Improving Antenatal and Postnatal Care Delivery Services in Remote Communities using Telehealth Pre-Analysis Plan*

Firoz Ahmad† Asad Islam‡ Umair Khalil§
Tabassum Rahman¶ Daniel Suryadarma‖

September 29, 2023

Abstract

Improving pregnant women and newborn babies' health care services in resource-poor settings, particularly in remote communities in many developed countries and in most developing countries, remains an important challenge. We implement a technology-assisted healthcare intervention focusing on pregnant women in rural Bangladesh. The technology involves a simple phone-based automated calling system, called interactive voice response (IVR), through which we offer telehealth support to women from their first/second trimesters of pregnancy. The telehealth support enabled participants to interact with the message system using the keypad of their mobile phones. In a separate treatment arm, these women are able to directly contact a doctor to discuss their health and the health of their babies once they

*This project was supported by funding from Laerdal foundation and Asian Development Bank.

†Technical University of Munich and Khulna University. Email: firoz.ahmed@tum.de

‡Corresponding author: Centre for Development Economics and Sustainability (CDES) and Department of Economics, Monash University, Australia, and J-PAL. Email: asadul.islam@monash.edu

§Department of Economics, Deakin University, Australia. Email: umair.khalil@deakin.edu.au

¶University of Melbourne, Australia. Email: tabassum.rahman@unimelb.edu.au

‖Asian Development Bank Institute, Kasumigaseki Building 8 F, 3–2-5 Kasumigaseki, Chiyoda-ku, Tokyo 100-6008, Japan. Email: dsuryadarma@adbi.org.

are born. We examine women's health-seeking behavior during the antenatal and post-natal period and its impact on maternal and child physical and mental health.

Keywords: parenting knowledge; early childhood development; RCT; pregnancy intervention; birth outcomes

JEL Codes: C93, D8, I12, I15, O15

1 Introduction

Research across multiple disciplines has established the critical importance of the intrauterine environment for child health and development (Barker, 1990; Almond and Currie, 2011). A substantial literature in economics has also extended this critical phase to the postnatal period with impacts leading well into adulthood (Hoynes et al., 2016; Almond et al., 2018). These studies have shown how proper nutrition and vitamin intake, lower maternal mental stress, physical exercise, and other such positive interventions can substantially improve maternal and child health.

Therefore, equipping pregnant women with adequate information and knowledge about how best to take care of themselves and the fetus in-utero and the baby postnatal has been of prime policy importance. However, in developing countries, access to this important resource is severely inadequate. For instance, between 2015-2021, only 37% of women in Bangladesh received the WHO-recommended four or more antenatal care (ANC) visits (UNICEF, 2022).[1] In contrast the median pregnant woman in the United States in 1980 went for 11 ANC visits (Buekens et al., 1993), while in 2015 the lowest 5th percentile of visits was five.[2] The WHO aims to have at least 90% of pregnant women attend four or more ANC visits by 2025 (WHO, 2021), a target that most developing countries including our country of study is woefully far from. A similar picture emerges for postnatal care as well although cross-country statistics are less well documented. A major impediment to universal access to pregnancy care in developing countries has been the inadequate availability of trained medical professionals. For instance, Bangladesh has only 0.6 doctors

[1]Globally, South Asia and Sub-saharan Africa have the lowest rates of ANC utilization with Pakistan and India at 52 and 58% and Somalia and Chad recording numbers as low as 24 and 31%, respectively.

[2]Based on authors' own calculation from the 2015 US Vital Statistics.

and only 0.4 nurses and midwives per 1000 people, the US in contrast has 2.6 and 15.7 of these professionals, respectively (WB, 2023).

Thus there is a need to implement innovative, technology-based solutions, which can be scaled up at a relatively low cost and at the same time can effectively deliver adequate antenatal and postnatal care, particularly in developing countries. In this project, we design and implement an interactive voice response (IVR) based reproductive health information service for pregnant women in rural Bangladesh.[3] The intervention is administered via cellphone through which respondents are able to connect to our virtual call center.

Bangladesh has been at the forefront of the telecom revolution, with cell phone subscriptions per 100 people jumping from 2 in 2004 to 107 in 2021 (WB, 2023).[4] We leverage this increased accessibility in telecom connectivity to provide access to maternal and child health-related information to a sample of rural women aged between 15-40 years, and measure the efficacy of our intervention. In particular, our study is conducted across 330 villages in southwestern Bangladesh and consists of two treatment arms: i) IVR + Tele-health: participants are able to interact with the message system using the keypad of their phone; ii) arm (i) + participants are able to contact a project doctor to discuss their health and the health of their babies once they are born. The third arm comprises the control group which is given no intervention. The three arms are distributed equally across the total number of villages in our study, i.e., 110 villages per arm. We aimed to recruit around 1400 women per arm but ended up with sample sizes of around 1350 in each arm. These are detailed below. The IVR system comprises 15 modules containing pregnancy and postpartum stage-specific information on antenatal and postnatal care, antenatal and postnatal mental health, and taking care of a newborn baby for up to 6 months. They are recorded in voice-over, form and participants are able to access this information any time they want, and

[3] IVR is an automated phone system technology that enables incoming callers to access pre-recorded information. It does not require any smart devices and allows callers to use the keypad of their mobile phone to select modules from a list of menu options. The information can be accessed at any time, allowing listeners to learn at their convenience and pace.

[4] In recent years, a similar rise in the telecom network has been experienced in other countries that have been lagging behind in the provision of adequate antenatal and postnatal care, like in Pakistan and India, and even some sub-Saharan African countries. Hence, our intervention can provide evidence for a readily scalable, and low-cost solution for increasing access to reproductive services in these countries as well.

can listen to it multiple times.

Due to the nature of the study, we are not able to recruit all participants into our study at the same time. For example, it is not practical to expect in the context of this study that we would be able to find enough pregnant women in each village within a static time window. Therefore, our recruitment in this study was dynamic, with field visits once every month after the first round of recruitment. However, there was a closing date (see Table 1 for more details). Furthermore, our focus is on recruiting mothers who are in the second trimester of their pregnancy at most. This will allow our intervention to have the potential of influencing birth outcomes given the sizable literature that has established the significance of interventions earlier on during the pregnancy (Almond and Mazumder, 2011; Grossman and Khalil, 2022). Section 2 provides further details about the exact nature of our intervention and the logistics involved as well as the timeline of the salient project milestones.

Our endline surveys are designed to measure objective outcomes revolving around health-care utilization. We conduct this around 6 months postpartum. We focus on a range of primary and secondary outcomes. The primary outcomes are - (i) knowledge and practices of maternal health, (ii) knowledge and practices of childcare, and (iii) health outcomes at birth and up to 6 months postpartum. The secondary outcomes are - (i) depression, (ii) the Edinburgh Postnatal Depression Scale (EPDS), (iii) maternal mental health awareness, (iv) context-specific mental health awareness, and (v) spillover effect on older children (food consumption and illness). We collect information on birth weight and height around 10-13 days after delivery to minimize measurement error induced by recall errors as detailed in milestone 13 given in Table 1.

A sizable experimental and non-experimental literature has established how nutritional support (e.g., Chorniy et al., 2020; Carneiro et al., 2021), cash transfers (e.g., Amarante et al., 2016; Parker and Todd, 2017), and knowledge-enhancing interventions (e.g., Doyle, 2020) both during pregnancy and immediately after delivery can improve maternal as well as child health and impacts even later life skills for the latter (e.g., Doyle, 2022). However, various surveys and cross-sectional studies have also consistently found how women in developing countries, particularly

poorer women, may still have inadequate knowledge regarding best practices and maternal needs during pregnancy (e.g., Choudhury et al., 2012; Withers et al., 2018; Abebe et al., 2021).[5]

The IVR-based intervention that we implement is designed to fill these health literacy gaps among rural Bangladeshi women and then measure the impacts this has on health behavior and birth outcomes. IVR based interventions have only recently been beginning to be applied in development economics, for instance, in agricultural advice to farmers (Walter et al., 2020; Van Campenhout et al., 2021); for remote learning among school children (Wang et al., 2023). The low-cost delivery of such interventions and thus their inherent scalability bodes well in tackling a number of pressing issues in the developing world, which would otherwise require access to substantial human capital resources. However, evidence is needed on the actual efficacy of this approach which our project endeavors to provide in the context of maternal and child health among a sample of pregnant women.

The rest of this document provides details on our intervention and research design (section 2), the econometric procedures that we will employ on the final data (section 3), and the various analyses that we hope to conduct by the end of the project (section 4).

2 Intervention and Research Design

2.1 Recruitment Details

Our intervention was focused on mothers aged between 15 and 40 years and those within the first two trimesters of their pregnancy. The intervention was implemented in collaboration with a local research partner Global Development and Research Initiatives (GDRI).[6] The study locations

[5] Maternal mortality rate in Bangladesh stands at an alarming 176 per 100 thousand live births (UNICEF, 2022). Neonatal deaths (30/1000 live birth) account for 67% of all under-5 deaths in Bangladesh (based on the Bangladesh Demographic Health Survey, 2020). Furthermore, during infancy, the risk of dying in the first month of life (30 deaths per 1,000 live births) is nearly four times greater than the rate of dying in the subsequent 11 months of their first year of life.

[6] GDRI is a research-focus local non-government organization with extensive experience in implementing RCTs on early childhood development, maternal and child health, and education. The organization has been working in the southwest part of Bangladesh since 2009.

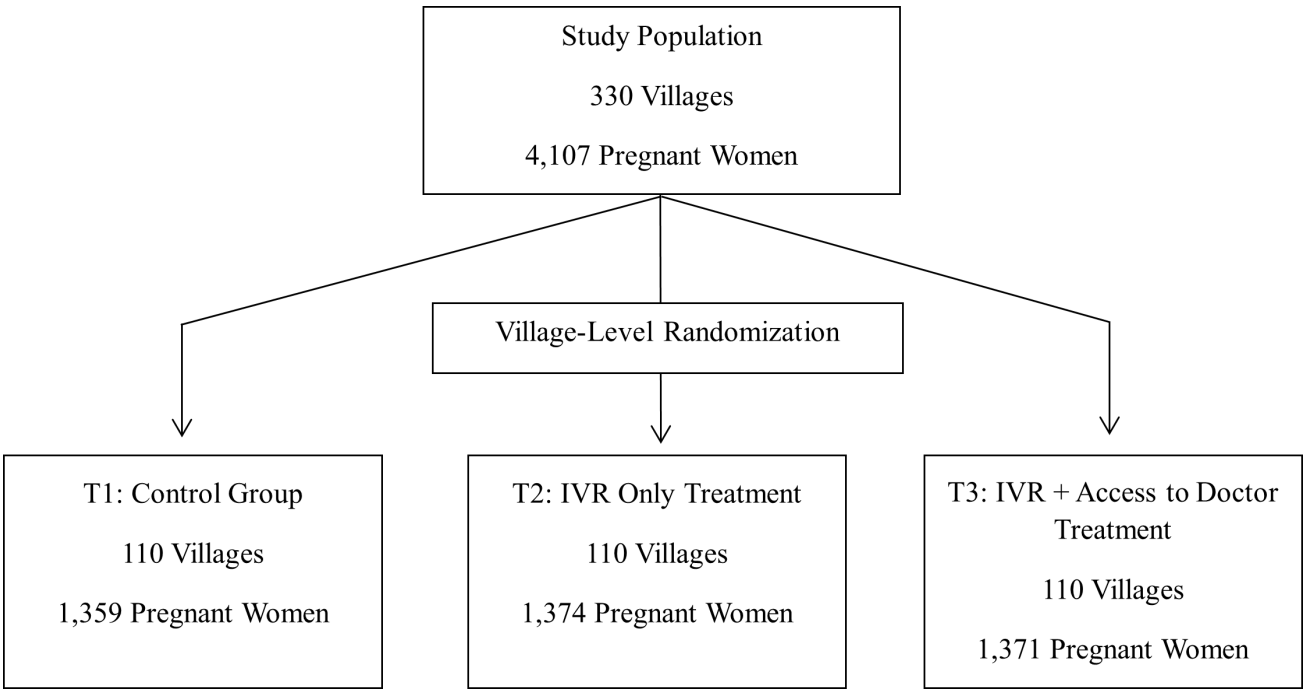

Figure 1: Randomization Design

were chosen based on GDRI's engagement and reach. We conducted our experiment in a total of 330 villages split equally across the three study arms. We identified 8-20 expectant mothers from each village who qualified based on the above criteria. In order to select the respondent, we made a list of pregnant women depending on their availability and we randomly selected them from this list. If we could not achieve the expected number of pregnant women from the initial list, we continued recruitment as an ongoing process. The participants reserved the right to choose and decide whether they want to take part in this study.

2.2 Treatment Details

Our treatment intervention provides knowledge on reproductive health issues via information modules delivered via IVR technology, which treated women are able to access it for free using their cell phones.

The intervention has two treatment arms and a control arm outlined below.

1. Arm 1: The control group was not administered any intervention.

2. Arm $T2$: Access to IVR-based modules via toll-free phone number

3. Arm $T3$: Access to IVR-based modules via toll-free phone number + telehealth support by a doctor for common health problems, at participant's own discretion. each.

Enumerators working with the two treatment arms will have different responsibilities. For the participants in arm $T2$, we will see them face-to-face a total of 4 times: first, at the time of the information and baseline survey, about 45-60 minutes; second, at the time of the demonstration of how to access IVR information modules, about 45-60 minutes; third, at the 10th to 13th day of delivery, about 30 minutes; and fourth, at endline, that is 6-6.5 months after delivery, about 45-60 minutes. The enumerators explained in a step-by-step manner, how the participant could access the IVR modules using a mobile phone.[7] They highlighted that participants can access the information free of cost and as many times as they like. Participants are expected to listen to each IVR module at least once for a potentially measurable effect. The duration of each module is about 5 to 8 minutes. However, for each participant, we will observe their interaction with the system and the time they spent dialed into each module.

Fifteen IVR modules are being delivered to the T2 and T3 participants. Intervention modules cover the following topics: (i) taking care of self during pregnancy (Module 1); (ii) diet during pregnancy (Module 2); (iii) signs of potential pregnancy complications (Module 3); (iv) antenatal care and birth plan (Modules 4-6); (v) postnatal care (Modules 7-8); (vi) maternal mental health (Modules 10-12); and caring for the baby in the first six-moths of life (Modules 9, 13-15).

The telehealth intervention in $T3$, is provided for individual, one-on-one counseling regarding any pregnancy-related issues including some general health problems only where a doctor can provide support without physically examining pregnant women. This includes referrals, advice on what to do for each individual circumstance, and over-the-counter drugs to support ongoing general health situations. If a physical examination is needed, the patient is referred to a doctor or a nearby hospital. The patient was always advised to visit a doctor or hospital and was told

[7]The nature of our intervention precludes us from recruiting households who do not have access to at least one mobile phone. However, as mentioned earlier Bangladesh has 107 mobile phones per 100 people and hence this restriction is not likely to be severely constricting.

not to think of this telehealth option as an alternative to seeing a doctor face-to-face. It provided flexibility for patients to seek immediate support for consultation. In most cases, women from rural areas do not have access to consulting a doctor, and a visit to a certified doctor is seen as the last option. The situation is even more acute for pregnant women because of trouble commuting far to see a doctor or nurse in a clinic or hospital.[8] Thus, telehealth consultations are less likely to replace any existing medical support they can receive.

Table 1 provides a detailed timeline of all important milestones during the course of this project along with the exact time period within which each would be conducted.

2.3 Data and Balance Checks

The data collected at baseline include (i) sociodemographic information of participants and their spouses; (ii) information on previous pregnancies (e.g., overall health, pregnancy complications, and outcomes, place of delivery); (iv) information on current pregnancy (e.g., antenatal visits, immunisation, diagnosis of health problems observed in pregnant women, preference for place of delivery; available health care facility); and (v) knowledge about maternal and child health.

In Tables 2 to 5 we report balance across the three arms for key maternal, paternal as well as current pregnancy and knowledge-related variables. The average age of the mothers in our samples is around 24 and they were married when they were just under 18 years.[9] The respondents have finished around 8-9 years of schooling on average as well. We record perfect balance in these characteristics as evidenced by columns (5) to (7) of Table 2. We see balance in other maternal and household characteristics in Table 2 as well.

Similarly, Table 3 presents balance checks for paternal characteristics. The fathers in our sample are on average six years older than the mothers. Around 28% work in agriculture, 40% are day laborers, 20% have their own business, while the remaining 12% have other government or private jobs. The last three columns again show that we achieve balance in all these covariates

[8]Factors such as transport cost, inconvenience, lack of a person accompanying pregnant women, not knowing the doctor, etc. play roles in not seeking an appointment with a doctor.

[9]While the legal age of marriage in Bangladesh is 18 for women and 21 for men, it is harder to enforce this in rural areas where we primarily draw our sample from.

Table 1: Timeline of Important Milestones for the Study

	Activity	Time Period
1	Village survey	02-17 August 2022
2	Randomization based on village survey under three treatment arms	23-25 August 2022
3	Audio lessons finalized (15 lessons)	04-09 September 2022
4	Enumerator selection and recruitment	24-29 September 2022
5	Enumerator training	08-11 October 2022
6	Baseline survey piloting	17-22 October 2022
7	Finalizing baseline questionnaires	02-07 November 2022
8	IVR setup and piloting	22-30 November 2022
9	Doctor's selection and recruitment	16-20 December 2022
10	Baseline survey and distribution of IVR operation manual & in-person one-to-one training (first phase)	21 January 2023-04 February 2023
11	Baseline survey and distribution of IVR operation manual & in-person one-to-one training (second phase)	19 February-04 March 2023
12	Baseline survey and distribution of IVR operation manual & in-person one-to-one training (third phase)	18-31 March 2023
13	Meeting with the women after 10-13 days of giving birth, and record the condition of the mother and newborn	March-October 2023
14	Endline survey	September 2023-March 2024

Notes: In case of any premature birth we also followed women and babies after around 10-13 days of giving birth, regardless of the timing of the birth of the child, beginning from March 2023.

as well.

In Table 4 we report salient characteristics of the current pregnancy, i.e., the one that our treatment targets. Around 30% of the mothers in our sample were in their first trimester at the time of enrollment in the trial, 65% were in the second trimester, while only 3% were in their third trimester. This is by design as we aimed to target mothers in the first two trimesters of their pregnancy given the extant literature on how interventions during the earlier phases of the pregnancy are more likely to have a beneficial impact. On average our respondents found within 2 months of conceiving that they were pregnant and once again we achieved balance for all the above variables across the three arms.

In Panel B of Table 4 we document the baseline rates of healthcare access in our sample. Less than 50% of our respondents had any consultation with a doctor even though well over half of our sample is already in their second trimester. This highlights the absolute need for increased access to innovative, low-cost, low-effort (on the part of the patient) options for healthcare. Our IVR-based intervention is aimed to provide evidence for the efficacy of precisely such options.

Finally, in Table 5 we document and show balance across treatment arms in the knowledge base of mothers in our sample across the domain of maternal and neonatal health. While baseline knowledge of post-birth practices is fairly high like breastfeeding initiation (around 70%), minimum duration in infancy (50%) as well as the introduction of additional foods (85%), existing knowledge about appropriate medical care during and immediately after is almost non-existent. Only 2% of mothers are aware of the WHO recommended 8 antenatal care visits and 4% are aware of the required 4 visits post-natal. On the other hand, Panel C shows that over one-third of the sample has visited and taken remedies from folk healers for their pregnancy complications. For all variables in the domain of knowledge as well we achieve balance except marginal significance for three variables as shown in columns (6) and (7).

To conclude this section, our design is successful in balancing 14 dimensions of maternal characteristics, 10 of paternal, and 4 household-level measures. Similarly, we achieve balance for baseline measures of current pregnancy as well as for healthcare access variables. Finally, we

report balance on 10 dimensions of knowledge measures among women in our sample regarding maternal, neonatal, and pregnancy and delivery domains.

2.4 Outcomes

We focus on a range of primary and secondary outcomes which we detail below. These revolve around objective outcomes for healthcare utilization at various points during and after the pregnancy as well as actual health outcomes of the baby. We also measure the direct impact of our intervention on maternal knowledge and health awareness regarding reproductive health. Our primary outcomes are in the following domains: (i) knowledge and practices of maternal health, (ii) knowledge and practices of childcare, and (iii) health outcomes at birth and up to 6 months postpartum. Secondary outcomes focus on: (i) Depression, (ii) the Edinburgh Postnatal Depression Scale (EPDS), (iii) maternal mental health awareness, (iv) context-specific mental health awareness, and (v) spillover effect on older children (food consumption and illness).

The indices for the outcomes that we will construct from individual questions will proceed with the following approach: i) standardize each question by the mean and standard deviation of the control group; ii) create weights based on the procedure outlined in Anderson (2008) to ensure that highly correlated questions are given small or negative weights, while less correlated outcomes receive higher weights (Schwab et al., 2020); iii) calculate the weighted, aggregate index based on ii); and finally iv) normalize the aggregate index again by the control group mean and standard deviation.

2.4.1 Knowledge and Practices of Maternal Health

We will create an index of maternal health knowledge based on the following 7-item questions. We are trying to understand if they know these seven options related to maternal health. We would assign 1 to items that are answered right and 0 for wrong choices.

1. If you need to eat food rich in vitamin C, what food would eat?

2. If your friend is pregnant and is suffering from a low level of iron, what food do you think she should eat?

3. It is good to have plenty of tea or coffee when someone is pregnant.

4. If your friend is pregnant, what advice would you give her in relation to food?

5. A pregnant woman needs to sleep 6-8 hours per night and should have 2 hours of rest during the day.

6. If your friend is pregnant, how often do you think she should see her doctor/midwife?

7. If your friend has just given birth, how often do you think she should see her doctor/midwife post-delivery?

In order to understand their knowledge and awareness of pregnancy-related complications, we will create an index of pregnancy-related complication knowledge based on the following 5-item questions. We are trying to understand if they know these five options or not. We would assign 1 to items that are answered, and 0 for others.

1. Excessive bleeding during pregnancy or delivery or after delivery and placenta accreta

2. Having a fever for more than three days during pregnancy or after delivery

3. Retaining water during pregnancy or delivery or after delivery, blurry vision, and headache

4. Convulsion during pregnancy or delivery or after delivery

5. Prolong labor over 12 hours and other parts of the baby's body being visible first instead of the head

To understand the standard maternal health practices specified by the World Health Organization (WHO), we will have two questions on a visit to a doctor/midwife during antenatal care (ANC) and postnatal care (PNC). Regarding the ANC visit, we will ask, "How often did you see

your doctor/midwife?". We consider at least 8 ANC visits as sufficient (=1) and 0 otherwise. Regarding the PNC visit, we will ask, "How many times did you see your doctor/midwife?" during the postpartum period. We consider at least 4 PNC visits as sufficient (=1) and 0 otherwise. Moreover, to understand maternal health practice, we will ask them a set of questions to know their practice after delivery. We will ask about their sleeping pattern and duration of rest as well as food consumption patterns after delivery. We will ask them three specific questions to understand their rest duration and sleeping patterns. These are:

1. How much sleep did you get per night, on average, in the last 7 days? (in hours),
2. How much sleep did you get last night? (in hours),
3. Other than sleep at night, how much time did you spend resting yesterday? (in hours).

To understand their food consumption pattern after delivery, we will ask several questions. First, we will ask them a single question on food consumption to know whether they are taking more food compared to before the pregnancy period or not. We will ask a question: "Thinking about your food consumption, what would you say about your current food consumption? 1= I eat the same volume of food I used to before becoming pregnant, 2 = I eat more since I became pregnant, and 3= I eat more since I had my baby".

In addition, we will ask them about a set of food items. Whether they ate the food items in the last 7 days: (i) Rice, (ii) Green leafy vegetables, (iii) Potatoes, (iv) Eggs, (v) Chicken, (vi) Beef/Mutton, (vii) Dal, (viii) food cooked with oil/ghee/butter, (ix) Cow's/goat milk, (x) Food prepared with milk, (xi) Seasonal fruits, and (xii) Banana. We will ask them about their consumption pattern on a 5-point scale: 0=Not at all or 0 days, 1= Rarely or 2 days in the last 7 days, 2= Sometimes or 3 days, 3=Often or 5 days, and 4= Always or 7 days. We code the activity as 1 if the respondent answered either 3 or 4, and 0 otherwise. Therefore, 1 corresponds to higher-quality food consumption. We will follow the above-detailed procedure to construct composite indices out of these questions.

2.4.2 Knowledge and Practices of Childcare

To assess maternal knowledge of childcare, we will use the following module based on 7 questions. We are trying to understand if they know these seven options related to childcare. We would assign 1 to items that are answered right and 0 for wrong choices.

1. What is the most appropriate first food for a baby less than 6 months old?
2. When a baby should be given complementary food along with breast milk?
3. How many times a baby should be fed?
4. For how long should a baby be breastfed along with complementary feeding?
5. What should be the first food for a baby?
6. To keep the child secure and disease-free, a child should be vaccinated
7. 10 vaccines should be provided to children before the age of 12 months

Similarly, our survey will also ask mothers several questions to understand their childcare practices related to breastfeeding and children's food consumption. These questions include: "Do you currently breastfeed your baby?", 1=yes, 0=no. If the answer is 'no', then we will ask: "When did you stop breastfeeding your child?", 0 =I never breastfed as such (regularly), 1 =When my baby was less than 3 months old, 2 = When my baby was less than 6 months old, and 3 = When my baby was 6 months old. If the answer to the last question is 'never breastfeed', "Why did you not breastfeed your baby?", 0=I was unwell, 1=I never had enough milk, 2=Elderly people in my family said so, 3=I had to start working soon after delivery, and 4=My baby always resisted having breast milk/ became sick after having breast milk.

Related to breastfeeding, we will also ask: "Does your child also bottle-feed formula or cow's/goat's milk?", 1=yes, 0=no. If the answer is 'yes', then we will ask the question: "When did you stop breastfeeding your child?, 1=When my baby was less than 3 months old, 2= When my baby was less than 6 months old, and 3= When my baby was 6 months old.

Finally, we will also ask them: "What else does your child have other than breast milk?", (i) Formula, (ii) Cow's/goat milk, (iii) Soft rice, (iv) Mashed vegetable, and (v) Dal (different types). We will ask them "How many times in a day do you breastfeed your child? (Number of breast-feeding)".

To understand the child's food consumption pattern, we will ask "How many times do you give semi-solid/solid food to your child?", '0' for under 6 months and 1 for over 6 months. In addition, to understand the quality of food consumption incidence, we will ask them: Whether their children ate the following items in the last 24 hours: (i) Rice, (ii) Green leafy vegetables, (iii) Potatoes, (iv) Eggs, (v) Chicken, (vi) Beef/Mutton, (vii) Dal, (viii) food cooked with oil/ghee/butter, (ix) Cow's/goat milk, (x) Food prepared with milk, (xi) Seasonal fruits, and (xii) Banana. The answer will be 1=Yes, 0=No.

To understand the intensity of food consumption, we will ask them about their food items in the last 7 days. How many days in the past 7 days your child had the following items: (i) Rice, (ii) Green leafy vegetables, (iii) Potatoes, (iv) Eggs, (v) Chicken, (vi) Beef/Mutton, (vii) Dal, (viii) food cooked with oil/ghee/butter, (ix) Cow's/goat milk, (x) Food prepared with milk, (xi) Seasonal fruits, and (xii) Banana. These questions will be answered on a 5-point scale: 0=Not at all or 0 days, 1= Rarely or 2 days in the last 7 days, 2= Sometimes or 3 days, 3=Often or 5 days, and 4=Always or 7 days. We will create an index by summing up the responses (so that it is between 0-48) and then taking the average (dividing by 48). For robustness, we code the activity as 1 if the respondent answered either 3 or 4, and 0 otherwise. Therefore, 1 corresponds to higher-quality food consumption. Using these food consumption indicators, we will construct composite indices following the above-detailed procedure.

2.4.3 Health Outcomes at Birth and up to 6 Months Postpartum

Pregnancy Complications We will create an index of pregnancy complications based on 5-item questions related to the dangers/complications that women may experience during their pregnancy. They will answer either yes (if she experiences the complication) or no (if she does

not experience it). We would assign 1 to items that are answered 'yes' and 0 for 'no', and then take the average. Therefore, this variable would range from 0 to 1, where a higher number would mean experiencing more pregnancy complications. We will also use this continuous score to construct the standardized index following the above-detailed procedure of constructing indices. For robustness, we will also try a complication index cut-off: equals 1 if the index score ≥ 0 and 0 otherwise. On pregnancy complications, we ask the following questions: Have you had any of the following when you were pregnant?

1. Headache that won't go away or gets worse over time
2. Dizziness or fainting
3. Changes in your vision
4. Fever over 100.4 degrees (F)
5. Extreme swelling of your hands and feet

If they experience any of the five times, then we will ask them: "What did you do if you experienced a complication?", 0= I rested, 1= I spoke to a family member /friend for help, 2= I consulted IVR project doctor, and 3= I consulted my midwife/doctor. For the five complications, if the answer is 2 or 3 then we will code them 1 and 0 for otherwise. Therefore, 1 corresponds to the right decisions to tackle these complications. Using these complication indicators, we will create a standardized complication index following the above-detailed procedure.

Baby's Illness We will create an index of a baby's illness based on 5-item questions related to specific illnesses that a baby may experience after birth. Index construction for this outcome based on these questions will follow the same format as above. We ask the following questions: Since birth, has your baby suffered from -

1. Common cold
2. Fever

3. Stomach problem

4. Pneumonia (your baby had a severe cold and difficulty breathing)

5. Jaundice (your baby turned yellow)

Birth Weight and Height We will also ask them about the baby's weight and height. If the birth weight is lower than 2500 g then this will be termed as 1 for low birth weight and 0 for otherwise. We will also be able to define the variable length/height-for-age to construct the z score based on World Health Organization (WHO) standards, and this score will differ by gender and birth weeks.

Preterm birth, Stillbirth, and Neonatal Death Our survey also includes questions about preterm birth and infant death outcomes. If the baby is born before 37 weeks of pregnancy, we will term it as preterm birth (=1) and otherwise (=0). We will also ask the question, "Did your baby die at birth?" If the answer is 'yes', then we will term this event as stillbirth (=1) and 0 for otherwise. Similarly, for neonatal death, we ask the following: "Did your baby die within 28 days after delivery?", Yes (=1) and No (=0).

Recovery After Delivery We will ask two questions to understand how speedy the mother's recovery was after the delivery: (i) "Are you feeling completely healthy after delivery?". If the answer to this question is 'yes' then we term it as full recovery (=1), and zero otherwise; (ii) "How long does it take to feel completely healthy after delivery?" to understand how speedy the recovery is. More days in the recovery process will indicate lower speedy recovery and vice-versa.

2.4.4 Secondary Outcomes

Depression Depression level would be measured using the 10-item version of the Center for Epidemiologic Studies Depression Scale (CES-D-10) (Andresen et al., 1994). The scale consists of 10 items that are answered on a 4-point scale (rarely or none of the time (less than 1 day) (= 0),

some or a little of the time (1-2 days) (= 1), occasionally or a moderate amount of time (3-4 days) (= 2), most of the time (5-7 days) (= 3)). Items 3 and 8 are reverse-scored. The CES-D-10 score is between 0 and 30, where a score greater than 10 means someone may have depression. Using this cut-off, we would create a binary variable that would equal 1 if the CES-D-10 score is above 10 and 0 otherwise. For robustness, we would assign 1 to items if the respondent's answer is either occasionally or a moderate amount of time, or most of the time, and assign 0 otherwise, and then take the average of these 10 responses. Therefore, this variable would range from 0 to 1, where a higher number would mean severe depression. Using these mental health-related questions, we will create a standardized CES-D index following the above-detailed procedure. The CES-D-10 questions are as follows:

In the last 7 days:

1. I was bothered by things that usually do not bother me.
2. I had trouble keeping my mind on what I was doing.
3. I felt depressed.
4. I felt like everything I did was an effort.
5. I felt hopeful about the future.
6. I felt fearful.
7. My sleep was restless.
8. I was happy.
9. I felt lonely.
10. I could not get going.

Edinburgh Postnatal Depression Scale (EPDS We would also measure the depression level of the mother by following the 10-item Edinburgh Postnatal Depression Scale (EPDS) (Cox et al., 1987). This scale consists of 10 items that are answered on a 4-point scale. Responses are scored 0, 1, 2, and 3 based on the seriousness of the symptoms that a mother felt during the last seven days. Items 3, 5 to 10 are reverse scored (i.e., 3, 2, 1, and 0). The total score is calculated by adding together the scores for each of the 10 items. Therefore, the EPDS-10 score is between 0 and 30, where a score more excellent than 12 means someone suffering from depression should seek medical attention.

Using this cut-off, we would create a binary variable that would equal 1 if the EPDS -10 score is above 12 and 0 otherwise. For robustness, we would also use another cut-off applicable to Bangladesh (Gausia et al., 2007); we would create a binary variable that would equal 1 if the EPDS -10 score is equal to or above 10 and 0 otherwise.

Moreover, we would create a binary variable that would assign 1 to items if the respondent's answer is either 2 or 3 and assign 0 otherwise, and then take the average of these 10 responses. Therefore, this variable would range from 0 to 1, where a higher number would mean severe depression. Using these mental health-related questions, we will create a standardized EPDS index following the above-detailed procedure.

The EPDS-10 questions are as follows: In the last 7 days:

1. I have been able to laugh and see the funny side of things.
2. I have looked forward with enjoyment to things.
3. I have blamed myself unnecessarily when things went wrong.
4. I have been anxious or worried for no good reason.
5. I have felt scared or panicky for no very good reason.
6. Things have been getting on top of me.
7. I have been so unhappy that I have had difficulty sleeping.

8. I have felt sad or miserable.

9. I have been so unhappy that I have been crying.

10. The thought of harming myself has occurred to me.

Maternal Mental Health Awareness In addition to measuring the CES-D score and EPDS scale, we will ask questions to understand the awareness of maternal mental health. We will ask the women four questions where the answer will be either true or false. The answer to the other three questions will be true except for the second question. Therefore, we will reverse the answer from question 2 to make it trend similar to the other three questions. Then, we will take the average of these four responses. Therefore, this variable would range from 0 to 1, where a higher number would mean more awareness about mental health. We will create a standardized mental health awareness index following the above-detailed procedure. The four questions on maternal health awareness are as follows:

1. If a woman is consistently feeling sad during pregnancy, she should discuss this with her doctor or midwife.

2. If a pregnant woman or a new mother seeks health from a doctor or midwife for her mental health, it is clear that she has gone crazy (pagol)

3. Talking to family and friends about her inner feelings can help a pregnant woman or a new mother cope with the sadness, stress, and anxiety they often experience.

4. If a new mother is consistently feeling sad, she might have difficulty bonding with her baby.

We will also ask some questions to understand the awareness of maternal mental health depending on different situations. We will ask 4 questions where the answer is option 4 out of five options. The options are "I will not have any advice for her because I don't know what to do in this case (=0)", "This is a personal matter, so I will not advise anything (=1)", "Everyone can feel sad sometimes, there is nothing that can be done (=2)", "I will advise her not to discuss anything

about how she feels as people might think she has gone mad (=3)", and I will try to talk to her and encourage her to talk to her family and if necessary to a doctor (=4).

We will term the answer as the right answer if they indicate option 4 (=1), and for the wrong answer, we will term it as no (=0). Then, we will take the average of these four responses. Therefore, this variable would range from 0 to 1, where a higher number would mean more awareness about mental health. The four questions are as follows:

1. Suppose your friend or neighbor is a new mother, and she is constantly feeling sad and often crying. What advice would you give her?

2. Suppose your friend or neighbor is not pregnant, and she is constantly irritable and agitated towards others (overly sensitive and crying without any serious issues). What advice would you give her?

3. If you come to know about a woman who is pregnant or has given birth recently, constantly having anxiety, always irritable, and often crying, what would you think of her?

4. Your friend or neighbor is a new mother. She is struggling to take care of her newborn. This is making her sad and anxious. What do you think she should be doing?

Food Consumption of Older Children in the Household To explore spillovers in terms of other children's food consumption patterns, we will ask the respondents to state the food consumption by other children aged 5 years or under (excluding the newborn). We will ask about this food consumption pattern separately for both boys and girls to understand the gender difference within the household (if any). We will ask them how often each of their older children eats the following items in a week: (i) Meat, (ii) Eggs, (iii) Vegetables, (iii) Lentils/pulses, (iv) Seasonal fruits, and (v) Milk.

These questions will be answered on a 5-point scale: 0=Never, 1= Seldom (1/2 day), 2= Sometimes (3/4 day), 3= Often (5/6 day), and 4= Always (7 days). We will create an index by summing up the responses (so that it is between 0-20) and then taking the average (dividing by 20) so that

this index is between 0 and 1 (where 1 indicates higher quality food consumption in the last 7 days). To check the robustness, we code the activity as 1 if the respondent answered either 3 or 4, and 0 otherwise, Therefore, 1 corresponds to higher quality food consumption. Using these food consumption indicators, we also will create a standardized food consumption index following the above-detailed procedure.

Illness of Older Children To explore spillovers in terms of other children's illness patterns, we will ask the respondents to state how often each of their older children has had these common illnesses in the last 6 months: (i) Common cold, (ii) Fever, and (iii) Stomach problem. We will ask for this information for both the boys and girls. We will take the frequency of these illnesses in the last 6 months and then we will count the total number of illnesses. Therefore, a higher number would mean experiencing more illness. We will also use this continuous score to construct the standardized index. We would also assign 1 to an illness if they experienced it at least once in the last 6 months and 0 for otherwise.

Moreover, we will also ask them about their doctor's visits for older children in the last 6 months. We will ask them: How often did it happen that any of your older children was unwell and needed treatment from a doctor, but it had to be skipped? This question will be answered on a 4-point scale: 0= Always, 1= Sometimes (3/4 day), 2= Once or twice, 3= Never. We will ask the same questions for all the other children and we will also track their gender.

2.5 Power Analysis

We perform a power calculation using our baseline data to determine the minimum detectable effect size (MDE). The key parameters are as follows: α = 0.05. (standard type I error), κ = 0.80 (standard power), J = 110 (number of clusters (villages) per treatment arm), number of respondents under three treatment arms T1 = 1359 (control), T2 = 1374 (IVR only Treatment), and T3 = 1371 (IVR + Access to Doctor Treatment), and we used the low ICC = 0.15 (intra-cluster correlation coefficient). For the actual power analysis, we round off the latter two numbers to 1370.

Using one of our main outcome variables, the required number of Ante-Natal Care (ANC) visits, we compute the MDE. According to our baseline, only 2% of respondents in the three treatment arms have access to the required number of ANC visits (see Panel A of Table 5). Therefore, our study is sufficiently powered to detect an effect size of 0.0331 or 3.31 percentage points. In case we experience very high attrition of roughly 30% (which would reduce the treatment sample to 959 and the control sample to 951), the minimum detectable effect size changes very little to 0.0359. For the MDE, we need less than 4 percentage points improvement in ANC visits, and given such a low base we expect the impact will be much higher.

We implement a similar exercise for a second outcome of interest, Post-Natal Care (PNC) visits. According to the baseline, only 5% of women (overall) have access to four required number of PNC visits (see Panel A of Table 5). Our study is sufficiently powered to detect an effect size of 0.0491, which is 4.91 percentage points (or 0.0528, with 30% attrition). For the MDE, we need less around 5 percentage points improvement in PNC visits, and given such a low base we expect a much higher treatment effect.

Similarly, we also compute the MDE of the treatment for some other outcome variables such as knowledge of five risk factors during pregnancy, providing additional food to the child after six months, breastfeeding duration, breastfeeding time, and going to Kabiraj (witch doctor) during complexity. Our study is sufficiently powered to detect an effect size of (i) 0.0381 for knowledge of five risk factors (or 0.0412 with 30% attrition); (ii) 0.0611 (or 0.0647) for knowledge of providing additional food to the child after six months of birth; (iii) 0.0882 (or 0.0939) for knowledge of breastfeeding duration (1= if continue breastfeeding until the age of 6 months); (iv) 0.0775 (or 0.0823) for knowledge of breastfeeding time (=1 if started within an hour); and (v) 0.0865 (or 0.0823). for knowledge of going to Kabiraj (witch doctor) during complexity (=1 if say yes).

3 Empirical Methodology

3.1 Baseline Analysis

Since we have a randomized design, the methods employed are straightforward. We estimate the following empirical specification using ordinary least squares (OLS) to explore the treatment effects of our intervention,

$$Y_{iv} = \beta_0 + \beta_1 T2_{iv} + \beta_2 T3_{iv} + \gamma \mathbf{X_{iv}} + \epsilon_{iv} \tag{1}$$

where Y_{iv} corresponds to an outcome measure like the number of ANC visits or an index of maternal health awareness. $T2_{iv}$ and $T3_{iv}$ are treatment indicators as defined in section 2.2. The coefficients on these variables, β_1 and β_2 give us the relevant treatment effects of our intervention. We will cluster the standard errors at the village level. We will estimate equation (1) both with and without a set of covariates specified in the vector $\mathbf{X_{iv}}$. These include standard predetermined demographic characteristics of the mother and father of the newborn baby including age at birth, age at marriage, length of current relationship status, occupation identifiers, and other wealth indicators. The balance of these variables was perfectly achieved at the baseline and their inclusion is not required for identification of the treatment effects but we may have some efficiency gains by adding these to our estimation models.

All estimates from equation (1) will also be accompanied by p-values based on procedures developed in Westfall and Young (1993) to deal with issues arising from multiple hypothesis testing.

We will undertake a number of key extensions of the above analysis to enrich our understanding of the impact of our treatment intervention and help generate potential policy implications that may have external validity. The next three sections detail our plan for implementing these extensions.

3.2 Intensity of Treatment

A key dimension that we want to explore in our setting is the variation in treatment effects as a function of the intensity of treatment received by the respondents. We will approach this in two ways. First, we will estimate how the effects vary by the trimester of enrollment in the program. Around 30% of women in our sample were in their first trimester at the time of enrollment while around 65% were in the second trimester. We will interact these identifiers with the treatment indicators and re-estimate a variant of equation (1). Our hypothesis is that women who were enrolled early on during their pregnancy will have a larger treatment effect size.

Second, we are carefully recording the intensity of interaction that our treated mothers are having with the IVR system. For instance, we will interact with the continuous variable measuring the number of times a mother placed a call to the IVR system with the treatment indicator. This interaction effect will provide an estimate of the association between per call placed and our outcomes of interests. These estimates, though not causal, can offer suggestive channels for understanding the efficacy of the intervention from a cost perspective. Since the marginal cost of an additional call is near zero, as the modules and the recordings are uploaded only once, a positive association will imply the value of having such telehealth services.

3.3 Heterogeneity Analysis

Our next extension revolves around using salient and policy-relevant pre-determined respondent characteristics to study treatment effect heterogeneity. These are standard in both the experimental and non-experimental literature in causal inference where researchers estimate either interacted models, i.e., treatment status multiplied by the pre-determined characteristic of interest, H_i or they split the sample by the latter and re-estimate equation (1) for both sub-samples.

$$Y_{iv} = \beta_0 + \beta_1 T2_{iv} + \beta_2 T3_{iv} + \beta_3 T2_{iv} * H_{iv} + \beta_4 T3_{iv} * H_{iv} + \beta_5 H_i + \gamma \mathbf{X_{iv}} + \epsilon_{iv} \qquad (2)$$

There are a number of key dimensions that we plan to study in this paradigm. First, we will look at differences in effects for mothers pregnant with boys compared to girls. There is considerable evidence in the literature about the prevalence of son-preference and conservative gender norms in South Asia (e.g., Jayachandran and Pande (2017); Khalil and Mookerjee (2019)).

Second, we will study differences by the order of the current pregnancy. In our sample, women on average have had 1.73 pregnancies. We will thus consider heterogeneity by whether the current pregnancy is a mother's first one as the latter may indicate larger treatment effects due to lower baseline knowledge of pregnancy-related information.

Third, since a number of our outcomes are self-reported we may run into an issue of conformation by the respondents also referred to as social desirability bias (SDB). We will adopt the 13-point SDB scale developed by Dhar et al. (2022), which in turn closely follows Reynolds (1982).

Following the recent approach in the literature, we will interact an index of SDB with the treatment indicators to study whether the respondents in our sample are being influenced by SDB-related channels.

Fourth, we will also study heterogeneity in effects by an index of baseline knowledge of maternal, neonatal, and pregnancy-related issues. As shown in Table 5, we have collected this information during the baseline survey. We will combine the 10 listed measures in Panel A and Panel B into an index and study whether treatment effects vary by this dimension. Mothers who started with a lower base of knowledge will stand to benefit more from our intervention.

3.4 Potential Mechanisms

Empirically establishing the precise mechanisms behind the baseline findings in any design, experimental or non-experimental, is fairly challenging. In our setup, we posit two or three distinct channels that can explain any estimated effects that we find. First, since only 2% of our sample is aware of the recommended number of ANC visits if the program is successful in inducing mothers to go for a higher number of visits then it provides a direct, almost mechanical, channel for improvement in post-natal birth outcomes. We will collect information on the total number of

ANC visits by the mothers in all three arms which will help us explore this dimension.

Second, although mentioned in the previous sub-section, improvements in the knowledge base in the relevant domain will also provide a distinct mechanism that can explain our treatment effects. If there are differences in effect sizes by this dimension then it will imply that more information and awareness is required to improve birth outcomes at least among rural populations in developing countries.

Third, over one-third of women in our sample have reported at baseline to have gone to traditional healers or witch doctors to avoid or seek treatment for pregnancy-related complexities. Our intervention may dissuade women from approaching such healers who can be directly detrimental to the health of the mother and the fetus. Therefore, if we document substitution effects towards practitioners of modern medicine it will provide a clean and distinct causal mechanism for explaining our treatment effects.

3.5 Attrition

We examine the attrition from the survey in a number of ways. First, we will compare the baseline characteristics of women who dropped out to the baseline characteristics of women who remained in the study (both within and across treatment and control groups) to check if attrition is selective. In addition, we will also check if the rate of attrition between treatment and control groups differs (differential attrition). In the case of differential attrition, we will use Inverse Probability Weighting and Lee (2009) bounds to address attrition bias concerns.

Table 2: Balance Check for Maternal and Household Characteristics across Treatment Arms

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	All	T1	T2	T3	T2-T1	T3-T1	T3-T2
Panel A	Characteristics of Women						
Age (in years)	24.11	24.05	24.11	24.18	0.062	0.124	0.062
	(5.87)	(5.89)	(5.91)	(5.81)	(0.250)	(0.235)	(0.235)
Age at marriage (in years)	17.90	17.85	17.88	17.98	0.032	0.130	0.098
	(3.25)	(3.35)	(3.08)	(3.30)	(0.155)	(0.153)	(0.147)
Age at first child (in years)	19.85	19.85	19.84	19.85	-0.014	-0.002	0.012
	(3.44)	(3.77)	(3.21)	(3.32)	(0.241)	(0.234)	(0.200)
Literacy (=1 can read and write)	0.98	0.98	0.98	0.97	0.002	-0.006	-0.009
	(0.14)	(0.14)	(0.13)	(0.16)	(0.005)	(0.006)	(0.006)
Schooling (in years)	8.84	8.86	8.92	8.72	0.062	-0.136	-0.198
	(2.94)	(2.98)	(2.85)	(3.01)	(0.143)	(0.153)	(0.145)
Occupation (=1 if housewife)	0.93	0.93	0.94	0.91	0.002	-0.024	-0.026
	(0.26)	(0.25)	(0.24)	(0.28)	(0.013)	(0.017)	(0.017)
Number of pregnancy	1.73	1.73	1.73	1.72	-0.005	-0.017	-0.012
	(0.80)	(0.81)	(0.82)	(0.76)	(0.037)	(0.034)	(0.034)
Number of children	0.68	0.66	0.69	0.69	0.025	0.029	0.004
	(0.76)	(0.74)	(0.78)	(0.76)	(0.033)	(0.031)	(0.031)
Desired number of children	2.10	2.09	2.08	2.13	-0.009	0.039	0.048
	(0.54)	(0.52)	(0.57)	(0.53)	(0.034)	(0.029)	(0.031)
Read Newspaper (=1 if $\geqslant$ once a week)	0.03	0.03	0.03	0.04	0.003	0.007	0.004
	(0.18)	(0.17)	(0.18)	(0.19)	(0.009)	(0.010)	(0.010)
Watch TV (=1 if $\geqslant$ once a week)	0.34	0.32	0.33	0.36	0.010	0.040	0.030
	(0.47)	(0.47)	(0.47)	(0.48)	(0.031)	(0.032)	(0.029)
Listen Radio (=1 if $\geqslant$ once a week)	0.02	0.01	0.01	0.02	-0.000	0.004	0.004
	(0.12)	(0.12)	(0.12)	(0.13)	(0.007)	(0.008)	(0.008)
Computer internet (=1 if $\geqslant$ once a week)	0.04	0.04	0.04	0.05	0.004	0.009	0.005
	(0.20)	(0.19)	(0.20)	(0.21)	(0.010)	(0.012)	(0.012)
Use Phone/internet (=1 if $\geqslant$ once a week)	0.37	0.36	0.39	0.37	0.026	0.004	-0.022
	(0.48)	(0.48)	(0.49)	(0.48)	(0.030)	(0.029)	(0.029)
Panel B	Household characteristics						
Monthly income (in 1000s BDT)	16.15	16.13	15.99	16.32	-141.6	186.3	327.8
	(10.87)	(10.52)	(10.88)	(11.18)	(583.7)	(679.3)	(678.1)
Number of people eat together	4.33	4.35	4.34	4.29	-0.009	-0.065	-0.056
	(1.66)	(1.64)	(1.73)	(1.59)	(0.079)	(0.083)	(0.086)
Number of female earners	0.11	0.10	0.10	0.13	0.002	0.031	0.029
	(0.35)	(0.31)	(0.35)	(0.38)	(0.021)	(0.023)	(0.025)
Number of male earners	1.40	1.39	1.40	1.42	0.016	0.029	0.013
	(0.64)	(0.63)	(0.65)	(0.64)	(0.031)	(0.030)	(0.031)

Note: The first four columns report the mean of the corresponding variable with standard deviations in brackets. The last three columns report the difference between treatments with standard errors in parentheses clustered at the village level. T1, T2, and T3 corresponds to Control, IVR only, and IVR with Doctor groups.

Table 3: Balance Check for Paternal Characteristics across Treatment Arms

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	All	T1	T2	T3	T2-T1	T3-T1	T3-T2
Age (in years)	30.71	30.57	30.79	30.76	0.215	0.183	-0.032
	(6.69)	(6.72)	(6.75)	(6.61)	(0.285)	(0.272)	(0.280)
Age at marriage (in years)	24.50	4.33	24.53	24.65	0.207	0.320	0.113
	(4.70)	(4.60)	(4.79)	(4.68)	(0.206)	(0.206)	(0.220)
Years of current marital status	6.20	6.25	6.25	6.11	0.008	-0.140	-0.148
	(5.41)	(5.42)	(5.56)	(5.25)	(0.227)	(0.212)	(0.229)
Literacy (=1 can read and write)	0.94	0.95	0.94	0.94	-0.009	-0.011	-0.002
	(0.24)	(0.23)	(0.24)	(0.25)	(0.010)	(0.011)	(0.012)
Schooling (in years)	8.54	8.70	8.53	8.39	-0.177	-0.315	-0.137
	(4.29)	(4.31)	(4.33)	(4.21)	(0.223)	(0.219)	(0.220)
Involvement in income (=1 if yes)	0.99	1.00	0.99	0.99	-0.004	-0.001	0.002
	(0.07)	(0.06)	(0.09)	(0.07)	(0.003)	(0.003)	(0.003)
Agriculture (=1 if yes)	0.28	0.29	0.28	0.27	-0.010	-0.019	-0.008
	(0.45)	(0.45)	(0.45)	(0.44)	(0.027)	(0.028)	(0.026)
Day laborer (=1 if yes)	0.40	0.40	0.41	0.39	0.012	-0.001	-0.013
	(0.49)	(0.49)	(0.49)	(0.49)	(0.028)	(0.029)	(0.028)
Owns Business (=1 if yes)	0.20	0.20	0.19	0.20	-0.001	0.006	0.008
	(0.40)	(0.40)	(0.40)	(0.40)	(0.019)	(0.020)	(0.020)
Govt. or private job (=1 if yes)	0.12	0.12	0.12	0.13	-0.001	0.013	0.013
	(0.33)	(0.32)	(0.32)	(0.34)	(0.013)	(0.013)	(0.014)

Note: The first four columns report the mean of the corresponding variable with standard deviations in brackets. The last three columns report the difference between treatments with standard errors in parentheses clustered at the village level. T1, T2, and T3 corresponds to Control, IVR only, and IVR with Doctor groups.

Table 4: Balance Check for Current Pregnancy Characteristics across Treatment Arms

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	All	T1	T2	T3	T2-T1	T3-T1	T3-T2
Panel A	Current Pregnancy Characteristics						
First trimester (=1 if yes)	0.32 (0.47)	0.31 (0.46)	0.32 (0.47)	0.32 (0.47)	0.006 (0.027)	0.010 (0.025)	0.004 (0.023)
Second trimester (=1 if yes)	0.65 (0.48)	0.65 (0.48)	0.65 (0.48)	0.65 (0.48)	-0.002 (0.028)	-0.001 (0.026)	0.001 (0.024)
Third trimester (=1 if yes)	0.03 (0.17)	0.03 (0.18)	0.03 (0.17)	0.02 (0.16)	-0.004 (0.009)	-0.008 (0.009)	-0.004 (0.007)
Tested at the diagnostic center (=1 if yes)	0.04 (0.18)	0.03 (0.18)	0.03 (0.18)	0.04 (0.20)	-0.000 (0.010)	0.009 (0.013)	0.010 (0.013)
Tested with test kit (=1 if yes)	0.76 (0.43)	0.77 (0.42)	0.74 (0.44)	0.76 (0.43)	-0.037 (0.039)	-0.018 (0.039)	0.019 (0.037)
Month of pregnancy confirmation	1.81 (0.59)	1.81 (0.57)	1.84 (0.60)	1.78 (0.61)	0.031 (0.041)	-0.028 (0.042)	-0.059 (0.039)
Panel B	Healthcare Access and Behavior (during current pregnancy)						
Doctor Consultation Govt. hospital (=1 if yes)	0.30 (0.46)	0.30 (0.46)	0.32 (0.47)	0.27 (0.44)	0.023 (0.034)	-0.030 (0.032)	-0.053* (0.027)
Doctor Consultation Private hospital (=1 if yes)	0.19 (0.40)	0.18 (0.38)	0.20 (0.40)	0.20 (0.40)	0.026 (0.023)	0.026 (0.023)	0.000 (0.024)
Midwife Consultation (=1 if yes)	0.03 (0.16)	0.03 (0.18)	0.03 (0.17)	0.02 (0.14)	-0.002 (0.011)	-0.013 (0.009)	-0.011 (0.008)
Blood test (=1 if yes)	0.54 (0.50)	0.56 (0.50)	0.52 (0.50)	0.53 (0.50)	-0.043 (0.033)	-0.031 (0.034)	0.011 (0.032)
Urine test (=1 if yes)	0.66 (0.47)	0.68 (0.47)	0.65 (0.48)	0.65 (0.48)	-0.023 (0.033)	-0.028 (0.039)	-0.005 (0.038)
Blood pressure test (=1 if yes)	0.51 (0.50)	0.53 (0.50)	0.49 (0.50)	0.51 (0.50)	-0.037 (0.031)	-0.014 (0.035)	0.023 (0.033)
	Tetanus Injection Dose						
First dose (=1 if yes)	0.69 (0.46)	0.67 (0.47)	0.71 (0.45)	0.70 (0.46)	0.045 (0.040)	0.031 (0.043)	-0.014 (0.042)
Second dose (=1 if yes)	0.61 (0.49)	0.59 (0.49)	0.60 (0.49)	0.63 (0.48)	0.015 (0.042)	0.039 (0.045)	0.025 (0.044)
Third dose (=1 if yes)	0.52 (0.50)	0.51 (0.50)	0.52 (0.50)	0.52 (0.50)	0.013 (0.042)	0.006 (0.044)	-0.006 (0.044)
Fourth dose (=1 if yes)	0.41 (0.49)	0.42 (0.49)	0.40 (0.49)	0.40 (0.49)	-0.024 (0.038)	-0.016 (0.041)	0.008 (0.038)
Fifth dose (=1 if yes)	0.35 (0.48)	0.32 (0.47)	0.36 (0.48)	0.38 (0.49)	0.040 (0.033)	0.064* (0.036)	0.023 (0.035)

Note: The first four columns report the mean of the corresponding variable with standard deviations in brackets. The last three columns report the difference between treatments with standard errors in parentheses clustered at the village level. T1, T2, and T3 corresponds to Control, IVR only, and IVR with Doctor groups.

Table 5: Balance Check for Pregnancy and Neonatal Knowledge across Treatment Arms

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	All	T1	T2	T3	T2-T1	T3-T1	T3-T2
Panel A	Knowledge related to mother and newborn						
Ante-Natal Care visits (=1 if say 8 visits)	0.02 (0.14)	0.02 (0.15)	0.02 (0.15)	0.01 (0.12)	0.001 (0.009)	-0.008 (0.008)	-0.009 (0.006)
Post-Natal Care visits (=1 if say 4 visits)	0.05 (0.21)	0.06 (0.23)	0.05 (0.22)	0.04 (0.18)	-0.006 (0.014)	-0.022 (0.015)	-0.016* (0.009)
Breastfeeding time (=1 started within an hour)	0.69 (0.46)	0.70 (0.46)	0.66 (0.47)	0.71 (0.45)	-0.038 (0.033)	0.012 (0.033)	0.050 (0.031)
Breastfeeding duration (=1 continue until age 6 months)	0.49 (0.50)	0.47 (0.50)	0.49 (0.50)	0.51 (0.50)	0.016 (0.050)	0.039 (0.051)	0.023 (0.050)
Additional Food (=1 started after 6 months)	0.85 (0.36)	0.83 (0.37)	0.84 (0.37)	0.87 (0.33)	0.002 (0.024)	0.038 (0.023)	0.036* (0.021)
Panel B (1 = yes)	Knowledge about Risk Factors during Pregnancy						
Excessive bleeding during pregnancy or delivery	0.29 (0.45)	0.30 (0.46)	0.30 (0.46)	0.26 (0.44)	-0.001 (0.037)	-0.041 (0.036)	-0.040 (0.033)
Fever for 3 days during pregnancy or delivery	0.17 (0.38)	0.16 (0.37)	0.17 (0.38)	0.18 (0.38)	0.008 (0.032)	0.018 (0.031)	0.010 (0.026)
Retaining water during pregnancy or delivery	0.17 (0.38)	0.18 (0.38)	0.18 (0.38)	0.16 (0.37)	-0.000 (0.030)	-0.013 (0.030)	-0.013 (0.026)
Convulsion during pregnancy or delivery	0.25 (0.43)	0.27 (0.44)	0.25 (0.44)	0.22 (0.41)	-0.015 (0.031)	-0.053* (0.030)	-0.037 (0.029)
Prolonged labor or breeched birth signs	0.14 (0.35)	0.16 (0.37)	0.14 (0.34)	0.13 (0.34)	-0.022 (0.024)	-0.028 (0.024)	-0.006 (0.022)
Know all five risk factors	0.04 (0.20)	0.03 (0.18)	0.05 (0.21)	0.05 (0.21)	0.016 (0.012)	0.016 (0.011)	0.000 (0.012)
Panel C	Belief in Alternative Medicine/Folk Remedies						
Used blessed water and oil (=1 if yes)	0.44 (0.50)	0.42 (0.49)	0.44 (0.50)	0.46 (0.50)	0.018 (0.031)	0.037 (0.031)	0.019 (0.031)
Using amulet to avoid complexities (=1 if yes)	0.36 (0.48)	0.35 (0.48)	0.38 (0.48)	0.36 (0.48)	0.031 (0.029)	0.020 (0.029)	-0.012 (0.029)
Go to Kabiraj (witch doctor) during complexity (=1 if yes)	0.37 (0.48)	0.36 (0.48)	0.38 (0.49)	0.37 (0.48)	0.025 (0.029)	0.011 (0.028)	-0.014 (0.027)

Note: The first four columns report the mean of the corresponding variable with standard deviations in brackets. The last three columns report the difference between treatments with standard errors in parentheses clustered at the village level. T1, T2, and T3 corresponds to Control, IVR only, and IVR with Doctor groups.

References

Abebe, H., G. A. Beyene, and B. S. Mulat (2021). Harmful cultural practices during perinatal period and associated factors among women of childbearing age in southern ethiopia: community based cross-sectional study. *PLoS One 16*(7), e0254095.

Almond, D. and J. Currie (2011). Killing me softly: The fetal origins hypothesis. *The Journal of Economic Perspectives 25*(3), 153–172.

Almond, D., J. Currie, and V. Duque (2018). Childhood circumstances and adult outcomes: Act ii. *Journal of Economic Literature 56*(4), 1360–1446.

Almond, D. and B. Mazumder (2011). Health capital and the prenatal environment: the effect of ramadan observance during pregnancy. *American Economic Journal: Applied Economics 3*(4), 56–85.

Amarante, V., M. Manacorda, E. Miguel, and A. Vigorito (2016). Do cash transfers improve birth outcomes? evidence from matched vital statistics, and program and social security data. *American Economic Journal: Economic Policy 8*(2), 1–43.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association 103*(484), 1481–1495.

Andresen, E. M., J. A. Malmgren, W. B. Carter, and D. L. Patrick (1994). Screening for depression in well older adults: Evaluation of a short form of the ces-d. *American journal of preventive medicine 10*(2), 77–84.

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

Buekens, P., M. Kotelchuck, B. Blondel, F. B. Kristensen, J.-H. Chen, and G. Masuy-Stroobant

(1993). A comparison of prenatal care use in the united states and europe. *American Journal of Public Health 83*(1), 31–36.

Carneiro, P., L. Kraftman, G. Mason, L. Moore, I. Rasul, and M. Scott (2021). The impacts of a multifaceted prenatal intervention on human capital accumulation in early life. *American Economic Review 111*(8), 2506–49.

Chorniy, A., J. Currie, and L. Sonchak (2020). Does prenatal wic participation improve child outcomes? *American Journal of Health Economics 6*(2), 169–198.

Choudhury, N., A. C. Moran, M. A. Alam, K. Z. Ahsan, S. F. Rashid, and P. K. Streatfield (2012). Beliefs and practices during pregnancy and childbirth in urban slums of dhaka, bangladesh. *BMC public health 12*, 1–6.

Cox, J. L., J. M. Holden, and R. Sagovsky (1987). Detection of postnatal depression: development of the 10-item edinburgh postnatal depression scale. *The British journal of psychiatry 150*(6), 782–786.

Dhar, D., T. Jain, and S. Jayachandran (2022). Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in india. *American economic review 112*(3), 899–927.

Doyle, O. (2020). The first 2,000 days and child skills. *Journal of Political Economy 128*(6), 2067–2122.

Doyle, O. (2022). Can early intervention have a sustained effect on human capital? *Journal of Human Resources*, 0321–11557R1.

Gausia, K., C. Fisher, S. Algin, and J. Oosthuizen (2007). Validation of the bangla version of the edinburgh postnatal depression scale for a bangladeshi sample. *Journal of Reproductive and Infant Psychology 25*(4), 308–315.

Grossman, D. and U. Khalil (2022). Neighborhood crime and infant health. *Journal of Urban Economics 130*, 103457.

Hoynes, H., D. W. Schanzenbach, and D. Almond (2016). Long-run impacts of childhood access to the safety net. *American Economic Review 106*(4), 903–934.

Jayachandran, S. and R. Pande (2017). Why are indian children so short? the role of birth order and son preference. *American Economic Review 107*(9), 2600–2629.

Khalil, U. and S. Mookerjee (2019). Patrilocal residence and women's social status: evidence from south asia. *Economic Development and Cultural Change 67*(2), 401–438.

Lee, D. S. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *The Review of Economic Studies 76*(3), 1071–1102.

Parker, S. W. and P. E. Todd (2017). Conditional cash transfers: The case of progresa/oportunidades. *Journal of Economic Literature 55*(3), 866–915.

Reynolds, W. M. (1982). Development of reliable and valid short forms of the marlowe-crowne social desirability scale. *Journal of clinical psychology 38*(1), 119–125.

Schwab, B., S. Janzen, N. P. Magnan, and W. M. Thompson (2020). Constructing a summary index using the standardized inverse-covariance weighted average of indicators. *The Stata Journal 20*(4), 952–964.

UNICEF (2022). Antenatal care.

Van Campenhout, B., D. J. Spielman, and E. Lecoutere (2021). Information and communication technologies to provide agricultural advice to smallholder farmers: Experimental evidence from uganda. *American Journal of Agricultural Economics 103*(1), 317–337.

Walter, T. F., M. Kremer, O. Reich, Z. Sun, S. van Herwaarden, and H. Yesigat (2020). Using data for development: Evidence from a phone system for agricultural advice.

Wang, L. C., M. Vlassopoulos, A. Islam, and H. Hassan (2023). Delivering remote learning using a low-tech solution: Evidence from a randomized controlled trial in bangladesh.

WB (2023). Global health workforce statistics.

Westfall, P. H. and S. S. Young (1993). *Resampling-based multiple testing: Examples and methods for p-value adjustment*, Volume 279. John Wiley & Sons.

WHO (2021). New global targets to prevent maternal deaths.

Withers, M., N. Kharazmi, and E. Lim (2018). Traditional beliefs and practices in pregnancy, childbirth and postpartum: A review of the evidence from asian countries. *Midwifery 56*, 158–170.