Community Health Educators and Maternal Health: Experimental Evidence from Northern Nigeria *

Martina Björkman Nyqvist, Jessica Leight and Vandana Sharma

July 30, 2019

Abstract

This paper shows the results from a randomized controlled trial of a community health worker program designed to enhance uptake of child and maternal health services, implemented in a conflict-affected region in northern Nigeria where change is desperately needed. Three interventions were evaluated: the deployment of community health educators; health educators with the provision of safe birth kits; and health educators with community dramas. The results suggest that the interventions increased utilization of antenatal, postnatal, and infant care, maternal and newborn health practices, and health knowledge. The community health worker program also worked better when supplemented with additional programs.

*Affiliations: Stockholm School of Economics, International Food Policy Research Institute, Harvard T.C. Chan School of Public Health. The evaluation was previously registered under the title "Strengthening the Midwife Service Scheme with Community Focused Interventions: A Randomized Controlled Trial." This project was the result of collaboration over many years among a large team. We would like to thank the Macarthur Foundation and USAID - Traction for funding, and particularly thank Judith Helzner, Erin Sines, Kole Shettima, and Oladayo Olaide at Macarthur for their ongoing leadership in the project. We would also like to extend our thanks to the Planned Parenthood Federation of Nigeria for collaborating on this evaluation, most particularly Ibrahim M. Ibrahim, Okai Haruna, John Adegbite, Damilola Toki, and Nafisatu Adamu. A large number of individuals contributed to the success of our field work, above all Laura Costica and Fatima Abdulaziz, without whose efforts the project would not have been possible; we would also like to acknowledge (in chronological order) Damilola Owolabi, Comfort Boman, Habiba Makanjuola, Salamatu Chiroma, Asinga Joshua Ungbo, Tsuya Halilu, Megan Kearns, Ahmed Sarki, Abdullahi Bulama, and Kelvin Chukwuemeka. We would also like to thank Jamin Dora Duba, Emmanuel Attah, and Kolapo Usman at the National Population Commission for their invaluable assistance, and the Ministry of Health of Jigawa state and particularly the Operational Research Committee (ORAC) chaired by Dr. Mohammad Abdullah Kainuwa for their guidance and leadership. For assistance in data analysis, we thank Luca Fachinello, Priya Iyer, and above all, Laura Costica.

1 Introduction

Enhancing health outcomes for pregnant women and newborns in developing countries has been a key policymaking goal over the last twenty years, but progress has been stubbornly slow. The World Health Organization estimates that more than 300,000 women and more than two million newborns die each year during pregnancy and childbirth (WHO, 2016). Although a sharp decrease in maternal and child mortality has been observed over the last decade in sub-Saharan Africa, the most affected region, progress has largely stagnated in areas characterized by political instability and violence. In order to identify strategies to target these adverse trends, it is crucial to evaluate health service programs delivered to women and newborns in communities where the challenges are most acute.

The majority of maternal and newborn deaths can be prevented via the appropriate provision of antenatal care in pregnancy, skilled care during childbirth, and postnatal care after delivery. However, availability and utilization of these services remains low in much of the developing world (WHO, 2016). Barriers to utilization of maternal health care are numerous, including low trust in available providers, inability to access services due to transportation, logistical or financial constraints, and disagreements within a household over care utilization (Gabrysch and Campbell, 2009; Simkhada et al., 2008).

In this study, we analyze a community-based health intervention designed to enhance maternal and neonatal health in rural northern Nigeria. The program relies on voluntary female community members who are recruited, trained and deployed as community health educators, deemed community resource persons (CORPs). CORPs conduct door-to-door visits to pregnant women in order to provide health information, encourage the utilization of facility-based care, and promote safe pregnancy and infant health practices; the goal is for each pregnant woman to receive six CORPs visits during the pregnancy and postpartum period.

The CORPs program was also supplemented with two additional health programs in two arms. The first additional program was the provision of safe birth kits distributed by

the CORPs to pregnant women. The kits contained sterile supplies that could be used during delivery at home or in a facility. While birth kits have been widely promoted by policymakers, including the World Health Organization, this is one of the first evaluations to generate rigorous experimental evidence around the effects of distributing birth kits to women outside of the formal health system. The second additional program was a series of community dramas implemented by a professional theatre group that sought to address misperceptions in the community and knowledge around maternal and child health. The dramas were conducted on a quarterly basis, promoting themes of safe motherhood and targeting a broad audience including men, elders, and traditional leaders.

The study was a large-scale cluster randomized controlled trial implemented in 96 rural villages/clusters in northern Nigeria between 2012 and 2016. The design included three treatment arms: the pure CORPs intervention, CORPs in conjunction with safe birth kits, and CORPs in conjunction with community drama. The evaluation sample included 7,000 women of reproductive age. Following the baseline survey, births were monitored on an ongoing basis for approximately two years; in 2016, a comprehensive endline survey was conducted. This evaluation was conducted in the state of Jigawa in the north of Nigeria, an overwhelmingly rural area characterized by extremely poor baseline maternal health outcomes. More importantly, since 2011 it has been extensively affected by violence linked to the Boko Haram rebellion.

This paper makes several contributions. First, although there exists some mixed evidence around the impact of community based health worker programs in developing countries, this is the first randomized controlled trial evaluating a community health educator program in West Africa and in particular, in a region characterized by one of the highest maternal mortality rates in the world, northern Nigeria (Sharma et al., 2017). Second, given that it is extremely challenging to operate and evaluate programs in areas affected by ongoing conflict, very little is known about the impacts of health interventions in these regions, and whether the programs in fact operate effectively. This evaluation was successfully conducted in an environment characterized by ongoing conflict linked to

Boko Haram, and thus generates the first evidence around a community health worker intervention in a conflict-affected region. Third, this paper contributes to the literature by analyzing how additional health interventions interact with a community health worker program, seeking to understand whether these interventions have an impact on program intensity. Finally, this evaluation provides the first experimental evidence around the efficacy of the distribution of birth kits outside of the formal health system, and provides additional evidence around the effects of community dramas designed to alter perceptions around safe motherhood and increase the take-up of formal health care services.

The first important finding generated by this evaluation is that it is feasible for community based health interventions to effectively reach beneficiaries even in areas characterized by ongoing civil unrest. In our core sample of around 4500 births observed during the two-year intervention period, 24% of women in treatment communities reported having had a visit from a CORP member. We also find that the CORPs were significantly more active in the experimental arms including supplemental interventions, a pattern that suggests that linking a traditional community health worker intervention to the provision of health inputs (such as birth kits) or educational outreach programming (such as community drama) may further motivate the voluntary health workers and hence, improve program performance.

The results show that in the pure CORPs arm, the probability that a pregnant woman receives a CORPs visit is 37% higher relative to a pregnant woman in the control arm; in the birth kits arm, this probability is 97% higher, and in the community drama arm, this probability is 77% higher. However, despite a higher level of CORPs engagement with women in the birth kits arm, only 10% of respondents in this arm reported receiving a kit. In the community drama arm, 36% of respondents report that a drama was conducted in their community, and 27% report attending such an event. In aggregate, the interventions effectively reached a relatively large number of pregnant women, despite the challenging environment in Jigawa state.

Our findings also reveal that the community based health interventions had significant

positive effects on a range of variables capturing care utilization and health practices, and the effects are proportionally large. We observe an increase in the probability of utilizing any antenatal care of 6–8 percentage points relative to 64% in the control group, or a proportional increase of around 11%, and an increase in the utilization of postnatal care of 3 percentage points relative to a control group mean of 7%, for a proportional increase of 45%. There is also evidence of greater intensity of use of antenatal care, and an increased probability that infants receive immunizations and check-ups in the first two months of life. The observed effects seem to be consistent across the two-year intervention period, suggestive of a persistent impact. There is, however, no significant effect on the location of deliveries, or on the probability of skilled attendance at birth.

Finally, we explore whether the interventions had any detectable effects on self-reported maternal and neonatal morbidity and neonatal and infant mortality, and anthropometrics of children born during the intervention period as observed at endline. There is no evidence of any significant effects on any measured health outcome, despite the observed increases in service utilization. We hypothesize that this may reflect low returns to utilization of formal health care in this setting. Alternatively, the increases in utilization may not be large enough in magnitude to generate any detectable shifts in health outcomes in the relatively short two-year period.

This paper links to several related literatures in economics and public health. There is increasing evidence around the effects of community health worker programs designed to provide information and/or health commodities to rural households, though previous evidence largely focuses on the effects on neonatal mortality and child health. Systematic reviews of randomized controlled trials of community health worker programs and related interventions have found mixed evidence, reflecting the wide heterogeneity in supervision and incentives provided to the health workers themselves (Lewin et al., 2010; Okwundu et al., 2013). Two recent papers evaluating the effect of community-based training for mothers found that information provision around curative treatments of malaria yielded a significant reduction in under-5 mortality (Kidane and Morrow, 2000), while more general

information about child care had no effect on neonatal and infant mortality (Sloan et al., 2008). Björkman Nyqvist et al. (2019) analyze a micro-entrepreneurship program in which community health workers were incentivized to provide education and sell health products during home visits to households with children younger than five years old, and report a significant reduction in child mortality over a three year period. This paper contributes to this broad literature by analyzing the effects of community health workers in a region where utilization of formal maternal and child care is particularly low, and in a context of ongoing civil unrest, making utilization of health services even more challenging.

There is also a small evidence base around the use of safe delivery kits; however, this literature primarily relies on non-experimental analysis, despite the fact this intervention has been widely promoted by the international health community (including the WHO) in recent years (WHO, 1996). The majority of the evaluations are small-scale, and focus on the impact of safe birth kits distributed via the formal health system. Two studies in Tanzania and Egypt find that women who received a birth kit at primary health clinics were significantly less likely to experience cord infection or puerperal sepsis (Winani et al., 2007; Darmstadt et al., 2009). Hundley and Avan (2012) provide an overview of evidence around safe delivery kits, and they note that only nine studies reported results of an intervention including birth kits; only one was a randomized controlled trial, and none were able to separately identify the effects of birth kits vis-a-vis other interventions. This study will be the first to evaluate take-up and usage of safe birth kits distributed via community channels outside the formal health system, and in a context where home-based delivery is the norm.

Finally, our paper also joins a small but growing debate around the returns to utilizing formal health care for delivery. Okeke et al. (2016) found no evidence that the expansion of the Midwives Service Scheme in Nigeria itself had a significant effect on maternal or neonatal outcomes in a quasi-experimental evaluation (described in more detail later in the paper). Chari and Okeke (2014) conclude that a policy-induced shock to the supply

of institutional deliveries did not have a significant effect in reducing newborn mortality in Rwanda. Evidence from Malawi suggests a government ban on the use of traditional birth attendants did result in a significant decrease in use of TBAs and an increase in utilization of formal sector care, but no overall decline in newborn deaths (Godlonton and Okeke, 2016). An earlier quasi-experimental analysis in Indonesia concluded that the deployment of midwives to rural villages led to improved nutritional status for children (Frankenberg et al., 2005). These papers analyze interventions encompassing the formal health system with educated health staff; given that they find no impact on maternal and child mortality, this evidence is suggestive of potentially low returns to care. In contrast, in this paper we are exploring the introduction of a community-based strategy designed to enhance maternal and child health: training volunteers with no formal education as community health workers in order to encourage pregnant women to utilize the existing formal health system.

The remainder of the paper proceeds as follows. Section 2 describes the context, and Section 3 describes the intervention and experimental design. Section 4 reports the primary experimental results, and Section 5 reports robustness checks. Section 6 concludes.

2 Institutional setting

This evaluation was conducted in 96 rural communities in Jigawa state in northern Nigeria, an environment that ranks among the most challenging in the world for maternal and child health. Though Nigeria accounts for only 2% of the world's population, it accounts for 10% of worldwide maternal deaths, and it is one of a handful of countries that have experienced stagnation and even slight increases in maternal mortality since 1980 (Hogan et al., 2010). Northern Nigeria is characterized by even more adverse health indicators for women of reproductive age. The maternal mortality rate in Jigawa state, the site of this evaluation, is estimated to be around 1,012 per 100,000 live births, compared to 576

per 100,000 nationwide (Sharma et al., 2017). Similarly, the neonatal mortality rate in the region of northwest Nigeria is 42 per 1,000 live births, compared to a national average of 37 (Akinyemi et al., 2012; NPC and International, 2014).

This region has also recently been vulnerable to ongoing violence linked to the Boko Haram rebellion. Between 2012 and 2016, the Armed Conflict Location and Events Dataset (ACLED) reported 23 violent events in Jigawa, of which at least six are linked to Boko Haram, and the neighboring states of Yobe and Bauchi reported 224 and 116 violent events respectively, of which 177 events in Yobe and 31 events in Bauchi were linked to Boko Haram. This evidence suggests that the region has been undergoing serious civil unrest in the period that coincides with this evaluation.

Jigawa is a poor, rural state with an overwhelmingly Muslim and Hausa-speaking population (EIU Canback, 2016), and it is characterized by very low levels of human capital attainment for women and children. In the 2013 Demographic and Health Survey, only 11% of women in Jigawa were literate, compared to regional (national) averages of 26% (53%). Among Nigerian states, Jigawa has the third lowest female literacy rate. Turning to maternal and child health, Jigawa also has the third lowest rate of facility births in Nigeria; 7% of women in the state reported their most recent birth was in a health facility, compared to a regional average of 12% and a national average of 36%. Similarly, only 4% of children age 12–23 months were fully vaccinated, compared to 10% in the northwest region and 25% nationwide, rendering Jigawa the state characterized by the fourth lowest vaccination rates (NPC and International, 2014).

These poor health outcomes are observed in conjunction with a pattern of household decision-making that is heavily dominated by men. In the baseline survey of this study (described in more detail in Section 3.3), more than 80% of the 7089 women surveyed reported that their husband was responsible for financial and health-care decisions in their household. Moreover, 94% of women concurred that their husband should be the sole decision-maker around the utilization of antenatal care during pregnancy, and 93% of women similarly reported that he should have sole decision-making power around

the utilization of facility care for delivery. Clearly, women in this region have very little decision-making power around health care decisions during pregnancy, an important finding to consider in the context of a program designed to increase utilization of maternal and child health services.

A survey of 24 primary health centers serving the evaluation sample was conducted to provide additional information about the formal health care services available in this region, and summary statistics are reported in Table B1 in the Appendix. Primary health centers generally provide basic preventive and curative care, including antenatal care, delivery and postnatal care, childhood immunizations and nutritional care, and minimal family planning services. While a majority of facilities report access to electricity and water over the last six months, only 4% report access to a telephone, and 21% report access to an ambulance. Assisted vaginal delivery and neonatal resuscitation are reported available in nearly all facilities, but fewer than 40% report access to a blood bank for transfusions, and no facility can conduct caesarean sections. Primary health centers report have on average three beds for labor and delivery, postpartum and newborn patients, relative to 44 deliveries conducted on average per month. Eight of the 24 primary health centers reported a general doctor on staff, but no health center was staffed by an obstetrician or other specialist. These stylized facts suggests that the formal health system available to the women in our sample is characterized by weak infrastructure, limited educated health staff, and a high volume of deliveries relative to the number of beds available.

Thus across a range of health indicators, evidence suggests that even in the Nigerian context, Jigawa is among the states characterized by the worst human development outcomes for women and children, and clearly the maternal health services available to the sample population are severely limited. Whether a community health worker program targeted to increase uptake of formal health services can be effective in this context is clearly an important empirical question that we seek to analyze in this paper.

3 Intervention and empirical design

3.1 Intervention

The objective of this research project is to evaluate the impact of three community-based health interventions in a challenging setting in northern Nigeria. These interventions were implemented in the context of the Midwives Service Scheme (MSS), a program that was launched by the Nigerian government in 2009 to fund the deployment of midwives to rural primary health centers (Okeke et al., 2016).

In northern Nigeria, our partner organization, the Planned Parenthood Federation of Nigeria (PPFN), rolled out three new interventions to stimulate utilization of these newly available maternal health services.[1] The core intervention entailed the deployment of community health educators, designated community resource persons (CORPs). CORPs were defined as "a bridge/link crucial to foster trust, confidence and acceptance between the midwives and their clients and . . . increased access to maternal health services" (PPFN, 2012b). Women between the ages of 20 and 45 who were married, widowed, or divorced were eligible to serve as CORPs provided they possessed a minimum of primary school education. Recruitment was managed by PPFN in conjunction with the ward health committee, a local committee with the responsibility of overseeing health services in each community.

CORPs were mandated to provide health information in a series of six home visits to pregnant women. The objective was to conduct visits at different points during pregnancy in order to monitor women's health status, encourage them to seek formal health care if needed, and promote the benefits of delivery at the primary health center. The first visit by the CORPs would be conducted in the first trimester, focusing on the benefits of antenatal care and adequate nutrition in pregnancy. The next two visits would be conducted in the second trimester, focusing on danger signs during pregnancy, and the

[1] Founded more than 25 years ago, PPFN is now one of the oldest indigenous organizations in Nigeria offering sexual and reproductive health services; however, the interventions proposed as part of this evaluation were new to the organization.

importance of birth preparedness. Two additional visits would be conducted in the third trimester, highlighting the benefits of facility delivery, and ensuring a birth plan has been formulated by the household. The sixth and last visit would be conducted postpartum, focusing on postpartum danger signs, the importance of exclusive breastfeeding and immunization, birth registration, and family planning. CORPs were encouraged to include husbands or other family members in the educational visits to the households, and they were also responsible for identifying newly pregnant women in the village.

At the launch of the intervention in 2012, the CORPs received a one-week training, and they received a small monthly stipend of 2000 naira (about $5) thereafter. Four CORPs were recruited for each community of approximately 500 households. Accordingly, each CORP was responsible for roughly 150 households, and given the estimated birth rate, would conduct approximately 15 household visits to pregnant women each month.

In the first treatment arm, the CORPs intervention was implemented in isolation, but in the other two treatment arms, it was supplemented by additional interventions. In the second treatment arm, CORPs were provided with birth kits (also known as safe delivery kits or clean delivery kits) to distribute at no cost to all pregnant women in their third trimester. The objective was for women to utilize the kits during delivery either at home or in the health facility. The kits included a plastic sheet for the woman to lie on during delivery, surgical gloves for the birth attendant, a sterilized razor and cord clamps to cut and tie the umbilical cord, methylated spirit, clean gauze, swabs and perineal pads to be used by the mother after birth, a gallipot, a mechanical suction tube to clear secretions from the baby's airways, and a wrapper and diapers (PPFN, 2012a). All materials are packaged in a single sterile unit, and had an estimated cost per kit of $8.[2] The kits distributed in this intervention were comparable to those promoted by the World Health Organization as a promising strategy to strengthen hygienic standards for deliveries conducted at home and in low-capacity facilities (Hundley et al., 2012).

The objective of the birth kits intervention was twofold: first, women who were ap-

[2]The kits were identical to those available to midwives working in the MSS primary health care centers.

prehensive that the primary health center would not have adequate supplies for their delivery could utilize the birth kit for a facility-based delivery. While in theory MSS-served facilities should have basic medical supplies available, in practice stock-outs were frequent (Okeke et al., 2015). Second, women who preferred to deliver outside a facility could utilize the birth kit to reduce their risk of infection during delivery.

In the third treatment arm, the CORPs program was implemented in conjunction with a series of community dramas conducted in the village in order to promote the importance of safe motherhood to a broader audience. The dramas were conducted quarterly by a local drama organization called YARAC (Youth, Adolescent, Reflection and Action Center), and included an entertaining performance followed by a community forum. The objective of this intervention was to shift norms around the utilization of maternal health care, and target community members who are primary decision makers but might not be well-informed about maternal health challenges (including elders, traditional leaders, and especially men who would not necessarily interact directly with a CORP member).[3] The drama was a script based on an intra-household conflict over utilization of maternal health care services in which a pregnant woman wishes to use facility-based care while her mother-in-law opposes her choice. The implementing organization reported that the dramas were widely attended by both men and women in the villages, and that was also confirmed in video recordings made available from some sample villages.

All three interventions were targeted to address the challenges around maternal health observed at baseline. Given low utilization rates of health services and perceived low levels of trust between communities and facility staff, the CORPs intervention was designed to provide information about the newly available midwife services, increase knowledge about the benefits of utilizing these services, and increase confidence in service quality. The CORPs also sought to enhance maternal and neonatal health by providing information about recommended health practices including nutrition during pregnancy, danger signs during pregnancy, delivery and the postpartum period, and breastfeeding and infant care.

[3]A mapping of health programming in the area suggests that similar events had not previously been conducted in the sample communities.

The birth kits intervention was designed to provide a safer, more sterile environment for women during delivery, regardless of whether they delivered at home or in the health facility. Finally, the drama intervention aimed to reach men as well as other community stakeholders and change social norms around safe motherhood, given the evidence that health care decision-making in this context is dominated by men.

The CORPs intervention and the birth kit intervention were launched in January 2013, while the drama intervention was launched in the third quarter of 2013. Data collection was conducted between 2012 and 2016. A timeline for the project is provided in Figure 1.

3.2 Evaluation design and sample

The study was a parallel-group, stratified cluster randomized controlled trial, where randomization was conducted at the community level. The evaluation design called for the inclusion of 96 clusters (communities), with 24 clusters in each of four evaluation arms and where each community included around 500 households. The surveyed sample comprises 15% of the households in the village that have a woman of reproductive age at baseline. Births among the sampled women are monitored continuously for a two-year period of intervention implementation, and all women are surveyed again at endline. Given this design, the evaluation was powered to detect a 25% decrease in the maternal morbidity and neonatal mortality rates with 90% power to detect an effect in a one-sided test at a ten percent significance level.[4]

In order to identify the sample communities, we first identified the target local government areas (a subdivision analogous to a district). Jigawa has 27 local government areas (LGAs), and all 24 LGAs that contained a primary health center receiving services under the Midwife Services Scheme were included in the evaluation. Within each of the

[4]This is assuming a baseline maternal morbidity rate of 35%, and a baseline infant mortality rate of 47 deaths per 1000 births. In addition, the design assumed a birth rate of 46 per 1000 population, and accordingly 21 sampled births per year per community of 3000 individuals; the intra-cluster correlation was assumed to be .2, and the power calculations calculated 90% power to detect an effect in a one-sided test.

24 LGAs, we identified all villages of the target size that were located in the catchment area of the MSS facility. Thereafter, a random subset of 96 communities in total were selected to be included in the evaluation sample. The villages frequently included more than one immediately adjacent settlement or hamlet, and we therefore utilize the term clusters to describe them. See Figure 2 for a description of the trial profile. Finally, we conducted a partial census in all communities that was employed as the basis for baseline sampling. More details on the sampling strategy for identifying communities and households are provided in Appendix Section A.1. The trial was registered with the American Economic Association registry (Leight et al., 2016), and registered at clinicaltrials.gov.[5]

3.3 Data

The evaluation included three primary phases of data collection: baseline, ongoing data collection (surveys conducted 3 days and 28 days after birth in addition to a sample-wide audit of recent births), and an endline survey. The data collection process is summarized in Figure 1.

The baseline survey was conducted between March and June 2012 and included 7069 households with a female household member of reproductive age (corresponding to roughly 15% of the listed households). More details are provided in Panel A of Table B2 in the Appendix. The respondent of the baseline survey was a woman in the household of reproductive age (between 15 and 49). If more than one eligible woman was present, the respondent was randomly selected utilizing an on-the-field randomization protocol. The baseline survey included information about the household's socioeconomic characteristics, the respondent's birth history, utilization of health services during the respondent's last birth (if she had given birth within the preceding 24 months), and health knowledge and attitudes. In addition, anthropometric measurements (height, weight, and mid-upper-arm circumference) were collected for the respondent's youngest three children under the age of five.

[5]The protocol number is NCT01487707.

Ongoing data collection entailed continuous monitoring of births among the baseline sample during the two-year study period. Female monitors in each community were recruited and trained to provide a cardboard chip to all baseline households and send a simple text message by SMS following a birth or infant death in the baseline households. The SMS messages were redirected to one of our enumerators, who had the responsibility of following up on the SMS messages by identifying the household and conducting surveys 3 days and 28 days after birth.[6]

The three-day survey included questions about utilization of antenatal care, the mother's health during pregnancy, the delivery itself (including the location of the delivery and any complications encountered), and the mother's and newborn's health since birth. The 28-day survey included questions about the mother's opinions around utilization of maternal health care, maternal and neonatal morbidity in the first month, as well as questions related to newborn care practices. The continuous monitoring and data collection was initiated in November 2013 and continued until November 2015. Unfortunately, the continuous monitoring survey was incomplete given that village monitors did not fully comply and failed to identify some births in the baseline sample households. More specifically, 1791 3-day and 28-day surveys were conducted, corresponding to 41% of all births reported for the sample during this period.

The survey team also conducted an audit between January and March 2015 (i.e., roughly at the halfway point of continuous data collection) in which all baseline households were scheduled to be revisited. The enumerators posed a brief series of questions about births in the household, and if they identified any births that had previously been missed, conducted a more concise survey that included some key questions from the 3-day and 28-day surveys. Approximately 70% of baseline households were reached in the audit survey (4674 households), and enumerators conducted detailed surveys on 802 additional births in 802 distinct households; the majority of these surveys were for births that had

[6]In the event an enumerator became aware of a birth more than three days after the birth, she was still instructed to conduct the three-day survey as soon as possible, and then return for the 28-day survey. In the event she became aware of a birth more than 28 days after birth, she was instructed to conduct the survey up to three months after birth.

not been included in the ongoing data collection.

Finally, the endline survey was conducted between February and July 2016.[7] We successfully re-surveyed 90% of the baseline sample, or 6350 households. Some households that could not be reached during the endline survey (households that had migrated, divorced couples, etc.) had already been surveyed previously in the post-birth or audit surveys. Therefore, the total number of women observed in any follow-up data collection was slightly larger than the endline sample (6494 women, or 92% of the original sample).

Given that the entire state was significantly affected by ongoing violence and instability linked to Boko Haram attacks during the evaluation period, we regard the observed pattern of attrition as low. In the analysis, we find no evidence that the correlation between baseline characteristics and a dummy for attrition differs significantly across treatment and control arms. More details on patterns of attrition are found in Section 5.2.

3.4 Outcomes and statistical framework

We analyze the impact of the interventions on five primary categories of outcomes: intervention exposure, utilization of maternal health care, maternal and newborn health practices, child anthropometrics, and maternal and child morbidity and mortality, as well as secondary outcomes related to knowledge and attitudes and fertility. These variables are found in the endline survey as well as in the post-birth survey (3-day or 28-day).

First, we analyze patterns of household exposure to the community health educator program, and identify whether the intensity of the CORPs program was different when additional interventions — safe birth kits and community dramas — were implemented. In particular, we report whether women who experienced a pregnancy during the follow-up period had received any visit from a CORP member, and how many visits. For the birth kits intervention, we report whether the woman had received a kit, if she knows how

[7] Some additional intensive data collection targeted to minimize attrition continued until October 2016.

to use the kit, whether she can name at least one item in the kit, and whether she reports utilizing the kit in her most recent delivery. Finally, we report whether the woman and her spouse attended the community drama.

Second, we analyze intervention impacts on utilization of maternal health care. Variables of interest capturing the utilization of maternal health care include the respondent's reported antenatal care visits, the number of visits, an index of antenatal care quality, and whether the woman received a postpartum check-up within two months. For delivery care, variables of interest include whether the program impacted facility-based delivery or skilled attendance at birth. We also analyze intervention effects for maternal and newborn health practices: whether the respondent developed a birth plan, breastfeeding practices, and immunizations and infant check-ups. We further analyze an index of maternal morbidity (during pregnancy, delivery, and postpartum), an index of neonatal morbidity and infant and neonatal mortality. Finally, we analyze child nutritional status as captured by height-for-age, weight-for-age, and MUAC-for-length, measured at endline for all children under five in households reporting a birth during the intervention period. In addition to these five primary outcome categories, we also analyze a set of additional secondary outcomes capturing knowledge and attitudes and fertility.

3.4.1 Sample for primary analysis

The sample for the analysis of the primary outcomes linked to maternal and newborn health is restricted to the subset of women who reported a birth during the intervention period, or women who were pregnant at endline. It is important to note that women who had not been pregnant during the intervention period do not report any pregnancy or delivery-related outcomes, and are thus not included in the analysis sample. Our definition of births includes live births, stillbirths, and contemporaneous pregnancies at endline. (No systematic data was collected on miscarriages, other than one question posed at the endline as to whether the respondent experienced at least one miscarriage in the preceding two years.) Specifically, 4420 births corresponding to 4290 women are

observed in the intervention period; thus 61% of the original sample reported at least one birth. 130 women, or 2%, report two births.

Selection into this subsample is potentially a source of bias if the intervention affected choices around fertility or birth timing. The CORPs were not mandated to target decisions around fertility, other than to encourage women to consider birth spacing following a birth. Given the relatively short evaluation timeline, the vast majority of women delivered only once, and thus there is very limited scope to observe any effect on timing following the first birth. Moreover, since CORPs did not interact with women until they were pregnant, there is similarly very little scope for an effect on birth timing for the first birth during the evaluation period.

We provide two sources of evidence consistent with the hypothesis that there is no selection into the subsample of women reporting births. First, we analyze whether there is a treatment effect on fertility, and find no evidence that the treatment shifted the probability that a respondent reports giving birth, or the number of births. (These results are reported in Table 9.) Second, we evaluate baseline balance in the subsample of respondents who report a birth. These results are reported in Tables B4 and B5, and also here we find no evidence of imbalance across arms in this subsample. We conclude that bias is not introduced by evaluating the sample of women who have given birth or are currently pregnant.[8]

Within the sample of 4420 births, 55% were observed in both a post-birth survey (conducted 3 and 28 days after birth) as well as in the endline survey. However, in 45% of cases, one survey was missing; more details are provided in Section A.3 in the Appendix. For the births for which one survey is missing, some outcomes of interest are not reported.

In addition, 411 of the pregnancies reported had not ended in a live birth at the point of the endline (either the woman was pregnant at the time of the endline survey, or she had

[8]An alternative experimental design would have entailed sampling only pregnant women at baseline. However, the size of the sample required would have rendered it impossible to identify the target number of pregnant women, survey them, and then expose them to CORPs programming all prior to delivery, especially given that the gestational age at verification of pregnancy can be quite late in this context.

suffered a stillbirth). For these cases, data on labor and delivery (for contemporaneous pregnancies), postnatal morbidity, anthropometrics, and newborn care is correspondingly unavailable. Our analysis suggests that baseline characteristics do not predict patterns of selection into the different surveys, and we therefore conclude that there is no bias introduced by missing data.[9] Further details are provided in Section A.4 in the Appendix.

3.4.2 Econometric specification

The primary specification used to estimate the effect of the interventions on an outcome variable X_{bicg} for birth b observed for respondent i in cluster c in LGA g can be written as follows:

$$X_{bicg} = \beta_1 CORPS_{cg} + \beta_2 Kits_{cg} + \beta_3 Drama_{cg} + \chi_{icg}\gamma + \mu_g + \epsilon_{bicg} \quad (1)$$

$CORPS_{cg}$, $Kits_{cg}$ and $Drama_{cg}$ are indicator variables denoting treatment assignment, and χ_{icg} is a vector of baseline control variables.[10] All regressions include LGA fixed effects μ_g, and standard errors clustered at the cluster level.[11]

For some outcomes, we have a group of related outcome measures. To assess the impact of the intervention on a set of K related outcomes, we follow Kling et al. (2004) and

[9] We also collected some limited information about the prevalence of miscarriages, defined as pregnancies terminating in the first two trimesters, and will briefly discuss this result below. However, detailed information about care utilization, maternal health, etc. was for obvious reasons not collected for those pregnancies that terminated in miscarriage.

[10] The control variables employed include all those reported in the balance tests in Panel A of Table 1: a dummy variable for whether the respondent is married, the number of co-wives, age at marriage, the number of marriages reported, age, a dummy for whether the respondent has ever attended school, a dummy variable for whether the respondent reads Hausa, a dummy variable for Muslim, current birth parity, and a wealth index. We also include dummy variables equal to one if the respondent is observed in the 3-day and 28-day surveys, and in the audit survey.

[11] Prior to initiating the analysis, the outcome measures of interest were identified and registered in a pre-analysis plan (Leight et al., 2016). There are three variables included in the analysis plan that were not included in the analysis here. The maternal mortality rate is not reported given that the verbal autopsies that would be required to identify the cause of death for women of reproductive age who died were ultimately not collected fully for all reported deaths. The perinatal mortality rate (calculated using data on stillbirths and deaths within the first seven days of life) was not calculated because information on the exact date of birth and death of newborns was frequently missing, rendering it challenging to identify a perinatal death rate. The under-five mortality rate was not calculated as the choice was made to report under-one mortality; however, both variables yield null results for the effect of the interventions estimated.

estimate a seemingly unrelated regression system. We then derive average standardized treatment effects, $\tilde{\beta} = \frac{1}{K}\sum_{k=1}^{K}\frac{\hat{\beta}_k}{\hat{\sigma}_k}$, where $\hat{\beta}_k$ is the point estimate on the treatment indicator in the k^{th} outcome regression and $\hat{\sigma}_k$ is the standard deviation of the control group for outcome k (Duflo et al., 2007).

The trial was not designed with sufficient statistical power to analyze differential treatment effects across the different treatment arms. Accordingly, while we do present tests analyzing differential treatment effects across arms, these tests should be interpreted with caution. More specifically, we report at the bottom of each table tests of equality across β_1, β_2, and β_3, and also report a joint test of the impact of any intervention $\beta_1 = \beta_2 = \beta_3 = 0$. In addition, we estimate a separate regression including a dummy variable for assignment to the CORPs only arm and a dummy variable for assignment to any "enhanced intervention" arm (birth kits or drama), and report a test $\beta_1 = \beta_{enhanced}$, where $\beta_{enhanced}$ denotes the effects of the enhanced interventions. This test allows us to identify whether there is the effect of the arms including enhanced interventions is significantly larger.

4 Results

4.1 Balance at baseline

Tables 1 and 2 present mean pre-treatment characteristics for the treatment (CORPs only, CORPs and birth kits, and CORPs and community dramas) arms as well as the control arm; in addition, we regress the outcomes of interest on dummies for each evaluation arm conditional on strata fixed effects in order to test balance in baseline characteristics. We report in the tables the p-value on the joint test $\beta_1 = \beta_2 = \beta_3 = 0$.

Demographic and socioeconomic characteristics of the households are presented in Table 1. The average respondent is age 28 at baseline, with a reported age at marriage of 15; around a third live in polygamous households. Only about 17% of the respondents ever attended school, and 10% are literate in Hausa. They are almost universally Muslim,

and have on average three living children. The primary occupation of about half of the household heads is own cultivation, while 29% are self-employed in non-farm work, and 12% are employed outside the household.

The baseline values of the main outcome variables of interest are reported in Table 2, including health care utilization, health practices, maternal and neonatal morbidity, and anthropometrics; these variables are reported for the subset of 4007 women who reported a birth during the two years preceding the baselines survey. Around 65% of women report utilization of antenatal care at baseline. However, only 9% of women report their most recent delivery was in a health facility, and only 11% report skilled attendance at this delivery. For children under the age of one, the percentage identified as underweight, stunted or characterized by low MUAC-for-age ranges between 16% and 25%. This evidence is consistent with the DHS evidence described in Section 2.

In the last row of Table 2 we report the p-value corresponding to the joint F-test of the hypothesis $\beta_1 = \beta_2 = \beta_3$, testing across all variables examined in both Tables 1 and 2. From this joint test we can conclude that there are no significant differences in observable characteristics in aggregate when comparing across treatment arms. In addition, we report in Tables B4 through B7 in the Appendix the same set of balance tests for respondents who report a birth in the intervention period and for the subsample of non-attrited respondents. Again, the joint tests do not reject the hypothesis of balance across arms, suggesting the randomization effectively generated a sample balanced with respect to observable characteristics at baseline.

4.2 Intervention exposure, care utilization and health practices

We first analyze the evidence around pregnant women's exposure to the interventions, patterns of utilization of maternal health services, and health practices. Table 3 reports the results of estimating equation (1) for variables capturing intervention exposure. Pregnant women in all three treatment arms were significantly more likely to report a visit by a CORP during their pregnancies compared to pregnant women in the control group,

despite the fact that we observe some contamination in the control arm (13% of women in the control group report a visit from a CORP).[12] Intervention coverage was 37% higher in the CORPs-only communities compared to the control group, when coverage is defined as the percentage of women reporting a birth who had a visit from a CORPs. We find even higher coverage in the treatment arms that had an additional health program, relative to the control arm: 97% increased coverage in the birth kit communities and 77% in the drama communities. The difference in program coverage between the pure CORPS program and the enhanced CORPs program (in conjunction with either the birth kit or the drama program) is statistically significant at the 1% level, suggesting that CORPs seem to have been more motivated to identify and visit pregnant women when other interventions entailing provision of health inputs or community education were also implemented. On average in the three treatment arms, 22% of women reporting births interacted with a CORP during the program period, and the difference between the treatment and the control arms is statistically significant; women in intervention communities report higher coverage rates. Conditional on reporting a CORPs visit, the average respondent reported 1.6 visits in total.

Within the sample of women who reported receiving a visit from a CORP, the most common topics reported discussed were the importance of antenatal care (70%) and healthy pregnancy practices (69%), followed by the importance of delivering at a clinic (63%). Other topics discussed included symptoms of a difficult pregnancy (54%), postpartum complications (44%), plans for delivery (43%), and newborn care practices (37%). These were all areas included in the CORPs curriculum and highlighted in the protocol for visits to pregnant women, and thus our survey results confirm that CORPs were in general disseminating the information they were trained to provide.

Next, we seek to explore the evidence around intervention implementation in the birth kits and drama arms. Column (3) in Table 3 reports results for the birth kit arm; here, we observe that among the 28% of women who reported they received a visit from a

[12]There is no evidence that contamination is higher in control communities that are geographically more proximate to treatment communities.

CORP, 10% reported receiving a birth kit. There was no evidence of any contamination in the control group. When asked specific knowledge questions about the birth kits, 7% of respondents stated they knew how to use the kit and 8% could name at least one item included in the kit. Clearly, the birth kit intervention was not as inclusive as intended, as only a small fraction of pregnant women received and used the kits. We were provided with reports from our NGO partner that they initially faced challenges in delivering the kits to the CORPs, but that this logistical challenge was subsequently resolved. The reasons why the CORPs did not distribute the birth kits to a larger extent is unclear since we did not survey the CORPs themselves.

By contrast, the drama intervention was more successful in achieving higher coverage rates, as reported in Columns (7) and (8) in Table 3. We find that 36% of respondents in the drama arm state that events were conducted in their community, and 27% state that they attended. The goal of the drama intervention was to disseminate information about maternal health risks and the importance of health care utilization in pregnancy to other members of the community, particularly men, and our data suggests that this goal was effectively achieved: among respondents who attended a drama, 32% report their husbands also attended, and 51% report that they discussed the content with a relative or friend. Conditional on attending at least one drama, the average number attended was 1.6, and the effect is statistically significant at the 1 percent level. We also observe some contamination of the drama intervention in the other arms, as 8-9% of the respondents in the control arm and 19% of women in the birth kits arm report exposure to a drama in their community. It is challenging to identify whether this reflects exposure to group educational events that may have been conducted by CORPs members, or if a drama event was in fact conducted.

Adding up the results on intervention exposure, it is clear that the CORPs program successfully reached the intended beneficiaries, and that the intervention worked better when it was accompanied by additional interventions (birth kits or dramas). The health educator program was able to reach the target population despite that the fact that the

NGO was working in a challenging environment with ongoing civil unrest, as well as in a region where women have very little power over health seeking behavior.

4.2.1 Utilization of maternal health care

The previous evidence demonstrated the intervention reached pregnant women in the target communities and educated women about the importance of maternal health. Our next objective is to understand if these interactions had a significant effect on women's health-seeking behavior.

Table 4 reports results analyzing women's utilization of maternal health care services during pregnancy. For antenatal care, we observe in Columns (1) and (2) an increase in the probability of using antenatal care in the birth kit and drama treatment arms of 6.3 to 8.6 percentage points that is statistically significant; relative to the control arm, these are increases of around 11%. We also observe an increase in the number of antenatal care visits of around 10% that is also statistically significant. In order to capture the quality of antenatal care, we construct an index equal to the mean of indicator variables for receiving important components of antenatal care: utilizing care in the first trimester, receiving more than half of available ANC services, receiving iron folic pills and the tetanus vaccine, and receiving advice on danger signs during pregnancy. The estimated coefficients reported in Column (3) suggest there was an increase in the quality of antenatal care in the birth kits and drama arms.[13] In addition, Column (4) reports an increase in the probability of utilizing postnatal care of around 2–3 percentage points in all treatment arms, though this coefficient is significant only for the CORPs and drama arms. Given the control mean of 7%, this is a proportionally large effect of 45%.

Column (5) reports the average standardized treatment effects for these four preceding outcome variables. Here, we observe a 0.12 standard deviation increase in utilization of non-delivery care, statistically significant for the birth kits and drama treatment arms.

[13]This effect is largely driven by increases in the receipt of iron pills and tetanus vaccines, as well as reported counseling about pregnancy danger signs.

This suggests that despite the fact that the interventions achieved coverage rates that were lower than intended, they generated a significant effect on utilization of maternal health care services. Moreover, the two interventions with the strongest coverage (birth kit and drama treatment arms) were also the two interventions with largest impact on utilization of maternal health services. The enhanced interventions (where the CORPs program is implemented in conjunction with another health program) had a larger impact on antenatal care and its quality compared to the treatment arm with the pure CORPs program, and the difference is statistically significant at the 1% level.

In Columns (6) through (8), we report treatment effects for the probability of a facility delivery, skilled attendance at birth, or the probability of delivering with another person present. Here we find uniformly null effect across all treatment arms.[14] In Column (9), we report the average standardized effect for the full family of outcomes capturing utilization of maternal health services (i.e. Columns 1-4 and 6-8); the estimated coefficients suggest there was an increase in care utilization of around 0.07 standard deviations that is statistically significant in the birth kits arm and drama treatment arm. This evidence suggests the intervention had a positive impact on maternal health care utilization in the two treatment arms characterized by the highest intervention coverage, the arms in which the CORPs program was rolled out in conjunction with birth kits and/or community dramas. Again, the health educator program was more effective when implemented in tandem with other health interventions.

4.2.2 Maternal and newborn practices

A second objective of the CORPs program was to enhance health practices related to maternal and newborn care, and Table 5 reports the estimated effects of the intervention on these outcomes. Two of the CORPs' specific learning objectives included encouraging pregnant women to develop a birth plan and providing information about the importance of breastfeeding. In Column (1), we find evidence of a significant and positive effect on

[14] In practice, facility delivery and skilled attendance at birth are almost equivalent, given that health personnel generally do not attend home births in this region.

the probability the respondent reported a birth plan in all treatment arms (an increase of between 21 and 40 percent relative to the control mean of 11%). We do not find that the intervention had any impact on breastfeeding behavior. In general, the sample reports high rates of breastfeeding initiation within 24 hours (89% in the control group), but relatively low rates of exclusive breastfeeding even within the first three days (58% in the control group); the average duration of breastfeeding is more than a year. These patterns are not shifted by the intervention.

However, we do find significant increases in the number of immunizations administered to newborns in the first month of life in the birth kit and drama arms (an additional 0.1 immunization administered on average, relative to a control mean of one), and a significant increase in the probability of an newborn check-up in the first month (an increase of around 12 percentage points, relative to the control mean of 29%). These are proportionally large effects, suggesting an increase of roughly 40% in infant check-ups. Again, we report the average standardized treatment effect for these variables in Column (7), and observe that there is an enhancement in maternal and newborn health practices of roughly 0.06 standard deviations compared to the control group, significant in all three treatment arms.

To sum up, the results presented in Tables 4 and 5 suggest that the interventions generated a substantial increase in health care utilization as well as enhanced health practices. These effects were generally larger in the treatment arms characterized by relatively higher coverage, the birth kit and the drama treatment arms.

4.3 Maternal and child morbidity, mortality and anthropometrics

Next, we investigate if these increases in utilization of maternal and newborn health services led to shifts in self-reported morbidity and mortality and anthropometrics Table 6 reports the results for maternal and neonatal mortality, stillbirths, and infant and neona-

tal mortality. There is no evidence of any effect of the interventions on self-reported measures of maternal morbidity during pregnancy, delivery, and the postpartum period or neonatal morbidity; it should be noted that the sample for neonatal morbidity is restricted, given that this information was reported only in the 28-day survey.[15] The average standardized effect for the morbidity indices is reported in Column (3), and also shows no significant effects. We also fail to identify any significant impact on the probability of stillbirths and infant and neonatal mortality; the mortality rates are calculated at the cluster level, yielding a sample of 96 observations.

We also evaluate the effects of the interventions on the anthropometric status of all children under five for respondents reporting a birth in the intervention period, as well as the subsample of children under one.[16] The final sample thus includes 5332 children with data on anthropometric measures reported at endline. In Table 7, we report the results of estimating equation (1) for height-for-age, weight-for-age, and mid-upper-arm-circumference-for-age. Here, we find evidence of a somewhat heterogenous pattern. Children in the CORPs-only and birth kit arms are characterized by higher height-for-age relative to sampled children in control communities, and this difference is statistically significant at the five percent level. However, the results for weight-for-age are insignificant or negative. The standardized effects reported in Column (5) are generally insignificant.[17]

[15]We also collected limited data on miscarriages at endline, and find no significant effect of the interventions on the rate of miscarriage, defined as the loss of a pregnancy in the first two trimesters.

[16]Endline anthropometric data is missing for children corresponding to births observed only in ongoing surveys (133 observations), as well as for an additional 650 observations; for the latter subsample, the adult respondent was surveyed in the endline, but the enumerator assigned to follow up with a separate anthropometric survey did not locate the household, the respondent declined to provide consent for measurement, or the child was not available.

[17]It is important to emphasize that the results should be interpreted cautiously given that the children observed in the anthropometric data are drawn from a subsample of respondents. However, we present evidence in Section 5 that respondents observed only in the endline and those observed in ongoing surveys are not characterized by significant differences in observable characteristics, and thus we cannot reject the hypothesis that the respondents observed in the anthropometric data constitute a random subsample.

4.4 Knowledge and attitudes

In addition to the primary outcomes enumerated above, we analyze a set of additional secondary outcomes: knowledge and attitudes around facility delivery care and knowledge about infant care and fertility. The CORPs intervention was designed to increase awareness and attitudes around maternal and infant health care. The results of estimating equation (1) for these variables are reported in Table 8; in Panel A, we use the full sample, and in Panel B, we use the subsample of respondents reporting a birth during the intervention period. The first four attitudinal indices (preference for facility-based delivery care, knowledge of relative risk around delivery options, knowledge of delivery complications, and knowledge of infant care) are reported only in the endline for 6350 respondents. The fifth index, attitudes toward the facility, was reported only in the 28-day survey for women who had recently given birth, and is thus a limited sample. The fertility variables are reported for all respondents represented in any follow-up data collection.

All the indices are coded such that a higher value indicates more knowledge or more positive attitudes. In Panel A, we find that the interventions increased knowledge of relative risk and pregnancy complications and enhanced attitudes toward health facility use. The average standardized effect for the attitudinal variables, reported in Column (6), shows increases of around 0.05 standard deviations that are statistically significant in all three arms.

In Panel B, we report the same outcomes for the sample of women reporting a birth during the intervention period, and the coefficients of interest are noisier; the average standardized effect is significant only for the birth kits arm and marginally insignificant at the 10 percent level for the CORPs-only arm. The fact that there is some effect observed among women who did not report a birth, and thus were not directly exposed to the CORPs intervention themselves, suggests that there were some positive spillovers of the intervention to other women of reproductive age.

In Table 9, we analyze whether the intervention generated any effects on fertility, including the probability of giving birth and the number of births. Here, the coefficients

of interest are small in magnitude and insignificant, consistent with the hypothesis that the intervention did not generate any differential patterns of selection into the subsample of women reporting a birth. Given that the CORPs were not trained or mandated to target any reductions in fertility, the absence of any significant effect here is unsurprising.

Summing up We conclude that the interventions achieved meaningful coverage rates in a challenging context characterized by ongoing civil unrest, and effectively targeted between 25% and 35% of women reporting births in this period. The intervention communities also exhibited increased utilization of antenatal and postnatal care, increased immunizations and newborn check-ups, and shifts in attitudes toward utilization of facilities for formal health care. These are important findings, particularly given the baseline utilization rates are extremely low. We find no evidence of any reduction in maternal or neonatal morbidity or mortality, and little evidence of enhancement in child anthropometric status.[18]

Comparing the coefficients estimated across the different treatment arms, in general we observe larger coefficients and more significant impact in the birth kits and drama arms compared to the pure CORPs treatment arm, though in many cases the difference between the estimated effects is not statistically significant. These arms also showed evidence of more active CORPs: health educators appeared to be more productive in contexts in which their program was linked to other interventions. Accordingly, the observed pattern is consistent both with a greater effect of a more intensive CORPs program, and an additional positive effect of the additional interventions.

We further explore whether there is any evidence of heterogeneous effects, utilizing the intent-to-treat specification and a pooled treatment dummy equal to one, and focusing on four characteristics of interest: parity of the birth, polygamous status of the household, an index of decision-making power for the mother, and a dummy for the straight-line

[18]All of the preceding results can be re-estimated including baseline measures of the outcomes of interest as additional controls when available, with the caveat that baseline measures of maternal care utilization and maternal and infant health are not reported at baseline for women who did not report a recent birth. The results are robust to including baseline control variables.

distance between the household and the primary health center exceeding 10 kilometers. We do not find any evidence of any heterogeneous response to the intervention.[19]

4.5 Intervention effects conditional on CORPs interaction

To evaluate the effects of the interventions on women who did in fact receive the interventions, we estimate treatment on the treated specifications. We focus on the CORPs intervention given that this intervention was implemented across all three treatment arms. The specification of interest can be written as follows

$$X_{icg} = \beta_1 CORPsDummy_{cg} + \chi_{icg}\gamma + \mu_g + \epsilon_{icg} \tag{2}$$

where χ_{icg} denotes the same vector of control variables utilized in the intent-to-treat estimates, and the dummy variable for exposure to the CORPS intervention is instrumented by a dummy variable for assignment to any treatment arm.[20]

The results of two-stage least squares estimation for care utilization, health practices, and attitudes and knowledge — the primary outcomes for which significant ITT estimates are observed — are reported in Table 10.[21] The estimated treatment on the treated coefficients are large in magnitude; the probability of accessing antenatal care services increases by more than 50 percentage points, and the probability of a postnatal care visit increases by 32 percentage points. Maternal and newborn health practices are also markedly improved for women reporting exposure to the CORPs intervention, with the probability of a birth preparedness plan increasing by 45 percentage points, a newborn receiving 0.7 more immunizations on average, and the probability of a neonatal check-up

[19]Distances are calculated using GPS coordinates of both locations. The decision-making index is constructed using five questions in the baseline survey asking who makes a series of decisions (the respondent alone, her husband alone, or both), with respect to how money is used, visits to the respondent's family, children's health, children's education, and children's discipline. We construct a series of dummy variables equal to one if the respondent has some decision-making power (i.e., if the husband does not make the decision alone), and then generate an index that is the mean of these five variables.

[20]In the first stage, the coefficient on the treated dummy is .101, significant at the one-percent level.

[21]Estimating the treatment on the treated specifications for the outcomes reported in Table 6 shows null effects. Limited power is available to estimate treatment on the treated for the anthropometric variables given the reduced sample.

increasing by 80 percentage points. There is also evidence of enhanced knowledge and attitudes toward maternal health.

These results provides suggestive evidence that a community health worker program that was more aggressively implemented could have very large effects on care utilization even in an extremely low-resource setting. In addition, this evidence highlights that given the coverage rate of the intervention, the effects on care utilization achieved were proportionately very substantial.

5 Robustness checks

Here, we report two robustness checks on the primary results. We analyze whether we observe non-random selection into the 3-day, 28-day, and audit surveys, and also present additional evidence around attrition.

5.1 Selection into the survey sample

While 90% of the baseline sample was re-surveyed in the endline survey, the surveys conducted during the ongoing data collection period (3-day, 28-day, and audit surveys) are available for a smaller sample. In addition, a reduced sample of households is represented in the endline anthropometrics data. Accordingly, we can compare the baseline characteristics of women included in these various data sources to evaluate whether they differ significantly on observables. We also analyze whether the probability of inclusion differed across treatment arms.

To implement this test, we utilize the full sample of all births observed and estimate the following specifications, regressing dummy variables for whether the birth was observed in a given survey ($Survey_{picg}$) on the three treatment dummies and two sets of baseline characteristics: baseline demographics as reported in Panel A of Table 1, and baseline care utilization as reported in Panel A of Table 2. The baseline care utilization variables are available only for women who reported a recent birth (within the last two

years) at baseline.

$$Survey_{bicg} = \beta_1 CORPS_{cg} + \beta_2 Kits_{cg} + \beta_3 Drama_{cg} + \beta\chi_{icg} + \mu_g + \epsilon_{bicg} \quad (3)$$

Table B3 presents the results. We can observe in Columns (1) to (3) that women in the CORPs-only arm were slightly less likely to be surveyed during ongoing data collection, while educated and married women characterized by higher birth parity at baseline are more likely to be surveyed. This pattern may reflect these women's higher status within the community; enumerators relied on local informants to assist in identifying the women in the sample, and they may have found it easier to locate women who were better known. In Columns (4) to (6), the results suggest that only significant differences between women who were and were not included in anthropometric data collection at endline is age and the probability of Muslim identity. Importantly, however, there is no evidence that baseline care utilization predicts selection into any form of data collection.

Data from the ongoing surveys was used to construct outcome measures employing two strategies. First, whenever the same outcome was measured in both the endline and an ongoing survey, we preferentially employed the latter report given that this data was collected closer to the birth and is presumably characterized by more limited recall bias. Second, for women who were observed only at endline, the outcome variables were constructed using the endline data. Given the characteristics of the women interviewed during the ongoing data collection, we can conclude that recall bias may be slightly lower for married and educated women who had slightly larger families. However, given that these women do not show evidence of significantly different care-seeking behavior at baseline, and all regressions include a large set of baseline demographic controls, we do not anticipate these differences will generate significant bias in the primary results.

5.2 Attrition

As previously noted, 10.2% of baseline respondents (719 women) were not observed in the endline survey, and 8.1% of respondents (575 women) were never observed again after the baseline survey. Given the challenging setting, characterized by ongoing civil unrest and violence, these attrition rates are relatively low. However, attrition may pose a threat to internal validity if respondents attriting from the sample in treatment communities differ systematically from those attriting in control communities. We conduct two tests to evaluate whether differential attrition could be a source of bias.

First, we report in Tables B6 and B7 in the Appendix the percentage of respondents reached at follow-up in each of the four experimental arms, as well as a comparison of baseline characteristics among non-attriters across arms. As before, we first analyze a series of basic socioeconomic characteristics, and then analyze the baseline values of the primary outcome variables of interest. It is evident that there are no significant differences between the proportion of respondents reached at follow-up across treatment arms. In addition, as with the balance analysis of the entire sample, there are in general very few significant differences in baseline characteristics across treatment arms, and the joint F-test again fails to reject the hypothesis of balance.

Second, we estimate the following specification to test whether baseline characteristics predict attrition from the sample, and whether this relationship differs significantly in treatment and control arms. A dummy for a respondent attriting from the sample $Attritted_{icg}$ is regressed on a treatment dummy, baseline characteristics, and the interaction between the two, again conditional on LGA fixed effects and using standard errors clustered at the cluster level. Again, we focus on baseline demographic characteristics and baseline reports of care utilization.

$$Attritted_{icg} = \beta_1 Treated_{cg} + \beta_2 BaseX_{icg} + \beta_3 Treated_{cg} * BaseX_{icg} + \mu_g + \epsilon_{icg} \quad (4)$$

The results are reported in Table B8 in the Appendix. Some demographic character-

istics are predictive of attrition on average. In general, however, there is limited evidence of any differences between attriters in treatment and control arms; only one estimated coefficient β_3 is significant at conventional levels, for postnatal care as reported in Column (5) of Panel C. This suggests the scope for bias due to differential attrition is limited.

6 Conclusion

Over the last twenty years, progress in reducing maternal and neonatal deaths in developing countries has largely stagnated, particularly in sub-Saharan Africa. Accordingly, identifying new interventions that can increase utilization of maternal health services and enhance health outcomes, particularly in challenging regions characterized by extremely low baseline human capital outcomes and persistent violence, is a key priority for policymakers and researchers.

This evaluation adds to the body of evidence around community-level interventions designed to improve maternal and child health in low resource settings, evaluating the effects of three interventions targeted to increase utilization of maternal health services in rural northern Nigeria. This evaluation was a randomized controlled trial analyzing interventions including the deployment of community resource persons to provide health education door to door to pregnant women, the distribution of safe birth kits, and the introduction of community drama activities promoting safe motherhood. To our knowledge, this was one of the first health-focused RCTs ever conducted in Nigeria, and one of the first community health worker programs evaluated in a conflict-affected region characterized by extremely poor health outcomes for women and children.

The results suggest that the interventions successfully reached approximately 22% of women reporting a birth during this period, and generated increased utilization of antenatal care, postnatal care, immunizations, and newborn check-ups in communities exposed to the intervention. We also find positive shifts in attitudes around facility-based delivery care. Given a challenging context and the presence of ongoing civil unrest, the coverage

rates achieved by the intervention are substantial, and clearly suffice to generate meaningful effects on health behaviors, knowledge and attitudes around maternal and health practices. However, these shifts in utilization did not yield any detectable improvements in anthropometric outcomes for infants or young children, or any decreases in maternal or neonatal morbidity or mortality. We hypothesize that the absence of any significant effect on health outcomes may reflect the previously documented low returns to utilization of formal health care in this setting. Alternatively, the increases in utilization may not be large enough in magnitude to generate any detectable shifts in health outcomes in the relatively short two-year period. We also find that the community health program was most effective when implemented in conjunction with additional health programs — the birth kits and the drama interventions.

Ultimately, the observed pattern is consistent with the hypothesis that demand-side interventions such as community health educator programs can effectively stimulate increased utilization and change health behaviors, even in a challenging context. Future research might explore further how to increase uptake and impact by linking community health worker programs to other interventions, and strengthening the implementation of community-based interventions in challenging environments.

References

Akinyemi, Joshua Odunayo, Elijah Afolabi Bamgboye, and Olusola Ayeni, "Trends in neonatal mortality in Nigeria and effects of bio-demographic and maternal characteristics," *BMC Pediatrics*, 2012, *15.*

Björkman Nyqvist, Martina, Andrea Guariso, Jakob Svensson, and David Yanagizawa-Drott, "Reducing child mortality in the last mile: Experimental evidence on community health promoters in Uganda," *American Economic Journal: Applied Economics*, 2019, *11* (3), 155–92.

Chari, A.V. and Edward Okeke, "Can institutional deliveries reduce newborn mortality? Evidence from Rwanda," 2014. RAND Working Paper.

Darmstadt, Gary L., Mohamed Hassan, Zohra P. Balsara, Peter J. Winch, Reginald Gipson, and Mathuram Santosham, "Impact of clean delivery-kit use on newborn umbilical cord and maternal puerperal infections in Egypt," *Journal of Health, Population and Nutrition*, 2009, *27*, 746–754.

Duflo, Esther, Rachel Glennerster, and Michael Kremer, "Chapter 61 Using Randomization in Development Economics Research: A Toolkit," in T. Paul Schultz and John A. Strauss, eds., *T. Paul Schultz and John A. Strauss, eds.*, Vol. 4 of *Handbook of Development Economics*, Elsevier, 2007, pp. 3895 – 3962.

EIU Canback, "C-GIDD (Canback Global Income Distribution Database)," 2016. Retrieved 1-16-2016.

Frankenberg, Elizabeth, Wayan Suriastini, and Duncan Thomas, "Can expanding access to basic healthcare improve children's health status? Lessons from Indonesia's midwife in the village programme," *Population Studies*, 2005, *59.*

Gabrysch, Sabine and Oona MR Campbell, "Still too far to walk: Literature review of the determinants of delivery service use," *BMC Pregnancy and Childbirth*, 2009, *9* (34).

Godlonton, Susan and Edward Okeke, "Does a ban on informal health providers save lives? Evidence from Malawi," *Journal of Development Economics*, 2016, *118* (1), 112–131.

Hogan, Margaret, Kyle Foreman, Mohsen Naghavi, Stephanie Ahn, Mengru Wang, Susanna Makela, Alan Lopez, Rafael Lozano, and Christopher Murray, "Maternal mortality for 181 countries, 1980–2008: A systematic analysis of progress toward Millennium Development Goal 5," *Lancet*, 2010, *375* (9726), 1609–23.

Hundley, Vanora A. and Bilal Avan, "Are birth kits a good idea? A systematic review of the evidence," *Midwifery*, April 2012, *2*, 204–15.

Hundley, Vanora, Bilal Avan, Haris Ahmed, and Wendy Graham, "Clean birth kits to improve birth practices: development and testing of a country level decision support tool," *BMC Pregnancy and Childbirth*, 2012, *12.*

Kidane, Gebreyesus and Richard Morrow, "Teaching mothers to provide home treatment of malaria in Tigray, Ethiopia: A randomized trial," *Lancet*, 2000, *12* (356), 550–555.

Kling, Jeffrey R., Jeffrey B. Liebman, Lawrence F. Katz, and Lisa Sanbonmatsu, "Moving to opportunity and tranquility: Neighborhood effects on adult economic self-sufficiency and health from a randomized housing voucher experiment," 2004. Princeton University Working Paper No. 5.

Leight, Jessica, Martina Björkman Nyqvist, and Vandana Sharma, "AEA RCT Registry," 2016. https://www.socialscienceregistry.org/trials/6/history/10566.

Lewin, Simon, Susan Munabi-Babigumira, Claire Glenton, Karen Daniels, Xavier Bosch-Capblanch, Brian E. van Wyk, Jan Odgaard-Jensen, Marit Johansen, Godwin N. Aja, Merrick Zwarenstein, and Inger B. Scheel, "Lay health workers in primary and community health care for maternal and child health and the management of infectious diseases (Review)," *The Cochrane Library*, 2010, *3.*

NPC and ICF International, "Nigeria Demographic and Health Survey," 2014.

Okeke, Edward, Peter Glick, Isa Sadeeq Abubakar, A.V. Chari, Emma Pitchforth, Josephine Exley, Usman Bashir, Claude Messan Setodji, Kun Gu, and Obinna Onwujekwe, "The Better Obstetrics in Rural Nigeria (BORN) Study: An Impact Evaluation of the Nigerian Midwives Service Scheme," 2015. RAND Research Report.

_ , _ , _ , _ , _ , _ , _ , _ , _ , _ , **and** _ , "The effect of increasing the supply of skilled health providers on pregnancy and birth outcomes: Evidence from the midwives service scheme in Nigeria," *BMC Health Services Research*, 2016, *16* (425).

Okwundu, Charles I, Sukrti Nagpal, Alfred Musekiwa, and David Sinclair, "Home- or community-based programmes for treating malaria (Review)," *The Cochrane Library*, 2013, *5.*

PPFN, "PPFN Programme Implementation Plan," 2012.

_ , "PPFN Training Manual for CORPs," 2012.

Sharma, Vandana, Willa Brown, Muhammad Kainuwa, Jessica Leight, and Martina Björkman Nyqvist, "High maternal mortality in Jigawa state, northern Nigeria estimated using the sisterhood method," *BMC Pregnancy and Childbirth*, 2017, *17* (163).

Simkhada, Bibha, Edwin R. van Teijlingen, Maureen Porter, and Padam Simkhada, "Factors affecting the utilization of antenatal care in developing coun-

tries: Systematic review of the literature," *Journal of Advanced Nursing*, 2008, *61* (3).

Sloan, Nancy L., Salahuddin Ahmed, Satindra N. Mitra, Nuzhat Choudhury, Mushtaque Chowdhury, Ubaider Rob, and Beverly Winikoff, "Community-based kangaroo mother care to prevent neonatal and infant mortality: A randomized, controlled cluster trial," *Pediatrics*, 2008, *121* (5).

WHO, "Essential newborn care: Report of a technical working group," Technical Report 1996.

WHO, "Maternal mortality fact sheet," 2016. Available at http://www.who.int/mediacentre/factsheets/fs348/en/, accessed February 23, 2018.

Winani, Samson, Siri Wood, Patricia Coffey, Tobias Chirwa, Frank Mosha, and John Changalucha, "Use of a clean delivery kit and factors associated with cord infection and puerperal sepsis in Mwanza, Tanzania," *Journal of Midwifery and Women's Health*, 2007, *52*, 37–43.

Figure 1: Timeline

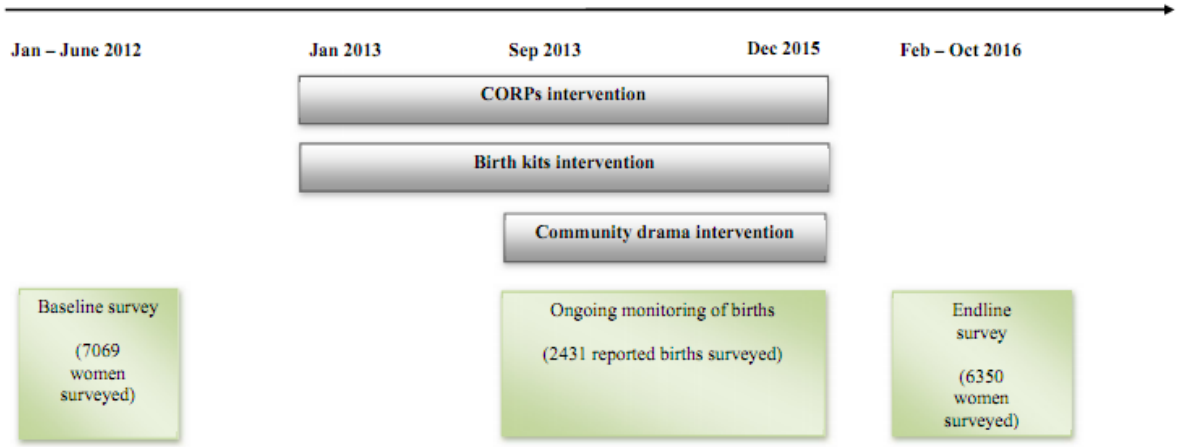

Figure 2: Trial profile

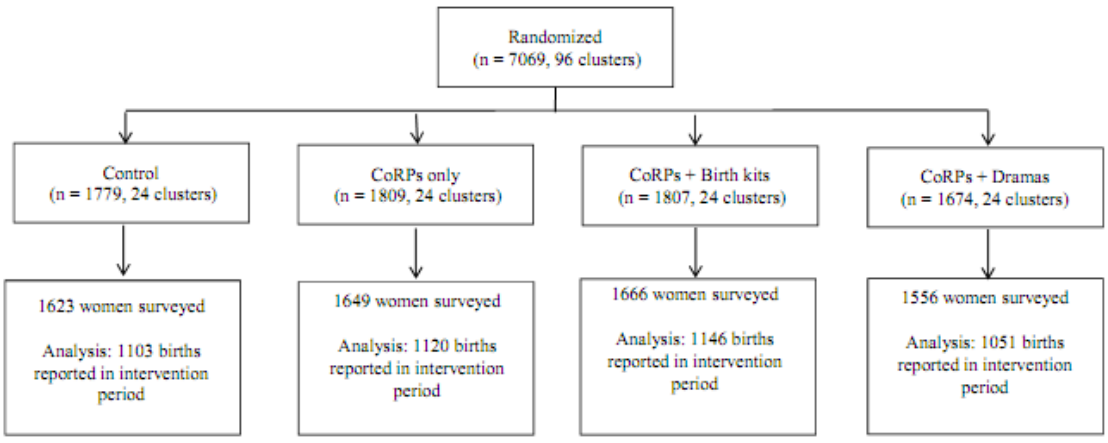

Table 1: Baseline socioeconomic characteristics

Variables	Control (N=1779)	CORPs only (N=1809)	CORPs + birth kits (N=1807)	CORPs + drama (N=1674)	Joint p-value	Obs.
Panel A: Demographic characteristics						
Married	0.99	0.99	1.00	0.99	.50	7069
Number of other wives	0.33	0.38	0.35	0.34	.36	7069
Age at marriage	15.02	15.17	15.19	15.23	.68	7052
Number of marriages	1.24	1.24	1.18	1.23	.00	7063
Age	27.99	27.67	27.79	27.98	.53	7069
Ever attended school	0.16	0.15	0.19	0.18	.20	7069
Literate in Hausa	0.08	0.09	0.11	0.10	.33	7069
Muslim	1.00	1.00	1.00	1.00	.01	7069
Birth parity	4.19	4.17	4.23	4.28	.60	7069
Wealth index	-0.03	0.04	-0.05	0.05	.71	7069
Panel B: Household income and consumption						
Head attended school	0.27	0.31	0.32	0.33	.89	6625
Head's highest educ. (primary)	0.13	0.14	0.13	0.14	.75	6625
Head's highest education - second. (secondary)	0.10	0.11	0.11	0.12	.59	6625
Head's occupation (Own cultivation)	0.51	0.51	0.51	0.49	.93	7031
Head's occupation (Non-farm self-emp.)	0.28	0.29	0.29	0.30	.49	7031
Head's occupation (Outside employment)	0.11	0.12	0.13	0.12	.94	7031
Owns land	0.95	0.94	0.94	0.94	.73	7069
Number plots	3.74	3.63	3.63	3.50	.50	6679
Cultivates any millet	0.90	0.89	0.87	0.88	.59	7069
Cultivates any guinea corn	0.87	0.85	0.84	0.84	.70	7069
Cultivates any beans	0.77	0.74	0.74	0.74	.91	7069
Cultivates any ground nut	0.48	0.44	0.40	0.48	.06	7069

Notes: This table reports the mean values of household demographic characteristics as reported at baseline for households in each experimental arm. We also estimate a regression in which each demographic characteristic is regressed separately on three dummy variables for assignment to each treatment arm, as well as LGA fixed effects; standard errors are clustered at the cluster level. The reported p-value is the p-value on the joint test $\beta_1 = \beta_2 = \beta_3$, where the three coefficients refer to the coefficients on each treatment dummy variable.

Table 2: Baseline health care utilization, health practices and health outcomes

Variables	Control (N=1779)	CORPs only (N=1809)	CORPs + birth kits (N=1807)	CORPs + drama (N=1674)	Joint p-value	Obs.
Panel A: Utilization of maternal health care						
Any antenatal care visit	0.47	0.53	0.56	0.57	.55	4007
Number of antenatal visits	1.95	2.23	2.42	2.41	.46	4007
Antenatal quality index	0.26	0.31	0.34	0.34	.41	4007
Care-seeking for complications	0.47	0.47	0.52	0.53	.26	2625
Facility delivery	0.07	0.10	0.09	0.10	.63	3742
Skilled attendant at birth	0.08	0.11	0.10	0.13	.16	3742
Post-natal check-up	0.25	0.25	0.30	0.28	.10	3742
Panel B: Maternal and newborn health practices						
Male involvement in pregnancy	0.57	0.56	0.59	0.59	.56	3837
Infant breastfed (first day)	0.82	0.82	0.82	0.83	.75	3642
Excl. breastfeeding (first three days)	0.31	0.30	9.34	0.28	.05	3642
Panel C: Infant and child anthropometrics						
Weight-for-age (< 1 year)	0.15	-0.21	-0.36	0.21	.01	2042
Height-for-age (< 1 year)	0.53	0.62	0.33	0.87	.36	2009
MUAC-for-age (< 1 year)	-0.74	-0.80	-0.90	-0.92	.41	1590
Weight-for-age (< 2 year)	-0.43	-0.68	-0.76	-0.30	.00	3382
Height-for-age (< 2 year)	0.29	0.24	0.16	0.58	.06	3312
MUAC-for-age (< 2 year)	-0.89	-0.86	-0.91	-0.94	.45	2915
Panel D: Maternal morbidity						
Any complication (during pregnancy)	0.55	0.57	0.56	0.54	.30	4007
Any complication (during delivery)	0.06	0.06	0.08	0.08	.17	3742
Any complication (during postpartum)	0.21	0.25	0.22	0.22	.59	3742
Joint F-test (All baseline characteristics in Table 1 and 2)					.59	

Notes: This table reports the mean values of household characteristics as reported at baseline for households in each experimental arm; the variables reported are constructed to be identical to the outcome variables of interest subsequently analyzed in Tables 4 through 7. We also estimate a series of regressions in which each demographic characteristic is regressed separately on three dummy variables for assignment to each treatment arm, as well as LGA fixed effects; standard errors are clustered at the cluster level. The reported p-value is the p-value on the joint test $\beta_1 = \beta_2 = \beta_3$, where the three coefficients refer to the coefficients on each treatment dummy variable.

Table 3: Intervention exposure

	(1) Any CORPs visit	(2) Number CORPs visits	(3) Kit received	(4) Kit knowledge	(5) Name an object	(6) Kit used	(7) Dramas conducted	(8) Dramas attended
CORPS only	.048***	.115***	.005	.003	.005	.007	.003	.011
	(.018)	(.044)	(.008)	(.006)	(.007)	(.006)	(.018)	(.017)
CORPs + Birth kits	.126***	.188***	.088***	.066***	.078***	.052***	.096***	.082***
	(.022)	(.046)	(.018)	(.012)	(.016)	(.012)	(.023)	(.021)
CORPs + Dramas	.100***	.152***	.002	.005	-.003	-.003	.250***	.209***
	(.018)	(.030)	(.010)	(.007)	(.009)	(.007)	(.026)	(.025)
Test $\beta_1 = \beta_2$	.000	.145	.000	.000	.000	.000	.000	.001
Test $\beta_1 = \beta_3$	.006	.328	.717	.84	.313	.131	.000	.000
Test $\beta_2 = \beta_3$	.214	.373	.000	.000	.000	.000	.000	.000
Test $\beta_1 = \beta_2 = \beta_3 = 0$	.000	.000	.000	.000	.000	.000	.000	.000
Test $\beta_1 = \beta_{enhanced}$	.000	.162	.001	.000	.003	.023	.000	.000
Control mean	.13	.16	.01	.01	0	0	.09	.05
Obs.	4420	4420	4420	4420	4420	4420	4420	4420

Notes: All regressions include LGA fixed effects and standard errors clustered at the cluster level. The sample includes all respondents reporting a pregnancy during the intervention period who were interviewed in at least one follow-up survey. Control variables include the following variables measured at baseline: a dummy variable for whether the respondent is married, the number of other wives the respondent's husband has, age at marriage, the number of marriages reported, age, a dummy for whether the respondent has ever attended school, a dummy variable for whether the respondent reads Hausa, a dummy variable for Muslim, birth parity at baseline, and a wealth index. We also include dummy variables equal to one if the respondent is observed in the 3-day and 28-day surveys, and in the audit survey. The variables reported include a dummy variable for a respondent receiving a CORPs visit and the number of visits; dummy variables for a respondent receiving a birth kit, reporting that she knows how to use the kit, naming at least one item in the kit, and reporting that she utilized the kit in her most recent delivery; and dummy variables equal to one if the respondent reports that PPFN conducted at least one community drama and if she attended at least one drama. We report tests of equality across the estimated coefficients; a test of the hypothesis that the joint effect of treatment is zero ($\beta_1 = \beta_2 = \beta_3 = 0$; and a test of equality of impact across the basic arms and the arms including additional interventions (drama and birth kits) $\beta_1 = \beta_{enhanced}$. Asterisks denote significance at the ten, five and one percent level.

Table 4: Utilization of maternal health care

	(1) Any ANC visits	(2) Number ANC visits	(3) ANC quality index	(4) Postnatal care	(5) Avg. std. effect Cols. 1-4	(6) Facility birth	(7) Skilled attendance	(8) Birth accompanied	(9) Avg. std. effect
CORPS only	.006	-.100	.002	.023*	.018	-.027	-.009	.010	-.001
	(.036)	(.168)	(.024)	(.013)	(.055)	(.021)	(.018)	(.023)	(.039)
CORPs + Birth kits	.063*	.281*	.053**	.026	.125	-.003	.018	-.003	.077
	(.034)	(.160)	(.021)	(.017)	(.050)**	(.024)	(.021)	(.022)	(.038)**
CORPs + Dramas	.086***	.352**	.059***	.033**	.154	-.026	.005	-.029	.072
	(.033)	(.171)	(.021)	(.013)	(.052)***	(.020)	(.019)	(.027)	(.037)**
Test $\beta_1 = \beta_2$	.084	.015	.016	.834	.036	.328	.213	.606	.047
Test $\beta_1 = \beta_3$	.017	.009	.01	.383	.012	.951	.453	.184	.059
Test $\beta_2 = \beta_3$	.461	.67	.753	.676	.552	.339	.56	.341	.896
Test $\beta_1 = \beta_2 = \beta_3 = 0$	.019	.016	.004	.110	.004	.45	.644	.606	.047
Test $\beta_1 = \beta_{enhanced}$	.021	.004	.006	.553	.009	.491	.218	.271	.026
Control mean	.63	2.66	.38	.07		.18	.12	.59	
Obs.	4420	4420	4420	3684		4009	3649	3649	

Notes: All regressions include LGA fixed effects and standard errors clustered at the cluster level. The sample includes all respondents reporting a pregnancy during the intervention period who were interviewed in at least one follow-up survey and who report the indicator of interest. The control variables in each regression are identical to those reported in Table 3. The variables reported in Columns (1) through (3) include a dummy variable for any reported ANC visits, the reported number of visits and an index of antenatal care quality. (The quality index is the mean of a dummy for utilizing antenatal care in the first trimester; a dummy for receipt of more than half of the possible components of antenatal care; a dummy for receipt of the tetanus vaccine; a dummy for the receipt of iron folic pills; and a dummy for receipt of advice around danger signs during pregnancy.) The variable reported in Column (4) is a dummy for care-seeking conditional on reported pregnancy complications; the variable reported in Column (5) is a dummy variable for receipt of a postpartum check-up within two months. Column (6) reports the ASTE for the preceding columns. The variables reported in Columns (7) through (9) include a dummy variable for facility-based delivery; a dummy variable for skilled attendance at birth; and a dummy variable equal to one if the woman reports delivering accompanied by another individual. Column (10) reports the ASTE for all variables reported in the table. We report tests of equality across the estimated coefficients; a test of the hypothesis that the joint effect of treatment is zero ($\beta_1 = \beta_2 = \beta_3 = 0$; and a test of equality of impact across the basic and enhanced arms $\beta_1 = \beta_{enhanced}$. Asterisks denote significance at the ten, five and one percent level.

Table 5: Maternal and newborn health practices

	(1) Birth plan	(2) Breast-feeding	(3) Breastfeeding exclusively	(4) Breastfeeding duration	(5) Immunizations	(6) Infant check-up	(7) Avg. std. effect
CORPS only	.046***	-.028	.016	-1.283	.015	.117***	.056**
	(.017)	(.018)	(.014)	(1.727)	(.064)	(.041)	(.027)
CORPs + Birth kits	.023	-.005	.004	-1.657	.091*	.119***	.067**
	(.014)	(.016)	(.018)	(1.854)	(.052)	(.044)	(.027)
CORPs + Dramas	.045***	-.014	.022	.006	.131*	.054	.073**
	(.014)	(.017)	(.017)	(1.672)	(.070)	(.042)	(.030)
Test $\beta_1 = \beta_2$	.17	.143	.458	.818	.234	.965	.670
Test $\beta_1 = \beta_3$	.957	.403	.697	.375	.133	.097	.568
Test $\beta_2 = \beta_3$	.111	.579	.37	.323	.523	.100	.839
Test $\beta_1 = \beta_2 = \beta_3 = 0$	.005	.398	.516	.679	.176	.013	.044
Test $\beta_1 = \beta_{enhanced}$	.418	.197	.788	.746	.137	.415	.538
Control mean	.11	.89	.58	14.22	1.03	.29	
Obs.	4152	3051	3516	2803	1385	1317	

Notes: All regressions include LGA fixed effects and standard errors clustered at the cluster level. The sample includes all respondents reporting a pregnancy during the intervention period who were interviewed in at least one follow-up survey and who report the indicator of interest. The control variables in each regression are identical to those reported in Table 3. The variables reported include a dummy variable equal to one if the respondent reported developing a birth preparedness plan prior to delivery; a dummy for initiation of breastfeeding within 24 hours after birth, a dummy for exclusive breastfeeding within the first 3 days, breastfeeding duration (in months) as reported at the endline, the number of immunizations administered to the newborn in the first 60 days of life; and a dummy variable equal to one if the newborn received a facility-based health check-up during the first 60 days. We report tests of equality across the estimated coefficients; a test of the hypothesis that the joint effect of treatment is zero ($\beta_1 = \beta_2 = \beta_3 = 0$; and a test of equality of impact across the basic and enhanced arms $\beta_1 = \beta_{enhanced}$. Asterisks denote significance at the ten, five and one percent level.

Table 6: Maternal and neonatal morbidity and mortality

	(1)	(2)	(3)	(4)	(5)	(7)
	Maternal morbidity	Neonatal morbidity	Avg. std. effect Cols. 1-2	Stillbirth	Under 1 mortality rate	Neonatal mortality rate
CORPS only	-.012	-.003	-.015	.005		
	(.018)	(.030)	(.037)	(.006)		
CORPs + Birth kits	-.033**	.054	.019	.003		
	(.017)	(.034)	(.035)	(.005)		
CORPs + Dramas	.010	-.028	-.017	.001		
	(.019)	(.034)	(.039)	(.005)		
Treated					4.291	6.761
					(7.226)	(5.985)
Test $\beta_1 = \beta_2$	.216	.075	.370	.794		
Test $\beta_1 = \beta_3$	.256	.44	.958	.533		
Test $\beta_2 = \beta_3$	.023	.024	.380	.696		
Test $\beta_1 = \beta_2 = \beta_3 = 0$	.097	.135	.770	.812		
Test $\beta_1 = \beta_{enhanced}$	.98	.475	.583	.636		
Control mean	.59	.5		.02	43.892	18.557
Obs.	4420	1288		4420	96	96

Notes: All regressions include LGA fixed effects and standard errors clustered at the cluster level. The sample includes all respondents reporting a pregnancy during the intervention period who were interviewed in at least one follow-up survey. The control variables in each regression are identical to those reported in Table 3. The morbidity indices are defined as dummy variables equal to one if the respondent reported any of a series of enumerated symptoms in each specified period (pregnancy, delivery and postpartum), or if she reports symptoms for her infant. For pregnancy, the enumerated symptoms include convulsions, swelling of legs, body, or face, excessive fatigue, vaginal bleeding, trouble with vision during daylight, night blindness, high blood pressure, and any other complication. For delivery, the enumerated symptoms include excessive bleeding, fits or convulsions not caused by fever, labor longer than 12 hours, headache / blurred vision / high blood pressure, a high fever with bad-smelling vaginal discharge, the baby's hands or feet coming out first, and any other complication. For the postpartum period (the first two months after birth), the enumerated symptoms include bleeding, convulsions, swelling in the legs, face or hands, blurring of vision, unconsciousness, high fever, abnormal or smelly vaginal discharge, and serious abdominal pain. For neonates, the enumerated symptoms include loose watery stools, blood in the stool, persistent vomiting, rash, high fever, cough, difficulty breathing, weight loss, convulsions, discharge from the umbilicus, and any other complication. The stillbirth dummy is a dummy variable equal to one if the respondent reports a stillbirth during the intervention period. Infant and neonatal mortality rates are defined at the cluster level as the total number of deaths observed over the total number of births observed in the sample households over the study period. We report tests of equality across the estimated coefficients; a test of the hypothesis that the joint effect of treatment is zero ($\beta_1 = \beta_2 = \beta_3 = 0$; and a test of equality of impact across the basic and enhanced arms $\beta_1 = \beta_{enhanced}$. Asterisks denote significance at the ten, five and one percent level.

Table 7: Child anthropometrics

	(1) Height-for-age	(2) Weight-for-age	(3) MUAC-for-age	(4) Avg. std. effect
Panel A: Children under five				
CORPS only	.230**	.008	-.024	.038
	(.097)	(.056)	(.043)	(.031)
CORPs + Birth kits	.228***	.003	.020	.058
	(.085)	(.059)	(.046)	(.028)**
CORPs + Dramas	-.022	-.120*	-.062	.035
	(.095)	(.063)	(.049)	(.033)
Test $\beta_1 = \beta_2$	.979	.853	.366	.727
Test $\beta_1 = \beta_3$	.013	.029	.422	.027
Test $\beta_2 = \beta_3$	.004	.047	.113	.014
Test $\beta_1 = \beta_2 = \beta_3 = 0$	.004	.121	.414	.070
Test $\beta_1 = \beta_{enhanced}$	.061	.296	.819	.275
Control mean	-1.54	-1.19	-.75	
Obs.	5332	5332	5332	5040
Panel B: Children under one				
CORPS only	.431**	.020	-.096	.026
	(.202)	(.118)	(.095)	(.031)
CORPs + Birth kits	.504***	.077	.128	.037
	(.170)	(.120)	(.084)	(.033)
CORPs + Dramas	.234	.047	.003	-.049
	(.235)	(.118)	(.098)	(.034)
Test $\beta_1 = \beta_2$	.810	.304	.005	.467
Test $\beta_1 = \beta_3$	.683	.381	.088	.914
Test $\beta_2 = \beta_3$	.499	.912	.239	.409
Test $\beta_1 = \beta_2 = \beta_3 = 0$	.045	.537	.029	.218
Test $\beta_1 = \beta_{enhanced}$	.044	.341	.019	.740
Control mean	-1.23	-1.42	-.84	
Obs.	1614	1614	1266	

Notes: All regressions include LGA fixed effects and standard errors clustered at the cluster level. The sample includes children of respondents born in the intervention period (in Panel A) or under age one at endline (in Panel B) who have anthropometric measurements collected at endline. The control variables in each regression are identical to those reported in Table 3. The variables reported are Z-scores for height-for-age, weight-for-age, and mid-upper-arm circumference-for-age (MUAC-for-age) constructed using anthropometric data collected at endline and the World Health Organization child growth standards. We report tests of equality across the estimated coefficients; a test of the hypothesis that the joint effect of treatment is zero ($\beta_1 = \beta_2 = \beta_3 = 0$; and a test of equality of impact across the basic and enhanced arms $\beta_1 = \beta_{enhanced}$. Asterisks denote significance at the ten, five and one percent level.

Table 8: Health knowledge and attitudes

	Delivery preference	Knowledge relative risk	Knowledge compli-cations	Infant. care know.	Attitudes toward facility	Avg. std. effect
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample						
CORPS only	.023	-.006	.006	.001	.026**	.025
	(.014)	(.018)	(.007)	(.010)	(.011)	(.035)
CORPs + Birth kits	-.00003	.032**	.008	.002	.022**	.035**
	(.011)	(.013)	(.006)	(.010)	(.009)	(.024)
CORPs + Dramas	-.004	.026	.011*	-.004	.014	.049*
	(.012)	(.016)	(.006)	(.009)	(.011)	(.029)*
Test $\beta_1 = \beta_2$	.204	.019	.749	.949	.723	.305
Test $\beta_1 = \beta_3$	.06	.089	.432	.582	.41	.501
Test $\beta_2 = \beta_3$	.417	.638	.567	.528	.489	.711
Test $\beta_1 = \beta_2 = \beta_3 = 0$	.305	.026	.299	.917	.040	.103
Test $\beta_1 = \beta_{enhanced}$	.087	.03	.538	.804	.525	.357
Control mean	.75	.86	.62	.55	.45	
Obs.	6350	6350	6350	6350	1393	
Panel B: Women reporting pregnancies during the intervention period						
CORPS only	.029*	-.020	.006	-.003	.026**	.009
	(.017)	(.020)	(.008)	(.010)	(.011)	(.037)
CORPs + Birth kits	-.013	.027*	.010*	.003	.022**	.067**
	(.016)	(.015)	(.006)	(.011)	(.009)	(.026)
CORPs + Dramas	-.002	.017	.007	-.012	.014	.028
	(.015)	(.018)	(.007)	(.011)	(.011)	(.031)
Test $\beta_1 = \beta_2$	.035	.007	.523	.641	.723	.080
Test $\beta_1 = \beta_3$	.164	.071	.896	.446	.41	.614
Test $\beta_2 = \beta_3$	.385	.505	.565	.227	.489	.121
Test $\beta_1 = \beta_2 = \beta_3 = 0$	.206	.031	.413	.642	.040	.037
Test $\beta_1 = \beta_{enhanced}$	.048	.015	.679	.897	.525	.225
Control mean	.76	.87	.63	.56	.45	
Obs.	4161	4161	4161	4161	1393	

Notes: All regressions include LGA fixed effects and standard errors clustered at the cluster level. The control variables in each regression are identical to those reported in Table 3. The first index measuring preference for delivery location is equal to the simple mean of seven dummy variables taking the value one if the woman said she would deliver at a facility in a series of hypothetical scenarios. These scenarios include if there was funding, if the husband said the choice was hers, if the mother-in-law said the choice was hers, if a female midwife was available at the facility, if the midwife was welcoming, if she was sure the facility was open at any hour, or if she was sure the facility was properly equipped. The second index measuring relative risk perception is an average of dummy variables equal to one if the woman reported that an unattended delivery is more dangerous than delivery with a traditional birth attendant (TBA), home delivery is more dangerous than delivering at a facility, and delivery with a TBA is more dangerous than delivering at facility. The third index measuring knowledge of pregnancy complications is a mean of three variables counting the number of complications correctly identified for the pregnancy, delivery and postpartum period, and of three dummy variables equal to one if the respondent reports a woman could die from said complications. The fourth index measuring attitudes around the use of a health facility is equal to the mean of 11 dummies coded to indicate positive attitudes toward the facility, its staff, and individuals who use the facility. The fifth index measuring knowledge of infant health practices is equal to the mean of dummies for correct answers to a series of questions; this includes whether an infant should be breastfed immediately after birth, when complementary feeding should be initiated, how much food and water should be provided to an infant or child suffering from diarrhea, how many immunizations an infant should receive in the first two months of life, and what the signs of pneumonia are in an infant. In Panel A, the sample includes all women surveyed in the endline survey. In Panel B, the sample is restricted to women who reported a birth during the intervention period. Accordingly, the fertility variables are omitted, as there is essentially no variation in fertility in this subsample. We report tests of equality across the estimated coefficients; a test of the hypothesis that the joint effect of treatment is zero ($\beta_1 = \beta_2 = \beta_3 = 0$; and a test of equality of impact across the basic and enhanced arms $\beta_1 = \beta_{enhanced}$. Asterisks denote significance at the ten, five and one percent level.

Table 9: Fertility

	Number of births	Any births
CORPS only	.013	.007
	(.015)	(.014)
CORPs + Birth kits	.002	-.006
	(.015)	(.014)
CORPs + Dramas	-.0005	-.008
	(.017)	(.016)
Test $\beta_1 = \beta_2$	.394	.319
Test $\beta_1 = \beta_3$	.374	.338
Test $\beta_2 = \beta_3$	.886	.904
Test $\beta_1 = \beta_2 = \beta_3 = 0$	.747	.720
Test $\beta_1 = \beta_{enhanced}$	.309	.251
Control mean	.67	.66
Obs.	6494	6494

Notes: All regressions include LGA fixed effects and standard errors clustered at the cluster level. The control variables in each regression are identical to those reported in Table 3. The dependent variables include a dummy variable for whether the respondent reported a birth during the intervention period, and the number of births reported. In Panel A, the sample includes all women surveyed in the endline survey. In Panel B, the sample is restricted to women who reported a birth during the intervention period. Accordingly, the fertility variables are omitted, as there is essentially no variation in fertility in this subsample. We report tests of equality across the estimated coefficients; a test of the hypothesis that the joint effect of treatment is zero ($\beta_1 = \beta_2 = \beta_3 = 0$; and a test of equality of impact across the basic and enhanced arms $\beta_1 = \beta_{enhanced}$. Asterisks denote significance at the ten, five and one percent level.

Table 10: Treatment on the Treated estimates: Care utilization, health practices and health knowledge and attitudes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Care utilization								
	Any ANC visit	Number ANC visits	ANC quality index	Care for complications	Postnatal care	Facility birth	Skilled attendance	Birth accompanied
CORPs	.561*	1.923	.411**	-.301	.315**	-.190	.054	-.084
	(.314)	(1.485)	(.201)	(.328)	(.150)	(.184)	(.178)	(.215)
Obs.	4420	4420	4420	1570	3684	4009	3649	3649
Panel B: Maternal and newborn health practices								
	Birth plan	Breast-feeding	Breastfeeding exclusively	Breastfeeding duration	Immunizations	Infant check-up		
CORPs	.446***	-.181	.157	-12.920	.695	.806***		
	(.165)	(.167)	(.144)	(19.269)	(.442)	(.309)		
Obs.	4152	3051	3516	2803	1385	1317		
Panel C: Health knowledge and attitudes								
	Delivery preference	Knowledge relative risk	Knowledge compli-cations	Infant. care know.	Attitudes toward facility			
	(1)	(2)	(3)	(4)	(5)			
CORPs	.107	.295	.147	-.002	.174**			
	(.189)	(.215)	(.092)	(.128)	(.070)			
Obs.	6350	6350	6350	6350	1393			

Notes: All regressions include LGA fixed effects and standard errors clustered at the cluster level; the CORPs dummy variable is instrumented by a dummy variable for assignment to a treatment arm, and the specifications of interest are estimated employing two-stage least squares. The sample in Panels A and B includes all respondents reporting a pregnancy during the intervention period who were interviewed in at least one follow-up survey and who report the indicator of interest; the sample in Panel C includes all respondents surveyed in the endline, and the intervention exposure variable is coded zero for those who did not report a pregnancy during this period. The control variables in each regression are identical to those reported in Table 3. Asterisks denote significance at the ten, five, and one percent level.

A Appendix: for on-line publication only

A.1 Sampling strategies

In order to identify the sample communities, we utilized the following strategy. First, we selected the sample local government areas and associated primary health centers. The local government areas of Gumel, Hadejia, and Jahun did not receive any MSS-funded services; anecdotally, this primarily reflects the fact that these three LGAs included larger population centers that were already served by well-equipped health facilities or hospitals. Accordingly, these three LGAs were excluded from the evaluation. In the remaining 24 LGAs, there were 35 facilities included in the MSS as of 2012; in those LGAs that had more than one MSS-served PHC, the PHC that had a larger number of communities of the target size in the catchment area was selected. Effectively, this strategy identified the health center located in a more densely populated area.

Second, we defined a catchment area around each PHC. Health authorities at the federal or state level had no systematic enumeration of the catchment areas of the health facilities of interest. While some PHCs did informally define their own catchment areas, this was not consistent across facilities, nor could all facilities provide such information. Accordingly, we imposed our own criteria, using a radius of 20 kilometers from the PHC's location (calculated employing a straight-line distance) to identify a catchment area.[22] This was the minimum radius that would allow us to identify sufficient sample communities of the target size.

Third, we selected any community in the defined catchment area that had a reported population between 500 and 850 households for inclusion in the sample. If there were more than four communities of this size, we randomly selected a subset of these communities. In the majority of LGAs, however, we could not identify four communities of the target

[22]Information about the population and geographical coordinates of localities in the LGAs of interest was provided by the Nigerian National Population Commission using data collected in the 2006 census. Census data at the locality level had not previously been analyzed or published by the NPC, but the agency produced a dataset with locality level information at our request. The coordinates of the MSS facilities in these local government areas were also provided.

size. In these cases, we sought to identify clusters of adjacent villages that could be jointly treated as a unit. The sampling procedure then entailed identifying "anchor villages" having a population of at least 150 households, and evaluating whether there were additional villages within 3 kilometers of the anchor village. The objective was to identify a cluster utilizing the smallest number of communities possible: accordingly, we first identified clusters including two separate villages, then three, and then four. If there were more than the target number of clusters within a certain subset, we randomly selected clusters for inclusion. Given that our sampling units frequently included more than one distinct village, we adopted the term "clusters" to describe them. The final sample included 30 clusters (31% of the total) that were comprised of one village; 47 (49%) that were comprised of two separate but adjacent villages; 14 (15%) that were comprised of three separate villages; and 5 (5%) that were comprised of four villages.

Randomization of the sampled clusters into the four experimental arms was conducted by the research team using Stata.

A.2 Identifying sample households

Following the identification of the 96 target clusters, we deployed a survey team to conduct a partial census in each sampled cluster. The objective of the partial census was to verify the population estimates provided by the NPC, and to generate a household roster that could be the basis of sampling for the baseline survey.

In order to conduct the partial census, the survey team worked with traditional leaders in each village (known as lambas, bulamas, or mai' angwans) to identify key informants knowledgeable about the population, typically assistants or relatives of the village chief. Leaders also assisted in subdividing the community into non-overlapping and easily recognizable areas to facilitate the partial census; enumerators then worked in pairs to walk on foot through their assigned area in conjunction with informants. As they walked, they created a map and a list of households, identifying each house and asking the informant to provide the name of the household head and the names and approximate ages of his

wives, if applicable.[23]

Following the conclusion of the partial census, the lists of households in each cluster were digitized. The sampled households for the baseline were then identified as follows: first, all households that informants had designated as including only single men or elderly couples were dropped. Second, we selected a simple random sample of 15% of households including women of reproductive age (15-49) as target households for the baseline; an additional 3.75% of households were selected to serve as substitutes. In the event a selected household in fact did not have a woman of reproductive age resident, or the identified woman declined to participate, a household would be randomly selected from the pool of substitute households to join the sample.

During the partial census, the number of households identified in our census often fluctuated quite substantially relative to the number postulated by the NPC. While on average the enumerated population was around 90% of the projected population, in some LGAs this was substantially lower (as low as 50% or 75%), and in some much higher (as high as 150%). These discrepancies presumably primarily reflect the fact that the census data was collected six years prior to the start of the evaluation. When the population identified was so low as to render a cluster ineligible for inclusion in the evaluation, two strategies were employed. The first strategy was to expand the cluster by including additional, adjacent settlements that had not originally been identified for inclusion. If this was infeasible, however, the cluster would be dropped and replaced with another cluster that had previously been identified as eligible, but had not been randomly selected to be part of the evaluation sample.

Some variation across LGAs in the average cluster size, and thus in the number of households sampled in each cluster, could not be eliminated. The total number of households in the four sample clusters in each LGA, and the total number of households

[23]Prior to the start of data collection, we conducted two pilots in communities outside of the sampling frame that compared this method of conducting a partial census with a regular census where enumerators interviewed the household head or a spouse directly in every household. The results of this pilot process suggested that the partial census method yielded comparable results at a fraction of the time and associated costs of a full census.

sampled for the baseline survey within those clusters, is reported in Panel A of Table B2. While the experimental design estimated that the total population of the four sampled clusters in each LGA would be 2000 households, in practice this number varied between 1600 and 2400.

A.3 Pregnancies observed in each survey

The evaluation design called for each pregnancy in the baseline sample to be observed in both immediate post-birth surveys (conducted 3 days and 28 days after birth), as well as the endline survey. However, the monitoring and follow-up system designed to identify pregnancies on an ongoing basis did not identify all pregnancies. This was primarily due to poor performance by monitors and enumerators, who could not be directly supervised due to their dispersion over a wide geographic area. Frequent absences by both monitors and enumerators due to travel, pregnancy, or illness rendered prompt follow-up impossible, and in some of the most remote LGAs, recruiting sufficient (female) staff with adequate skills was challenging.

During the two-year period of continuous data collection, 55% of the pregnancies were thus observed in at least one of the ongoing surveys (3-day or 28-day survey or the audit). There are 1989 pregnancies, or 45% of the total sample, observed only in the endline.[24] This information is also summarized in Panel B of Table B2 in the Appendix.

A.4 Missing data

The objective of this section is to provide a brief overview of patterns of missing data and the reasons for these patterns in the primary results. First, intervention exposure (as reported in Table 3 is reported for all pregnancies observed during the intervention period, a sample of 4420 births.

[24]More specifically, there are 1562 pregnancies that were observed in the endline and in a 3-day or 28-day survey. 583 pregnancies were observed in the endline and the audit survey. 67 (57) pregnancies were observed only in an ongoing (audit) survey. 9 pregnancies were observed in both an audit and ongoing survey but not the endline, and 153 pregnancies were observed in all three surveys.

For care utilization variables, as reported in Table 4, variables for utilization of antenatal care are reported for all pregnancies. Care conditional on complications during pregnancy (Column (4)) is reported for the subsample of 1571 women who did report a complication during pregnancy. Delivering in a facility (Column (5)) is reported for the full sample of pregnancies that had concluded in a live birth prior to the endline, excluding the 411 pregnancies corresponding to stillbirths and pregnancies ongoing at endline. Dummy variables for postnatal care, skilled attendance at birth, delivering accompanied by another individual, and postnatal care (in Columns (5), (8) and (9)) are available for a slightly reduced sample, given that these variables were not reported in the 28-day survey or the audit.

For health practice variables, as reported in Table 5, the sample is limited for various reasons. Data in Columns (1) and (2) is missing for pregnancies reported in the audit, as well as for some pregnancies recorded only in the endline due to non-response. Data in Columns (3) and (4) is available in the 3-day survey and the endline, while data in Column (5) is reported only in the endline. Some respondents also stated that they could not recall their breastfeeding decisions at endline, and this response was coded as missing. Data in Columns (6) and (7) is available only in the 28-day survey.

For variables capturing maternal and neonatal morbidity, as reported in Table 6, variables capturing maternal morbidity and stillbirth are reported for all pregnancies in the sample. However, the variable capturing neonatal morbidity (reported in Column (4)) corresponds to data reported only in the 28-day survey, and is thus available for a reduced sample.

For variables capturing child anthropometrics, as reported in Table 7, the sample is restricted to households included in anthropometric data collection at endline reporting a child born in the intervention period.

For variables capturing additional outcomes around attitudes, knowledge, and fertility, as reported in Table 8, data availability is as follows. The variables reported in Columns (1) through (4) capturing knowledge and attitudes were reported in the endline, and are

available for the 6350 women observed in the endline. Attitudes toward the health facility (Column (5)) was collected in the 28-day survey, and is available for the reduced sample of women included in that survey. Fertility variables as reported in Columns (6) and (7) are available for all 6494 women observed in follow-up data collection.

Clearly, the fact that some outcomes are reported for a restricted sample may raise questions about bias driven by selection into the survey sample. Detailed evidence is presented in Section 5.1 that baseline demographics do not predict selection into the sample of respondents included in particular surveys within the evaluation, and there is also little evidence of differential selection across treatment arms. This suggests that despite the limited sample, there is minimal risk of bias introduced by selection into surveys.

B Appendix tables

Table B1: Characteristics of sample health facilities

	Mean	St. dev.	Min.	Max.
Facilities reporting access to electricity (last 6 months)	0.67	0.48		
Facilities reporting access to water (last 6 months)	0.92	0.28		
Facilities reporting access to refrigerator (last 6 months)	0.63	0.49		
Facilities reporting access to telephone (last 6 months)	0.04	0.20		
Facilities reporting access to ambulance (last 6 months)	0.21	0.41		
Facilities reporting ability to perform assisted vaginal delivery	0.92	0.28		
Facilities reporting ability to perform neonatal resuscitation	0.92	0.28		
Facilities reporting access to a blood bank for transfusions	0.38	0.49		
Facilities reporting ability to conduct a caesarean section	0	0		
Number of labor and delivery beds	2.6	2.9	1	16
Number of postpartum beds	3.6	4.6	0	22
Number of newborn beds	2.1	2.3	0	8
Number of deliveries in last 12 months	532.1	401.6	19	2109
Number of sick newborns treated in last 12 months	112.1	327.7	0	1554

Notes: This table reports summary statistics for the 24 primary health centers that serve the local government areas included in the primary sample.

Table B2: Sample composition

Baseline sample by LGA

LGA name	Total households in sampled clusters	Total sampled households	Average villages per cluster
Dutse	1782	307	3
Gwaram	1727	294	2.25
Miga	2211	355	2.8
Birniwa	1979	261	2.33
Kaugama	1509	183	1.67
Mallam Madori	1673	257	2.25
Babura	1689	288	2.25
Gagarawa	1677	278	2.33
Garki	1744	259	2.25
Maigatari	1694	223	2.56
Ringim	2127	351	2.78
Roni	1818	323	2.60
Birnin Kudu	2505	353	1
Buji	2230	332	2.25
Kiyawa	2379	347	1.67
Auyo	2218	289	1.86
Guri	2543	315	1
Kafin Hausa	1964	284	1.86
Kirikasama	2189	313	1.86
Gwiwa	2038	272	3
Kazaure	1851	316	3
Sule Tankarkar	2286	257	1
Taura	2203	337	1.67
Yankwashi	2359	275	1.86

Pregnancies observed by source

Endline - only	1989
Ongoing survey - only	67
Audit - only	57
Endline and ongoing	1562
Endline and audit	583
Audit and ongoing	9
All three surveys	153
Overall total	4420

Notes: In Panel A, we report for each LGA the total number of households identified in the four sampled clusters; the total number of sampled households within each cluster; and the average number of villages constituting each cluster. In Panel B, we report the number of pregnancies observed in the surveys conducted.

Table B3: Selection into surveys

	Ongoing survey			Endline anthropometrics		
	(1)	(2)	(3)	(4)	(5)	(6)
CORPs only	-.075**			.012		
	(.034)			(.022)		
Birth kits	-.005			.002		
	(.035)			(.023)		
Media	-.046			.016		
	(.040)			(.021)		
Married		.288***			-.114	
		(.096)			(.089)	
Has other wives		.013			.006	
		(.013)			(.012)	
Age at marriage		.005			.002	
		(.004)			(.004)	
Num marriages		-.012			-.017	
		(.017)			(.013)	
Age		-.001			-.005***	
		(.002)			(.002)	
Ever attended school		.065***			.0009	
		(.023)			(.016)	
Reads Hausa		-.026			-.005	
		(.029)			(.021)	
Muslim		-.186			-.241***	
		(.165)			(.021)	
Wealth index		.007			.004	
		(.005)			(.006)	
Birth parity		.011**			.007	
		(.005)			(.005)	
Any ANC visits			.040			.044
			(.062)			(.040)
Number of visits			-.007			-.008
			(.010)			(.006)
ANC quality index			-.029			.006
			(.089)			(.061)
Care comp.			-.009			.0006
			(.039)			(.024)
Facility birth			-.085			-.032
			(.087)			(.062)
Skilled birth			.129			.0006
			(.086)			(.054)
Postnatal care			.001			-.016
			(.031)			(.024)
Obs.	4420	4420	1312	4420	4420	1312

Notes: This table reports a series of regressions using the sample of all respondents reporting a pregnancy during the intervention period who were interviewed in at least one follow-up survey. The dependent variable is a dummy variable equal to one if the respondent was included in any ongoing survey (3-day or 28-day postnatal survey, or the audit) or the endline anthropometric data collection, and the independent variables are dummies for treatment assignment, demographic characteristics, and variables capturing care utilization as observed at baseline. All regressions include LGA fixed effects and standard errors clustered at the cluster level. Asterisks denote significance at the ten, five and one percent level.

Table B4: Baseline socioeconomic characteristics *for households reporting a birth during intervention period*

Variables	Control (N=1623)	CORPs only (N=1649)	CORPs + birth kits birth kits (N=1666)	CORPs + drama drama (N=1556)	Joint p-value	Obs.
Percentage of respondents observed at follow-up	.912	.912	.923	.929	.233	7069
Panel A: Respondent demographic characteristics						
Married	0.99	1.00	1.00	1.00	.43	4290
Number of other wives	0.32	0.36	0.35	0.31	.31	4290
Age at marriage	15.10	15.27	15.29	15.39	.45	4282
Number of marriages	1.19	1.21	1.15	1.17	.03	4288
Age	25.97	25.76	26.21	25.93	.17	4290
Ever attended school	0.16	0.17	0.20	0.20	.33	4290
Literate in Hausa	0.08	0.10	0.12	0.11	.78	4290
Muslim	1.00	1.00	1.00	1.00	.03	4290
Birth parity	3.76	3.61	3.86	3.77	.03	4290
Wealth index	-0.05	0.09	-0.07	0.09	.33	4290
Panel B: Household income and consumption						
Head attended school	0.30	0.34	0.34	0.36	.82	4007
Head's highest educ. (prim.)	0.15	0.14	0.14	0.15	.87	4007
Head's highest educ. (sec.)	0.11	0.13	0.12	0.15	.29	4007
Head's occupation (Own cultivation)	0.49	0.48	0.49	0.48	.96	4280
Head's occupation (Non-farm self-emp.)	0.31	0.29	0.30	0.31	.65	4280
Head's occupation (Outside employment)	0.12	0.15	0.13	0.12	.62	4280
Owns land	0.95	0.94	0.94	0.93	.49	4290
Number plots	3.72	3.66	3.52	3.33	.10	4045
Cultivates any millet	0.90	0.89	0.88	0.86	.27	4290
Cultivates any guinea corn	0.87	0.84	0.85	0.84	.99	4290
Cultivates any beans	0.76	0.74	0.75	0.73	.61	4290
Cultivates any ground nut	0.46	0.44	0.40	0.47	.14	4290

Notes: This table reports the mean values of household demographic characteristics as reported at baseline for households in each experimental arm; the sample is restricted to households reporting a birth during the intervention period. We also estimate a regression in which each demographic characteristic is regressed separately on three dummy variables for assignment to each treatment arm, as well as LGA fixed effects; standard errors are clustered at the cluster level. The reported p-value is the p-value on the joint test $\beta_1 = \beta_2 = \beta_3$, where the three coefficients refer to the coefficients on each treatment dummy variable.

Table B5: Baseline health care utilization, health practices and health outcomes *for households reporting a birth during intervention period*

Variables	Control (N=1623)	CORPs only (N=1649)	CORPs + birth kits (N=1666)	CORPs + drama (N=1556)	Joint p-value	Obs.
Panel A: Utilization of maternal health care						
Any antenatal visit	0.47	0.53	0.58	0.60	.28	2661
Number of antenatal visits	1.96	2.24	2.49	2.52	.32	2661
Antenatal quality index	0.26	0.31	0.35	0.36	.17	2661
Care-seeking for complications	0.45	0.45	0.52	0.52	.12	1782
Facility delivery	0.07	0.11	0.08	0.10	.39	2528
Skilled attendant at birth	0.07	0.11	0.09	0.12	.25	2528
Post-natal check-up	0.25	0.24	0.29	0.29	.05	2528
Panel B: Maternal and newborn health practices						
Male involvement in pregnancy	0.57	0.56	0.60	0.60	.19	2584
Infant breastfed (first day)	0.82	0.83	0.83	0.85	.49	2464
Excl. breastfeeding (first three days)	0.31	0.31	0.34	0.28	.19	2464
Panel C: Maternal morbidity						
Index of complications (pregnancy)	0.56	0.58	0.57	0.53	.27	2661
Index of complications (delivery)	0.05	0.05	0.07	0.08	.07	2528
Index of complications (postpartum)	0.21	0.24	0.21	0.21	.67	2528
Joint F-test (All baseline characteristics in Table B4 and B5)					.52	

Notes: This table reports the mean values of household characteristics as reported at baseline for households in each experimental arm; the variables reported are constructed to be identical to the outcome variables of interest analyzed in Tables 4 through 7, and the sample is restricted to households reporting a birth during the follow-up period. We also estimate a series of regressions in which each demographic characteristic is regressed separately on three dummy variables for assignment to each treatment arm, as well as LGA fixed effects; standard errors are clustered at the cluster level. The reported p-value is the p-value on the joint test $\beta_1 = \beta_2 = \beta_3$, where the three coefficients refer to the coefficients on each treatment dummy variable.

Table B6: Baseline socioeconomic characteristics *for households observed at follow-up*

Variables	Control (N=1623)	CORPs only (N=1649)	CORPs + birth kits (N=1666)	CORPs + drama (N=1556)	Joint p-value	Obs.
Percentage of respondents observed at follow-up	.912	.912	.923	.929	.23	7069
Panel A: Respondent demographic characteristics						
Married	0.99	0.99	1.00	0.99	.43	6494
Number of other wives	0.34	0.38	0.35	0.35	.31	6494
Age at marriage	15.02	15.18	15.23	15.24	.45	6481
Number of marriages	1.23	1.24	1.18	1.23	.03	6489
Age	28.09	27.69	27.77	28.02	.17	6494
Ever attended school	0.16	0.16	0.19	0.18	.33	6494
Literate in Hausa	0.08	0.08	0.11	0.10	.78	6494
Muslim	1.00	1.00	1.00	1.00	.03	6494
Birth parity	4.25	4.19	4.24	4.32	.03	6494
Wealth index	-0.04	0.05	-0.05	0.06	.33	6494
Panel B: Household income and consumption						
Head attended school	0.27	0.31	0.32	0.33	.82	6094
Head's highest educ. (prim.)	0.13	0.14	0.13	0.14	.87	6094
Head's highest educ. (sec.)	0.10	0.12	0.11	0.12	.29	6094
Head's occupation (Own cultivation)	0.50	0.50	0.51	0.50	.96	6466
Head's occupation (Non-farm self-emp.)	0.29	0.29	0.29	0.30	.65	6466
Head's occupation (Outside employment)	0.11	0.13	0.13	0.12	.62	6466
Owns land	0.95	0.94	0.94	0.94	.49	6494
Number plots	3.75	3.64	3.60	3.50	.10	6130
Cultivates any millet	0.90	0.89	0.88	0.88	.27	6494
Cultivates any guinea corn	0.88	0.85	0.84	0.84	.99	6494
Cultivates any beans	0.78	0.75	0.75	0.75	.61	6494
Cultivates any ground nut	0.48	0.45	0.41	0.48	.14	6494

Notes: This table reports the mean values of household demographic characteristics as reported at baseline for households in each experimental arm; the sample is restricted to households observed at follow-up. We also estimate a regression in which each demographic characteristic is regressed separately on three dummy variables for assignment to each treatment arm, as well as LGA fixed effects; standard errors are clustered at the cluster level. The reported p-value is the p-value on the joint test $\beta_1 = \beta_2 = \beta_3$, where the three coefficients refer to the coefficients on each treatment dummy variable.

Table B7: Baseline health care utilization, health practices and health outcomes *for households observed at follow-up*

Variables	Control (N=1623)	CORPs only (N=1649)	CORPs + birth kits (N=1666)	CORPs + drama (N=1556)	Joint p-value	Obs.
Panel A: Utilization of maternal health care						
Any antenatal visit	0.47	0.53	0.56	0.57	.28	3697
Number of antenatal visits	1.97	2.22	2.43	2.40	.32	3697
Antenatal quality index	0.27	0.31	0.34	0.34	.17	3697
Care-seeking for complications	0.47	0.47	0.52	0.53	.12	2427
Facility delivery	0.07	0.10	0.09	0.10	.39	3454
Skilled attendant at birth	0.07	0.11	0.10	0.13	.25	3454
Post-natal check-up	0.25	0.24	0.29	0.28	.05	3454
Panel B: Maternal and newborn health practices						
Male involvement in pregnancy	0.58	0.56	0.59	0.59	.19	3542
Infant breastfed (first day)	0.83	0.82	0.83	0.83	.49	3365
Excl. breastfeeding (first three days)	0.32	0.30	0.33	0.27	.19	3365
Panel C: Maternal morbidity						
Index of complications (pregnancy)	0.55	0.58	0.57	0.54	.27	3697
Index of complications (delivery)	0.05	0.06	0.07	0.08	.07	3454
Index of complications (postpartum)	0.21	0.24	0.21	0.21	.67	3454
Panel D: Infant and child anthropometrics						
Weight-for-age (< 1 year)	0.13	-0.18	-0.33	0.24	.02	1892
Height-for-age (< 1 year)	0.48	0.62	0.36	0.92	.41	1861
MUAC-for-age (< 1 year)	- 0.73	-0.76	-0.90	-0.93	.16	1478
Weight-for-age (< 2 year)	-0.44	-0.65	-0.77	-0.27	.00	3144
Height-for-age (< 2 year)	0.25	0.26	0.16	0.63	.05	3078
MUAC-for-age (< 2 year)	-0.88	-0.83	-0.91	-0.94	.20	2718
Joint F-test (All baseline characteristics in Table B6 and B7)					.47	

Notes: This table reports the mean values of household characteristics as reported at baseline for households in each experimental arm; the variables reported are constructed to be identical to the outcome variables of interest analyzed in Tables 4 through 7, and the sample is restricted to households observed at follow-up. We also estimate a series of regressions in which each demographic characteristic is regressed separately on three dummy variables for assignment to each treatment arm, as well as LGA fixed effects; standard errors are clustered at the cluster level. The reported p-value is the p-value on the joint test $\beta_1 = \beta_2 = \beta_3$, where the three coefficients refer to the coefficients on each treatment dummy variable.

Table B8: Analysis of differential attrition predictors

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Respondent demographic characteristics							
Treated	.163	.163	-.011	.073	.019	-.041	-.005
	(.106)	(.106)	(.010)	(.077)	(.018)	(.027)	(.010)
Covariate	-.036	-.036	-.007	-.0002	.025*	-.001*	.003
	(.061)	(.061)	(.012)	(.005)	(.013)	(.0008)	(.017)
Treated x covariate	-.173	-.173	.006	-.005	-.022	.001	-.021
	(.107)	(.107)	(.014)	(.005)	(.015)	(.0009)	(.020)
Covariate	Married	Number other wives	Age at marriage	Number of marriages	Age	Attended school	Literate in Hausa
Obs.	7069	7069	7069	7052	7063	7069	7069
Panel B: Respondent demographic characteristics, cont.							
Treated	.134	-.025*	-.009				
	(.131)	(.015)	(.008)				
Covariate	.089***	-.006***	-.00002				
	(.020)	(.002)	(.009)				
Treated x covariate	-.143	.004	-.005				
	(.130)	(.002)	(.009)				
Covariate	Muslim	Birth parity	Wealth index				
Obs.	7069	7069	7069				
Panel C: Utilization of maternal health care							
Treated	-.012	-.006	-.005	-.022	-.006	.006	.008
	(.014)	(.014)	(.014)	(.017)	(.012)	(.011)	(.011)
Covariate	-.024	-.003	-.022	-.008	-.016	.008	.054**
	(.017)	(.003)	(.023)	(.021)	(.017)	(.030)	(.028)
Treated x covariate	.024	.003	.017	.015	.040*	-.014	-.050
	(.020)	(.004)	(.027)	(.024)	(.022)	(.034)	(.034)
Covariate	Any ANC visit	Number of visits	ANC quality index	Care comp.	Postnatal care	Facility birth	Skilled attendance
Obs.	4007	4007	4007	2625	3742	3742	3742

Notes: This table reports the results of a series of regressions in which a dummy for attrition is regressed on baseline covariates, a dummy for assignment to a treatment arm, and the interaction between the two; the covariates included are household demographic characteristics as previously reported in the balance tests. The attrition dummy is equal to one if a baseline respondent is not observed in any survey post-baseline. All regressions include LGA fixed effects and standard errors clustered at the cluster level. Asterisks denote significance at the ten, five and one percent level.