midwife to be generous/caring than to be selfish/uncaring.

[16] This is partly informed by causal observation after talking to some midwives who participated on our initial focus group discussions that informed the choice of non-monetary incentives.

We also elicited the same information as in questions 2.16.1 to 2.16.4, but in response to a non-monetary incentive (questions 12.1.1-12.1.4). We will repeat the estimation above but using the answers to these questions

H6: Monetary incentives have a higher effect on retention rates for midwives with lower discount rates. This kind of heterogeneity does not necessarily occur for non-monetary incentives.

Monetary incentives are paid quarterly. For midwives with a strong preference for the present (high discount rates) the present discounted value of the entire incentives plan is lower; this is likely to reduce the power of this incentive.

Non-monetary incentives have virtually no economic value, hence their power may not depend on midwives' discount rates. We hypothesise, however, that, for the trial arm that receives both types of incentives, the additional prospect of non-monetary reward may compensate midwives partly for the lower present discounted value of the monetary incentives. Thus, midwives' response to monetary incentives should depend less on time preferences for those in this arm of the trial.

H7: Monetary incentives have higher impact on retention rates for midwives with lower levels of risk aversion. This kind of heterogeneity does not necessarily occur for non-monetary incentives.

A large body of evidence suggests that behavioural responses to incentive schemes that incorporate risk, such as variable pay upon performance, depend significantly on the workers risk aversion (see Deckop et al, 2004 and references therein). Although the incentives packages provided in SURE-P MCH do not explicitly incorporate risk, changeability in public policies has been part of the Nigerian political context for decades. It is thus possible that midwives believe that there is a small but positive possibility that the incentives contract is discontinued. In this case, monetary incentives are expected to be more highly powered for midwives with lower levels of risk aversion. It is unclear whether preferences over risky financial stakes (used to elicit risk preferences) applies to the case of non-monetary

incentives. We hypothesise, however, that, for the trial arm that receives both types of incentives, the additional prospect of non-monetary reward may compensate midwives for part of the risk associated to a possible lack of payment of monetary incentives. Thus, midwives' response to monetary incentives should depend less from risk preferences for those in this arm of the trial.

5.3 Group C: Hypotheses on causal mechanisms

M1: Retention incentives affect retention rates because they carry information about the job

One of the explanations why incentives can backfire is because they might signal that the task that they incentivize is more costly for the individual than previously thought (Benabou and Tirole, 2003). In our context, midwives might think that a retention bonus is provided to them because working in the SURE-P PHC will become more unpleasant as time goes by, and they might quit their job sooner if a good enough job opportunity arrives.

To test this hypothesis, we asked midwives' for their views on how certain features of their current job and livelihoods will evolve in the future (questions 2.1 to 2.7 in the post-contract baseline questionnaire). This information was collected just after midwives were given their retention bonus contract.

Each question has three possible categorical answers. For each question, we will build two cumulative binary variables, and build an index using all the binary variables simultaneously. The index will be built using the methodology of Anderson (2008). Regression model (1) will be estimated using the computed index as dependent variable. We will also report the estimates on the individual questions but prefer the results on the index because it deals with problems associated with multiple hypotheses testing.

M2: Monetary and non-monetary incentives carry different information about the job

We will test this hypothesis by using the analysis above and testing whether the coefficients on the treatment dummies are different one from each other.

M3: The effect of monetary retention incentives depend on midwives concern with their image amongst co-workers. Non-monetary incentives do not harm image motivation.

Another explanation why incentives might not work is because they do not allow the individual to signal that they are altruistic when they engage in the incentive task (Benabou and Tirole, 2006). If a co-worker knows that a midwife is receiving a retention bonus, they do not know if the midwife is staying in her post because she is altruistic or because she wants to receive the bonus.

To elicit each midwife's image motivation, we play a private and a public version of an altruism game. In each version, each midwife is endowed with N2,000 and is free to decide how much to donate to the Nigerian Red Cross Society. In the private version of the game, they make this choice in a tablet and nobody can see it. Before choosing the quantity to donate in the public version of the game, the enumerator tells the midwife that the quantity that she donates will be publicly announced to all her co-workers (nurses, CHEWs, etc) at the end of the data collection.

We will define a midwife as having image motivation if she donates more in the public version of the game than in the private version of the game. We will interact this binary variable with the treatment variables in regression model (1); as well as in the duration model. There will be evidence of image motivation if any of the type of retention incentives is less effective for midwives who exhibit image motivation according to the binary indicator that we have built following the definition given above.

M4: Midwives compensate the potential harm to their image caused by incentives by engaging in more volunteering work

It could happen that midwives do have the image motivation that we refer to above, but it is not evident on the analysis outlined above because they manage to compensate the harm to their image with non-incentivized volunteering activities. To test that, we plan to estimate regression model (1) using as dependent variable whether the midwife dedicates any time

(binary variable) and the number of hours dedicated to unpaid volunteering activities at follow-up. For the purpose of this analysis, we will augment the vector X_{ifc} with the dependent variable collected at baseline (questions 1.a.25 and 1.a.26). In case the binary indicator that we have defined does not appropriately elicit image motivation, we will estimate two versions of model (1): one in which we do not interact the treatment dummies with the image motivation binary indicator, and another one in which we do. A problem with this analysis is that the opportunity cost of contributing to voluntary activities has probably changed for midwives who have moved community, and that this might be related to the retention bonus contract received. However, the analysis might still be valid if the differences across treatment arms are small.

M5: Retention incentives change *social norms*; they change midwives' perception about what constitutes socially acceptable behaviour in terms of length of service.

Fuster and Meier (2010) argue that incentives can change the social norm. In our case, because those midwives that quit their SURE-P job earlier are leaving the retention bonus on the table, it might be thought that it is not so bad to quit the job earlier. We also hypothesize that this effect will be more attenuated for non-monetary incentives.

We elicit the social norm on when to quit the SURE-P job by using an incentivized activity following Krupka and Weber, 2013 (see section on data collection). Midwives respond whether it is socially acceptable or not to quit the program at the following thresholds: two months, five months, eight months, eleven months, fourteen months, seventeen months, twenty months, or twenty-four months after joining the program.

We will test this hypothesis by estimating model (1), using as dependent variable the lowest time threshold (in months) for which they find acceptable to quit the SURE-P job.

M6: Retention incentives increase the number of midwives available in the health care facility.

Retention bonuses improve retention and hence they decrease midwife attrition. It takes considerable time until a new midwife replaces the one who dropped out. Hence, the retention bonuses increase the average number of midwives available in the facility.

To test this hypothesis, we will use the available information to compute the average number of midwives per day who are employed at the health care facility between baseline and follow-up. Using this as the dependent variable, we will estimate a linear regression as in (1), but aggregated at the facility level overall and by trimester (because attrition is more likely to happen in the latter trimesters). The regression will be augmented with the number of midwives in the facility at baseline.

M7: **Retention incentives lead midwives to consider that they are supported adequately for their work and that their employer appreciates the work they do.**

The provision of incentives might lead midwives to feel that they are better supported, and that their employer appreciates the work that they do. This is important because midwives might reciprocate by working harder (Fehr et al 2009 and Kube and Zultan, 2008). This same effect might also translate into improved retention. To test this hypothesis, we asked midwives "What sentence describes best how you feel about your relationship with the National Primary Health Care Development Agency". The answer is a four-point scale ranging from "extremely well supported" to "poorly supported". This information was collected just after midwives were given their incentive bonus contract.

In addition, to test whether this hypothesis holds in the longer run question 17.3.2 (follow up questionnaire) midwives are asked about their confidence in several institutions, amongst which are the National Primary Health Care Development Agency, SURE-P MCH and Nigeria's Federal Ministry of Health. To test this hypothesis, we will estimate an ordered probit version of regression model (1) in which the dependent variable is the answer to this question

References

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

Andreoni, J. and C. Sprenger. 2012. Estimating Time Preferences from Convex Budgets. *American Economic Review*, 102(7), 3333-3356.

Bénabou, R. and Tirole, J. 2003. Intrinsic and Extrinsic Motivation. *The Review of Economic Studies* 70 (3): 489–520.

Bénabou, R. and Tirole, J. 2006. Incentives and Prosocial Behavior. *The American Economic Review* 96 (5): 1652–78.

Deserranno, E. 2014. Financial incentives as signals: Experimental evidence from the recruitment of Health Workers.

Fehr, E. Lorenz, G. and Zehnder, C. 2009. A Behavioral Account of the Labor Market: The Role of Fairness Concerns. *Annual Review of Economics* 1 (1): 355–84.

Fuster, Andreas, and Stephan Meier. 2010. Another Hidden Cost of Incentives: The Detrimental Effect on Norm Enforcement. *Management Science* 56 (1): 57–70.

Giné, X., Goldberg, J., Silverman, D., Yang, D. 2014. Revising Commitments: Field evidence on the adjustment of prior choices.

Hatt, L., Stanton, C. Makowiecka, K., Adisasmita, A., Achadi, E. and Ronsmans, C. 2007. Did the strategy of skilled attendance at birth reach the poor in Indonesia?" Bulletin of the World Health Organization 85(10), 774–82.

Jenkins, Stephen P. 1995. Easy Estimation Methods for Discrete-Time Duration Models. *Oxford Bulletin of Economics and Statistics* 57 (1): 129–38.

Krupka, Erin L., and Roberto A. Weber. 2013. Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary? *Journal of the European Economic Association* 11 (3): 495–524.

Kube, S and Zultan, R. 2010. Treating Equals Unequally: Incentives in Teams, Workers' Motivation, and Production Technology. *Journal of Labor Economics* 28(4), pages 747-772, October.

Liu, E. 2013. Time to Change What to Sow: Risk Preferences and Technology Adoption Decisions of Cotton Farmers in China, *Review of Economics and Statistics* Vol. 95(4), 1386-1403.

Powdthavee, N. Boyce, C. and Wood, A. 2013. Is Personality Fixed? Personality Changes as Much as "Variable" Economic Factors and More Strongly Predicts Changes to Life Satisfaction, *Social Indicators Research*, 111(1), 287-305.

Rotter, J. B. 1966. Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General & Applied* 80(1) 1966, 1-28.