IMAGINE Endline Evaluation Report

Methodology and findings from CARE IMAGINE programs in Bangladesh and Niger.

July 11, 2023

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

LIST OF ACRONYMS

ANCOVA	Analysis Of Covariance
CHW	Community Health Worker
DiD	Difference In Differences
FDR	False Discovery Rate (Benjamini-Hochberg)
FIML	Full Information Maximum Likelihood
FP	Family Planning
IMAGINE	Inspiring Married Adolescent Girls to Imagine New Empowered Futures
IPW	Inverse Probability Weights
IUD	Intrauterine Device
LCS	Latent Change Score models
MCAR	Missing Completely At Random
MSM	Marginal Structural Models
SEM	Structural Equation Model
VSLA	Village Savings and Loan Association
WLSMV	Weighted Least Squares Means and Variance adjusted estimation

EXECUTIVE SUMMARY

This report presents the findings on the CARE's Inspiring Married Adolescent Girls to Imagine New Empowered Futures (IMAGINE) intervention in Bangladesh and Niger. The goals of the IMAGINE program were broadly to provide knowledge and resources to married adolescent girls to improve a wide variety of health and economic outcomes.

Overview of IMAGINE

Ninety percent of adolescent pregnancies in the global South are to married girls, and complications from pregnancy and childbirth are a leading cause of death among this population. Empowering adolescents to delay the timing of their first pregnancy can result in positive health and economic outcomes, such as healthy pregnancies and infants, completing their education, and economic prosperity. Unfortunately, married adolescent girls face numerous social and structural barriers that impact their ability to delay childbearing.

In 2017, CARE embarked on formative research, qualitative and quantitative approaches to: 1) understand the barriers and facilitators that influence a married girl's ability to delay pregnancy, and 2) identify alternative futures that could inspire girls, families and communities to support this delay.

The IMAGINE intervention started in January 2019 and ended in May 2021, with a primary goal of postponing the first birth after marriage by 6 months. Girls aged 15-19 with no current or prior pregnancies in Bangladesh (four geographic regions) and Niger (six geographic regions) were selected as the target population for the IMAGINE intervention. In the IMAGINE program, Girls' Collectives (or small groups led by trained female facilitators from the region) served as a platform for enhancing married and unmarried adolescent girls' social support and access to information, including sexual and reproductive health and rights, decision-making, communication skills, and gender and social norms. The Collectives also functioned as a platform for training and knowledge transfer on essential vocational and financial topics.

Evaluation Design

The IMAGINE evaluation was carefully constructed to examine the impact of the two-year intervention on a) the timing (delay) of first birth among married adolescents and b) numerous additional outcomes targeting sexual and reproductive health, health service utilization, economic factors and personal agency. Geographic regions in Bangladesh and Niger were matched and assigned to either the IMAGINE treatment condition or no-intervention comparison (control) condition. The IMAGINE evaluation also featured an extensive village enumeration and sampling process to ensure sufficient sample size for evaluation of the intervention.

Sampled respondents completed a baseline survey (in November - December 2018) prior to the start of the intervention and a follow-up survey after the two-year intervention (in December

2021 - January 2022). A complex data weighting process was also carried out to improve the representativeness of the data and balance of treatment and control groups across several demographic factors.

Analyses were conducted using the baseline and follow-up data to evaluate the impact of the IMAGINE intervention on delaying first childbirth, along with many other priority outcomes such as contraceptive use. In addition, a dose-response analysis among the treatment areas examined the effect of higher participation in the IMAGINE intervention on delay-of-birth and family planning use outcomes.

Key Findings

In Bangladesh…

- For the primary outcome analysis, there was no significant difference between treatment and control on timing of birth after marriage. However, rates of marriage declined in the treatment group (relative to control) during the study, which could have impacted the ability to detect a delay-of-birth outcome.
- Despite the above finding, a dose-response analyses (within the intervention areas only) revealed that higher levels of participation in the IMAGINE program were significantly associated with a delay of first birth.
- The treatment group also showed greater improvements in family planning perceptions, reproductive health knowledge, and psychosocial outcomes than respondents in the control group.
- There were no intervention differences in family planning use outcomes and few differences in health and economic indicators.

In Niger…

- For the primary outcome analysis, no significant effect of treatment on delay of first birth was observed in Niger.
- In the dose-response analysis (within the intervention areas only), higher levels of participation in the IMAGINE program were not associated with a delay of first birth.
- Niger treatment respondents showed significantly higher rates of current, lifetime, and effective contraception use compared to the control group. For example, 34.1% of Niger treatment respondents reported lifetime use of contraception, compared to 19.1% in the control group.
- The Niger treatment group also reported greater health service utilization, social and economic mobility, and engagement in income-generating activities than respondents in the control group.

OVERVIEW

In low- and middle-income countries, approximately 90% of adolescent pregnancies occur among married girls (UNICEF, 2014). This puts their highest future potential, and in some cases their lives, at risk. Yet, sexual and reproductive health initiatives often fail to reach married girls. Typical initiatives focus on either preventing child marriage by targeting unmarried adolescents, or by serving adult married women. This approach tends to ignore married adolescents who thereby lack the services and support they need to lead healthy and productive lives.

To address this challenge, CARE partnered with the Bill & Melinda Gates Foundation to develop and implement the IMAGINE project to examine how to support married adolescents and their families. Implemented between 2018 and 2021, the project aimed to help married adolescents in Niger and Bangladesh delay their first birth and envision, value, and pursue alternative life trajectories. IMAGINE's goal was two-fold:

1. to identify, design, and test interventions that hold promise for delaying the timing of first birth among married adolescents, and
2. to document and share learning from this initiative with the wider development community of others working to address the issue of adolescent childbearing.

The multifaceted IMAGINE program included components geared toward enabling married adolescents to delay first births and encouraging economic and social agency in their life courses. The intervention was designed to effect change in specific domains:

1. The health system structure
2. The alternative future opportunity structure
3. Relations and community social norms and values
4. Individual agency and control.

The following goals, questions, and hypotheses were evaluated separately in Niger and Bangladesh.

STUDY GOALS, HYPOTHESES, AND MEASURES

Goal 1: Delay timing of first birth

The primary goal of the CARE IMAGINE project was to delay first birth for married adolescent girls in Niger and Bangladesh. The timing of first birth was defined as a measure of the difference from date of marriage to the date of first birth among married adolescents. The following evaluation question addressed the impact of the intervention on the timing of first birth:

- Will the intervention delay the time to first birth (by a minimum of 6 months) for married adolescents in the treatment group compared to controls?

Goal 2: Improve family planning behaviors

Additional outcomes of interest included contraception use and unmet family planning need. Table 1 includes timing of first birth after marriage and all other reproductive health outcomes, with associated hypotheses and measures.

Table 1: Goals, hypotheses, and measures for timing of first birth after marriage and family planning behaviors

Goals	Hypotheses	Measures
Delay **time to first birth** for married adolescents exposed to the intervention.	On average, Married adolescents in the treatment group will have first births at least 6 months later than the control group.	• Difference in days between date of marriage and date of first birth.
Improve **family planning behaviors** for married adolescents exposed to the intervention.	Married adolescents in the treatment group will have higher contraceptive prevalence compared to married controls.	• Survey item indicating **current** use of any form of contraception.
	Married adolescents in the treatment group will have a higher modern contraceptive prevalence compared to married controls.	• Survey items asking about **current** method use for the following types of **modern** contraceptives: ○ Female Sterilization ○ Male Sterilization ○ IUD ○ Injectables ○ Implants ○ Pill ○ Condom ○ Female Condom ○ Emergency Contraception ○ Lactational Amenorrhea method ○ Other modern method
	Married adolescents in the treatment group will have a higher prevalence of contraceptive use across modern methods compared to married controls.	• Method-mix table composed of survey items asking about **current** method use for modern contraceptives as well as: ○ Standard days method ○ Rhythm method ○ Withdrawal ○ Other traditional methods
	Married adolescents in the treatment group will have higher contraceptive ever-use (i.e., lifetime) rates compared to married controls.	• Survey item indicating that the respondent has **ever used** some type of contraception.
	Married adolescents in the treatment group will have a lower prevalence of unmet family planning need compared to married controls.	• Unmet family planning need calculated with **three** conditions[1]: ○ Desire to delay birth by at least 24 months[2] OR ○ Desire to avoid pregnancy AND ○ Not currently using a contraceptive method.

[1] Unmet family planning need was calculated when a respondent indicated 1) a desire to avoid pregnancy indefinitely or for 24 months or more and 2) no current contraceptive use.

[2] Baseline unmet family planning was calculated using a desire to delay birth by 12 months. Endline reporting re-calculated the baseline measure of unmet family planning need using the recommended 24-month delay desire cutoff per recommendations from CARE based on more recent literature.

Goal 3: Improve reproductive health service use and knowledge, communal and personal efficacy, and economic outcomes

The intervention was also expected to impact reproductive health knowledge, service use and satisfaction, collective and personal efficacy, and social and economic mobility for married and unmarried girls in the treatment areas. All reproductive health knowledge and psychosocial outcomes with associated evaluation questions, hypotheses, and measures are listed in Table 2.

Table 2: Goals, hypotheses, and measures for reproductive health knowledge and psychosocial outcomes

Goals	Hypotheses	Measures
Improve **reproductive health service use and knowledge** for married adolescents exposed to the intervention.	Married adolescents in the treatment group will believe fewer family planning myths compared to married controls.	• Composite score[3] of ten survey items asking about family planning myths.
	Married adolescents in the treatment group will show significantly greater pregnancy risk knowledge compared to married controls.	• Composite score of four survey items asking about pregnancy risk.
	Married adolescents in the treatment group will show significantly greater satisfaction with SRHR services compared to married controls.	• Survey item asking about satisfaction with services received from the SRHR provider.
	Married adolescents in the treatment group who currently use family planning will report higher-quality interactions with service providers compared to married controls who currently use family planning.	• Composite score of thirteen survey items asking about the quality and nature of the interaction with the service provider.
	Married adolescents in the treatment group will report discussing family planning with CHW workers more frequently compared to married controls who were visited by a CHW.	• Survey item indicating family planning discussion with a CHW that has visited in the last 6 months.
	Married adolescents in the treatment group will report discussing family planning more frequently at health facilities they visited in the past 6 months compared to married controls who visited a health facility.	• Survey item indicating family planning discussion with personnel at a health facility visited within the last 6 months.
	Married adolescents in the treatment group will report longer durations of continuous usage of a current family planning method compared to married controls who currently use family planning.	• Survey item asking about the number of months of continuous usage for the most recent family planning method.
	No known hypothesis and is expected to be exploratory and descriptive.	• Survey item asking about reason for discontinuation of usage for the most recent family planning method.
	Married adolescents in the treatment group who currently use family planning will show significantly greater willingness to recommend the service to a friend compared to married controls who currently use family planning.	• Survey item asking about willingness to recommend service to a friend.
	Married adolescents in the treatment group will report their husbands being involved in family planning use and decision making more often compared to married controls.	• Survey item asking about husband's involvement in the use and decision making for family planning.
	Married adolescents in the treatment group will report significantly higher rights-based family planning scores compared to married controls.	• Composite score of six survey items asking about rights-based family planning.

[3] See Appendix E for a list of all composite measures, associate individual survey items, and coding strategies.

Goals	Hypotheses	Measures
Improve **communal and personal efficacy** for married adolescents exposed to the intervention.	Married adolescents in the treatment group will report higher self-efficacy to discuss and use family planning compared to married controls.	• Composite score of four survey items asking about self-efficacy to discuss and use family planning.
	Married adolescents in the treatment group will report higher self-efficacy to refuse sex compared to married controls.	• Composite score of five survey items asking about self-efficacy to refuse sex.
	Married adolescents in the treatment group will report higher self-efficacy to go to a health facility compared to married controls.	• Composite score of five survey items asking about self-efficacy to go to a health facility.
	Married adolescents in the treatment group will report higher self-efficacy to engage in income generating activities compared to married controls.	• Composite score of six survey items asking about self-efficacy to go to engage in income generating activities.
	Married adolescents in the treatment group will have higher mobility scores compared to married controls.	• Composite score of nine survey items asking about mobility of respondent.
	Married adolescents in the treatment group will report higher social cohesion scores compared to married controls.	• Composite score of eight survey items asking about social cohesion of respondent.
	Married adolescents in the treatment group will report higher collective efficacy scores compared to married controls.	• Composite score of five survey items asking about perception of collective efficacy.
	Married adolescents in the treatment group will report greater involvement in household decision making compared to married controls.	• Composite score of five survey items asking about who participates in general household decisions.
	Married adolescents in the treatment group will report significantly greater interspousal communication compared to married controls.	• Composite score of six survey items asking about the frequency of interspousal communication.
Improve **economic outcomes** for married adolescents exposed to the intervention.	Married adolescents in the treatment group will report greater involvement in financial household decision making compared to married controls.	• Composite score of eight survey items asking about who participates in household decisions involving finances.
	More married adolescents in the treatment group will report income generating activities in the past 12-month period compared to married controls.	• Survey item asking about participation in work during past 12-months.
	Married adolescents in the treatment group will report more monthly income compared to married controls.	• Survey item asking the participant to report their average monthly income.
	Married adolescents in the treatment group will report larger savings amounts compared to married controls.	• Survey item asking the participant to report their current savings amount.
	More married adolescents in the treatment group will report having a savings accounts compared to married controls.	• Binary indicator of participant having saving amount greater than 0.
	More married adolescents in the treatment group will report participation in VSLA groups compared to married controls.	• Survey item asking about participation in a VSLA/saving group.
	Married adolescents in the treatment group will indicate participating in more vocation training activities compared to married controls.	• Survey item asking about participation in vocation training activities.
	Married adolescents in the treatment group will have a higher rate of capital asset ownership compared to married controls.	• Survey item asking about owning capital assets.

Goal 4: Improve family planning perceptions

The intervention was also expected to affect family planning perceptions and beliefs, including ideal timing and number of children, and familial and gender roles. All goals, hypotheses, and measures to assess changes in family planning perceptions are listed in Table 3.

Table 3: Goals, hypotheses, and measures for family planning perceptions

Goals	Hypotheses	Measures
Improve outcomes in the domain of **ideal family norms** for married adolescents exposed to the intervention.	Married adolescents in the treatment group will report an older ideal age of first birth compared to married controls.	• Survey item asking about ideal age for first birth.
	Married adolescents in the treatment group will report a longer ideal time between marriage and first birth compared to married controls.	• Survey item asking about ideal time after marriage for first birth.
	Married adolescents in the treatment group will prefer a smaller family size compared to married controls.	• Survey item asking about the preferred number of children.
	Married adolescents in the treatment group will report their husband prefers a smaller family size compared to married controls.	• Survey item asking about the respondent's husband's preferred number of children.
Improve **perceived expectations** for married adolescents exposed to the intervention.	Married adolescents in the treatment group will report significantly higher scores for normative expectations about girls' roles compared to married controls.	• Composite score of ten survey items asking about normative expectations about girls' roles.
	Married adolescents in the treatment group will report significantly higher scores for expectations of family planning use compared to married controls.	• Composite score of two survey items asking about expectations regarding family planning.
	Married adolescents in the treatment group will report significantly higher scores for expectations for delaying first childbirth after marriage compared to married controls.	• Composite score of four survey items asking about expectations for delaying childbirth.
	Married adolescents in the treatment group will report significantly higher scores for expectations for engaging in income generating activities compared to married controls.	Composite score of three survey items asking about expectations about engaging in income generating activities.

Intervention Program

To achieve program objectives, the program designers worked from an explicit Theory of Change based on prior intervention work, early field research, and focus groups in similar areas of each country. Table 4 summarizes the program components and intervention strategies as related to the intermediate (secondary) and primary outcomes.

Table 4. IMAGINE Theory of Change

Problem Statement	Adolescents lack the skills, capacity, and support (from their families, communities, and health system) they need to be able to delay first birth and pursue alterative futures to early motherhood.
Because of….	
Analysis of Key Dynamics	**Health Workers** lack the skills and capacity to tailor services to adolescent's specific needs, and often hold values, beliefs, and norms that act as barriers to the equitable provision of FP services to adolescents. **Adolescent girls** have a limited awareness of the alternatives to early motherhood available to them, lack access to financial capital and control over resources, and are often neglected by existing vocational training opportunities and positive economic secular trends. **Newlyweds** face significant pressure from their families and **communities** to have a child soon after marriage, with the **husband** as the primary decision maker in unions often characterized by a lack of gender equity. **Adolescent girls** lack the capacity to **envision** and **pursue** alternative life trajectories other than early motherhood. **Adolescent girls** lack the knowledge, skills, capacity, and links to formal health sector needed to make the healthy timing of pregnancy a reality.
However, if we do…	
Interventions / Strategies	Engage in **reflective dialogue** practice and counseling skills-building activities with health workers; Offer transformative **vocational opportunities** in IT, mobile technology, and handicraft sectors; Provide in-home **couples counseling services** to newlywed couples and mothers-in-laws (Bangladesh only); Engage the wider public in visible, positive **events**; And deliver a comprehensive curriculum to **girls' collective solidarity groups**.
Then we expect that…	
Intermediate Outcomes	Health care workers will adopt supportive behaviors toward married adolescents who wish to delay first birth; Improved engagement in alternative opportunities among married adolescents; Increased support to delay first birth among young men / husbands; Increased support to delay first birth among mothers-in-laws; Increased support to delay first birth among young men / husbands and mothers-in-laws; Married adolescent girls will be able to envision and perceive value in alternatives to early first birth; Married adolescent girls will have enhanced agency and assets relevant to delaying first birth and pursuing alternative futures; Increased use of and satisfaction with sexual and reproductive health services among married adolescents.
And, as a result…	
Primary Outcomes	The timing of first birth will be delayed by 6 months or more above the average among married, 15-19 year-olds in intervention areas.

EVALUATION DESIGN

Key messages:

- The primary aim of the intervention was to extend the time between marriage and first birth by six months or more.
- The program targeted adolescent girls aged 15-19 year in Bangladesh and Niger. Both countries had a treatment and control group.
- Comprehensive information was collected to monitor girls' trajectories over the 2-year intervention which began in January 2019 and ended in May 2021.
- The baseline measurement occurred during November/December 2018 and the endline survey was completed in December 2021/January 2022

The central question for this evaluation was whether exposure to the intervention would extend the time between marriage and first birth by 6 months or more (compared to the control group). To test this hypothesis, the impact of the intervention was assessed using a longitudinal before-and-after design where the same respondents were surveyed at baseline and endline. A sample comprised of both married and unmarried nulliparous 15- to 19-year-old adolescent girls was obtained from both treatment and comparison communities. Treatment areas were purposively selected; control areas were then selected and matched by subject matter experts to major regions in each country based on available geographic and demographic characteristics. Treatment and control samples were matched on baseline covariates (e.g., religion, ethnicity, literacy, etc.) to ensure comparability.

METHODS

Sample and Data Collection

The study population came from two geographic regions within Niger and Bangladesh. Participant households within Bangladesh came from the far northeast region of the Kurigram Sadar Upazila, situated in Kurigram District. Bangladesh treatment unions included Belgaccha and Punchgachhi and control unions include Bhogdanga and Kanthalbari. Participant households within Niger come from the south-central Mirriah Department, situated in the Zinder Region. Treatment communes in Niger include Dogo and Kolleram and the control areas include Gaffati, Gouna, Hamdar, and Zermou.

Figure 1 shows the geographic distribution of surveyed households in Kurigram Sadar. Households with respondents in the treatment arm are plotted in green against a light-yellow background; those assigned to the control arm are plotted in blue against a light blue background.

Figure 1. Kurigram Sadar Upazila, Bangladesh – geographic distribution of survey households

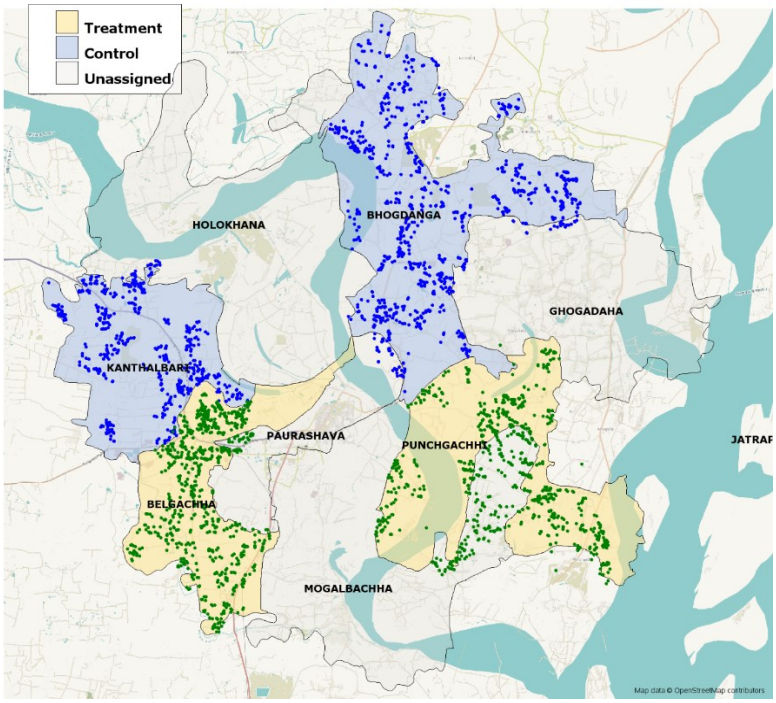
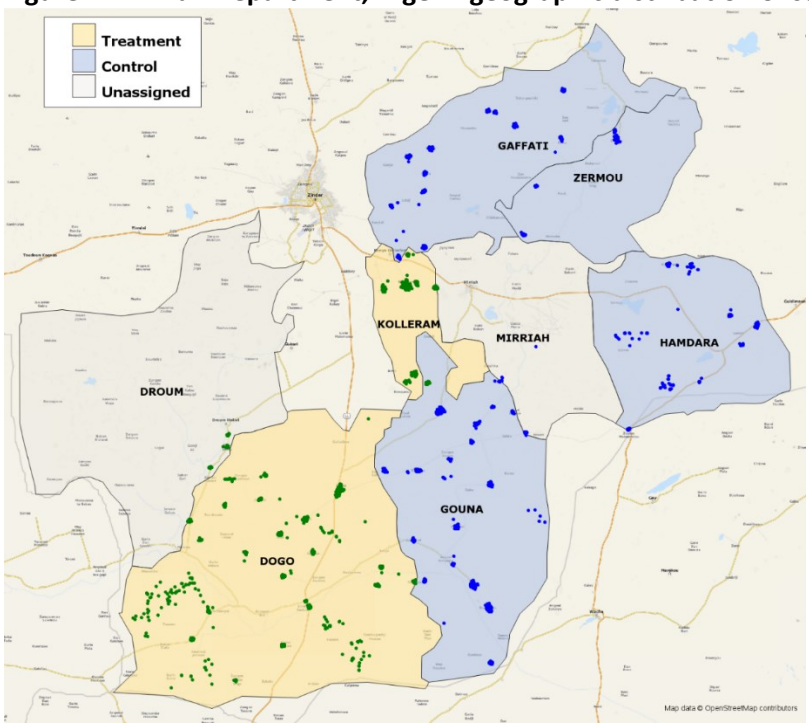


Figure 2 shows the geographic distribution of surveyed households in Mirriah Department. Households with respondents in the treatment arm are plotted in green within a light-yellow background; those assigned to the control arm are plotted in blue within a light blue background.

Figure 2. Mirriah Department, Niger – geographic distribution of surveyed households



The IMAGINE intervention occurred over a two-year period from January 2019 to May 2021, with eligible intervention participants contributing responses to surveys at baseline in November to December of 2018. Participants were then followed up with at endline in December 2021, and January 2022. Data collection began with an initial enumeration of all treatment and control villages estimated to have at least 19 eligible adolescent girls. In Bangladesh, this included 35 villages in the treatment areas and 58 in the control areas; in Niger, it included 41 villages in treatment areas and 45 in control areas. At baseline, all eligible girls were surveyed by in-person field teams.

At baseline, three questions were used to assess eligibility:

1. How old were you at your last birthday? IF AGE IS BETWEEN 15-19, CONTINUE WITH SURVEY
2. Have you ever been pregnant? IF 'NO' OR 'UNSURE', CONTINUE WITH SURVEY
3. Are you currently pregnant? IF 'NO' OR 'UNSURE', CONTINUE WITH SURVEY

If age at last birthday (question 1) was unknown, interviewers probed the respondent by asking what year and season that the respondent was born. If at least 15 years since the earliest point in that season-year and no more than 20 years since the latest point in that season-year, the participant was considered eligible for participation. Since data collection occurred in the autumn of 2018, girls born as early as the autumn of 1998 or as late as the autumn of 2003 were eligible.

Every village was assigned a target sample size proportionate to the estimated eligible population from publicly available records. To ensure random selection of sample-eligible participants from within villages, the survey team conducted "on-the-ground" village-level enumeration, to identify all households containing at least one adolescent girl aged 15 to 19 years old. During this enumeration phase, the survey team generated a sampling frame of all study-eligible girls found within all households. The sampling frame also contained the following information for each household:

- The GPS coordinates
- The name of the head of the household
- A listing of all adolescents in the household ages 15-19 years old (identified by first name) that were eligible for the primary or the secondary (childbearing) samples, disaggregated by eligibility status (primary vs. secondary sample).

At the end of the enumeration phase, the survey team generated a sampling frame of randomly selected households for every village. The survey team then returned to households containing eligible subjects according to the randomized frame order until achieving the target sample size for each village. If a village was found to have fewer than the target eligible size, teams attempted to collect data on every eligible girl within the village.

Table 5. Summary of study design and methods used in Niger and Bangladesh.

Study Design	Study Population	Sampling Strategy	Country	Regions
Longitudinal before-and-after study	*Primary*: Adolescent girls (aged 15-19) who had never been pregnant *Secondary*: Childbearing adolescent girls (aged 15-19)	Two-stage design purposive selection of regions. Random selection among all eligible adolescents from all villages.	Bangladesh	*Treatment*: Belgachha and Punchgachhi *Control*: Bhogdanga and Kanthalbari
			Niger	*Treatment*: Dogo and Kolleram *Control*: Gaffati, Gouna, Hamdara, and Zermou

Baseline and Endline Sampling

The baseline data for the primary and secondary surveys were collected from the field during the months of November and December 2018. Starting sample sizes were computed under the assumption of a simple random sample (SRS). Samples were then inflated to account for women likely to be ineligible at endline because of a) married women unaware that they were pregnant at baseline (i.e., within their first two trimesters) or b) unmarried women remaining unmarried at endline. After estimating the number needed for power of .90 for the primary delay-of-birth outcome under the assumption of an estimated sample (SRS), then adjusted for ineligibility determined at baseline, the sample size was then inflated by another 25% to conservatively

accommodate for an estimated 20% attrition that could result from outmigration or for enumerators who were unable to locate baseline participants at follow-up. Finally, since the sample is not actually SRS, the sample size was multiplied by two (2.0) to adjust for design characteristics such as clusters (i.e., villages) and weights. These design factors tend to increase variance estimates in statistical models, requiring larger sample sizes than SRS (Hsieh et al., 2003).

Bangladesh

The planned baseline primary sample size was 3,250 adolescent girls total in 93 villages across four regions (two treatment and two comparison). During enumeration prior to the baseline survey, the data collectors learned that three villages from the original frame (all in the control areas) were destroyed due to flooding, resulting in lower-than-expected sampling. In addition, enumeration of households in many villages was discovered to be smaller than the published census estimates. Since the analysis was powered based on sample sizes for each village, the sampling rates increased. Also, due to smaller overall enumeration and unexpected loss of three flooded villages, the final sample size was approximately 19% less than expected. Field teams attempted to visit 2,892 girls from the frame across 90 villages (55 control and 35 treatment) to reach the remaining village target sample sizes. A total of 2,629 girls completed the survey and were the sampling frame for endline data collection. Of 2,508 households included in the baseline survey sample, 101 had either two or three respondents, composing 7.8% of the respondent sample.

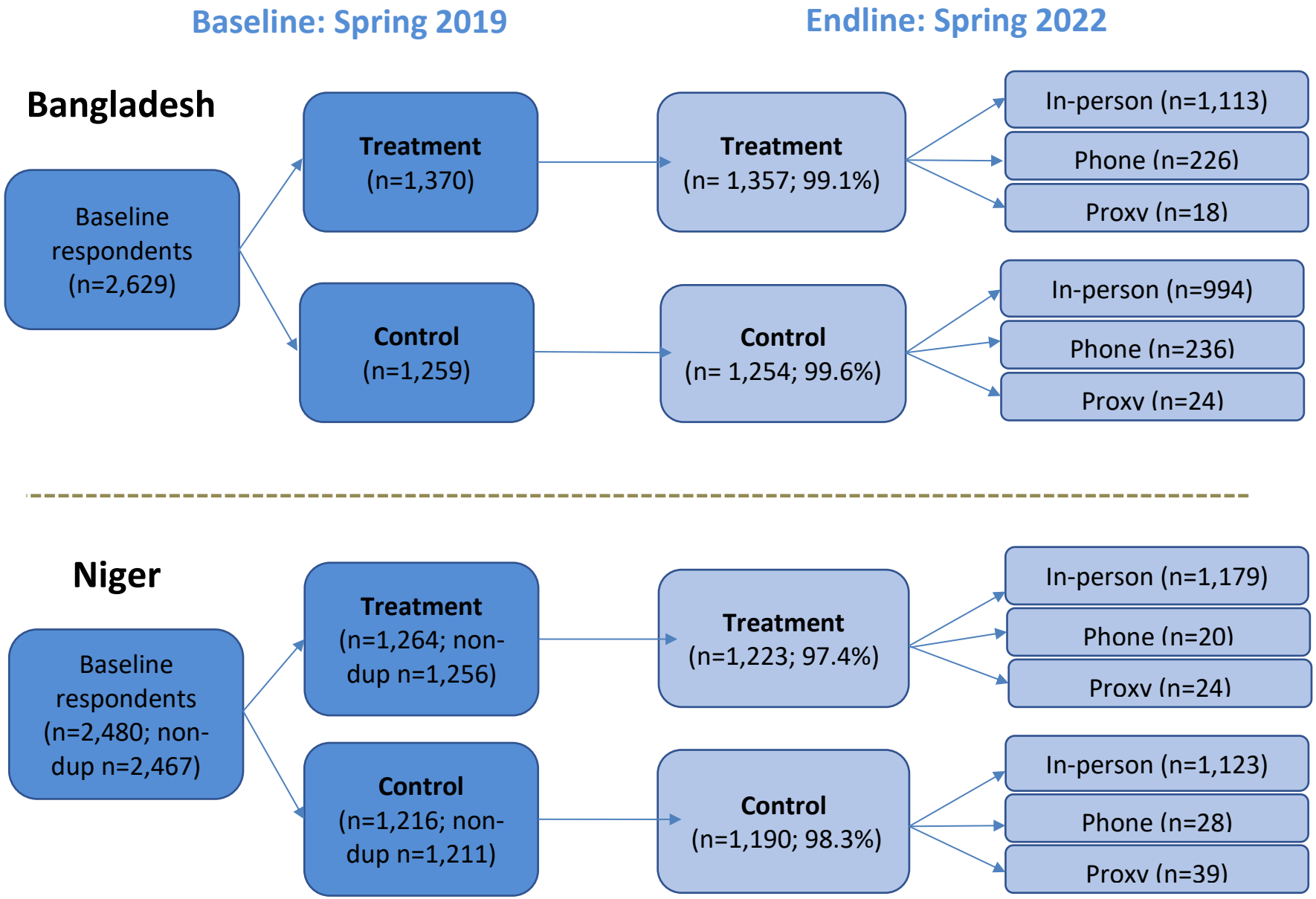
Niger

The initial target sample size was 2,750 in 86 villages across six regions. As in Bangladesh, due to challenging field conditions, difficult enumeration, and higher than estimated refusal rates, the sample was about 10% less than the target. Field teams attempted to visit 3,164 girls from the frame across 86 villages (45 control and 41 treatment) to obtain the target sample size. In the end, 2,480 survey responses were collected in 85 villages and were the sampling frame for endline data collection (in data provided by CARE, one control village had no corresponding survey responses). Thirteen girls in the Niger sample completed the baseline survey twice, and thus the actual baseline sample for Niger was 2,467. Of 2,249 households included in the baseline survey sample, 149 had two, three, or four respondents, composing 13.0% of the respondent sample.

Given these challenges in sampling at baseline, the field teams were strongly encouraged to focus on retention to maintain power for the endline analysis. This effort was successful, yielding more than sufficient power for the endline analysis (see next section on ‘Attrition and Missing Data’ for more details).

Figure 3 shows baseline to endline sample size by country and endline response modality.

Figure 3: Timeline and longitudinal respondents by country



Endline Data Collection

At endline, in-field interviewers attempted to contact baseline respondents in person or by phone using the contact information collected during enumeration and baseline. If the respondent could not be reached in person or by phone, a proxy respondent such as a parent could answer questions on the adolescent's marital status, approximate date of marriage, whether she had given birth to her first child, and the approximate date of birth of her first child.

The phone and proxy surveys only included essential variables for calculating the duration between first birth and marriage and estimating the primary family planning behaviors: current contraceptive use, current use of a modern contraceptive, method-mix rate, ever-use of family planning methods, and unmet family planning need.

Data Preparation

Sample Exclusions

Based on consultation with the CARE team, respondents who gave birth within 266 days (a full-term pregnancy) of the country-specific *start of intervention* were excluded from all endline analyses. The rationale for this was that the intervention could not have affected time of first pregnancy for these respondents, because they became pregnant before the intervention could have made an impact. These participants (n = 114 [4.4% of endline respondents] in Bangladesh and n = 290 [12.0%] in Niger) were considered "left-censored" and excluded from all analyses.

Attrition and Missing Data

Given minimal attrition (0.7% in Bangladesh and 2.2% in Niger), the primary sources of missing data were restricted question sets in the phone and proxy surveys and sporadic item nonresponse.

As noted above, phone and proxy surveys were shortened to prioritize collecting the necessary variables for the primary outcomes. These alternate collection modes minimized overall attrition but resulted in a pattern of item-level non-response, especially for items for which a proxy respondent could not reasonably be expected to have knowledge of the topic.

Methods and conclusions for investigating risk of mode bias are detailed in Appendix A. In summary, there was no affirmative reason to reject the assumption that data that were missing based on response mode were Missing Completely At Random (MCAR: the likelihood of a datapoint being missing is unrelated to its true value; Little & Rubin, 1987).

Taking together the low rate of attrition, the lack of evidence for non-ignorable missingness based on endline data collection mode, and the low rate of item missingness for the key indicators of time to first birth (which were collected across all modes), endline data were assumed to be MCAR. Descriptive statistics were based on listwise deletion and inferential analyses on either listwise deletion or methods that use estimation directly from respondent-

level data (e.g., "full information" maximum likelihood [FIML] or scaled weighted least squares [WLSMV] estimators).

Sample Weighting

Baseline weights accounted for differential response rates across villages and were constructed and scaled to represent the full primary populations (Seaman & White, 2013; Mansournia & Altman, 2016). Baseline village weights summed to the primary frame counts. Due to frame demographics being limited to age and marital status, baseline weights could not account for other dimensions of individual-level demographic variations in response propensity. Even so, as previously reported, treatment and comparison groups were well balanced at baseline on demographic variables. Applying the baseline weights, no key demographics differed significantly between condition in Bangladesh, and in Niger there was a statistically significant difference only for ethnicity (87.4% of participants in the treatment group were of the Hausa ethnicity, versus 95.7% of control group participants).

Analytic weights that reflected both the sampling process and timing of marriage for use in marginal structural models (MSMs). MSMs with Inverse Probability of [Marriage] Weights (IPW) allowed appropriate adjustment for a time-varying exposure (i.e., marriage) that could be confounded with intervention assignment or other baseline variables in the causal chain to the outcomes of interest (Robins, Hernán, & Brumback, 2000). Appendix B provides a detailed description of the endline weighting methodology and results.

The resulting analytic weights were designed to balance the sample at endline on both baseline characteristics and time of marriage.[4] Each respondent's weight reflected a composite of their likelihood of being in the baseline sample (relative to population characteristics) *and* their likelihood of being married during a given period in the study. Analytic weights were scaled to a unit mean to reflect statistical power corresponding to the respondent sample size.

Data Consistency and Imputation

Consistency was imposed *post hoc* on indicators of marital status for analyses reported herein. Respondents who reported being married at baseline were assigned a value of "married" in the endline variables regardless of their actual response. This preserved the data of all married respondents even in cases of widowing or other life events. Further, participants who reported being married at endline and gave a date of marriage before their baseline interview were assigned a value of "married" in the baseline variables (i.e., relying on the date rather than the binary item).

[4] Though possible with the MSM-IPW approach (Robins *et al.* 2000), loss to follow-up was not incorporated in the calculation of analytic weights because of the likelihood of extreme weights and minimal benefit of the added complexity.

Missing responses to date of marriage would have resulted in decreases in the sample size for the models of time to first birth, potentially in a biased fashion (for example, where a date was missing due to the respondent being widowed during the study). Dates of marriage were algorithmically imputed in cases in which they were not reported. The imputation rules and SAS code applied (including the random number seed) are presented in Appendix C. Dates of marriage, raw or imputed, were not used in any other analyses.

Analytic and Modeling Strategy

The longitudinal design of the study had several characteristics that made it ideal for evaluating the primary research question. First, the entire marriage to birth process could be observed in both treatment and control samples. This measurement was especially important given that the intervention could delay marriage timing among the treated unmarried, which might then confound the impact on duration of marriage to first birth if not properly modeled. The design thus enabled evaluation of change over and above any change in age at marriage. Second, individual participants could be matched on covariates measured at a common baseline timepoint to ensure comparability (such as religion, ethnicity, age, or education). Third, the longitudinal study enabled a richer modeling of individual change (in beliefs, norms, knowledge, etc.) rather than depending on the simple group mean comparisons inherent in a repeated cross-sectional study. Finally, for secondary outcome measures that were not modeled as time-to-event (such as contraception use), the longitudinal design allowed for difference-in-difference estimators in the evaluation of the program's impact. This allows for direct calculation of how the change in these measures from baseline to endline differs between the treatment and control groups.

Outcome Modeling

For purposes of this section, the analysis strategy is described in general terms according to the measurement type of the outcome. The *Results* section presents analyses ordered by the CARE team's priorities.

Clustered Design

All models, except where noted, accounted for the complex design of respondents clustered within villages in a random-intercepts model. Given that the large majority of households (96.0% in Bangladesh and 93.4% in Niger) included only one interview participant, a third level of nesting of individual within household would have been highly resource intensive relative to likely benefits and was not modeled.

Marginal Structural Models

Except for the primary outcome of delay-of-birth, analyses applied marginal structural models (MSM). These models rely on the sequential weighting process used for calculating the analytic weights. MSMs have substantial added complexity in the calculation of parametric standard errors (Robins et al., 2000). Empirical standard errors and confidence intervals presented in this report were derived using non-parametric resampling techniques (the jack-knife, percentile

bootstrap, or bias-corrected bootstrap, as appropriate to the specific analysis), with corresponding statistical inference methods.

Combined with the weighting strategy, these non-parametric confidence intervals were intended to result in optimal estimates and tests. Resampling-based confidence intervals have a second advantage in that the process can be applied to subsample ("domain") analysis (e.g., effects among only married women) without any procedural change to the weighting or calculation of degrees of freedom (West, Berglund, & Heeringa, 2008; West, Sakshaug, & Aurelien, 2018).

Covariate Set

All analyses, except where otherwise noted, used a standard set of covariates selected *a priori* from the baseline survey (i.e., variable value at baseline, when change over time is possible). These included:

- age,
- education,
- marital status,
- having a healthcare visit in 6 months prior to baseline,
- knowing of a place to access family planning,
- having income generation activity in 6 months prior to baseline,
- having a personal savings,
- household having capital assets,
- religion [in Bangladesh], and
- ethnicity [in Niger],

Exclusions

Numerous outcome indicators, including time to first birth, were based on items asked only of respondents who indicated at endline that they were married. These items were considered not applicable to unmarried respondents, and those respondents were omitted from analyses predicting these outcomes. Certain others were asked only of respondents who reported using contraception at endline. The relevant sample exclusions are noted with the presentation of each outcome in the *Results* section.

Analysis of Time to First Birth

The analytic approach for the primary outcome—a delay in time to first birth--was a time-dependent sequential Cox proportional-hazard survival model (versus an MSM hazard model), estimating the effect of intervention on the "hazard" of giving birth in any time increment after marriage (Gran et al., 2010). These models incorporated the common covariate set (except for baseline marital status), weights, and clustering, with standard errors calculated from bootstrap weights with 500 replicates (Rao, Wu, & Yue, 1992). The time windows established in the weighting (see Appendix B) served as strata. Following Gran *et al.* (2010), weights were computed separately for each window. Within each time window / stratum, the inverse

probability weights of marriage within the window (among respondents not yet married at its beginning) were estimated from logistic regressions with the same predictor sets used in the creation of analytic weights. The resulting weight within stratum was the product for each participant of the stratum-specific marriage IPW and the original baseline weight. The composite likelihood estimates from a stratified weighted Cox regression are the parameter estimates for the aggregated causal effect of treatment on the hazard of first birth over all time windows (Gran et al., 2010).

Certain additional data considerations were relevant to this analysis. Every effort was made during endline data collection to record the exact dates of both marriage and first birth, providing a continuous measure for time to event. In practice, month and year were available for nearly all relevant dates. Exact day of birth for the first-born child was known for all births in Bangladesh. In Niger, exact day of birth was missing or unknown for 611 (58.9%) respondents who reported having given birth by the endline interview. The day of the month was set as the 15th when it was not reported or indicated as unknown. This was a trade-off between precision and accuracy, as it allows the use of actual days for a continuous-time model. As random variation in actual dates of marriage or birth is expected to center around the middle of the month, this strategy was expected to minimize bias.

For inclusion in the time to first birth survival analysis, study participants were required to have been married at least 266 days (the duration of a full-term pregnancy) before the endline survey. Participants who reported a date of first birth less than 154 days after date of marriage were also excluded from the time to first birth analysis, as a live birth was unlikely to have occurred in this timeframe. In such cases, or when a respondent who reported a live birth before marriage or indicated she was unmarried at both baseline and endline, the likely scenarios are either a reporting/recording error or that the child was conceived outside of marriage and thus the respondent was not in the "at-risk" set. Forty (1.9%) respondents in Niger were excluded from the time-to-birth analysis due to a pregnancy of less than 154 days; 1 (<0.1%) respondent was excluded for this reason in Bangladesh.

Regression Models for Additional Outcomes

Program effects on categorical outcomes were evaluated using multiple logistic regressions, adjusted for clustered sampling and analytic weights, the common covariate set, and, where applicable, the baseline measure of the variable[5]. The cumulative logit model with the proportional odds assumption was applied to ordered categorical outcomes, and the multinomial logit model to nominal outcomes. Several outcomes involving numeric responses that were not amenable to normal-theory methods (e.g., ideal family size) were binned into *ad*

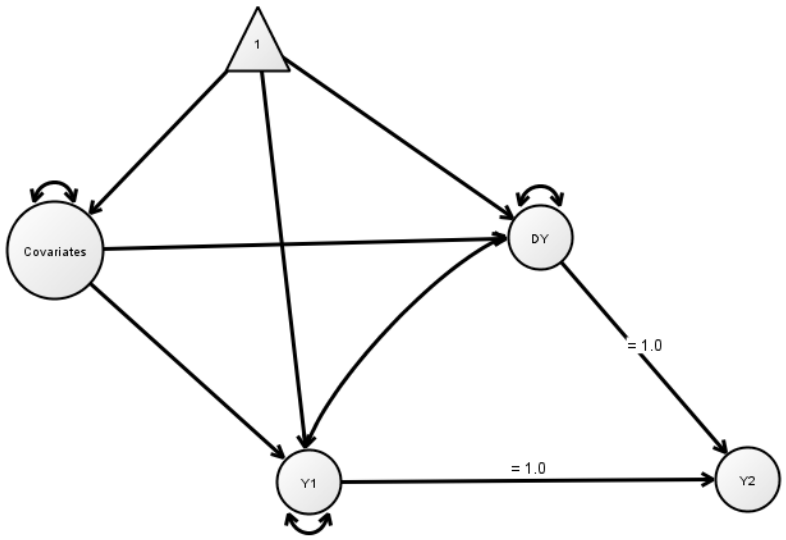
[5] The baseline measure of the endline outcome was excluded from the analysis model only when its inclusion would substantially lower the analysis sample due to missing predictors. In most cases, these outcomes were those where the baseline measurement was conditional on being married or having used contraception.

hoc categories according to the actual response frequencies and analyzed as ordinal. In all cases, jackknife standard errors were used for inferential testing.

Two-group latent change score models (LCS; McArdle, 2009 were estimated for continuous outcomes that had both baseline and endline measurements (e.g., social cohesion), accounting for the analytic weights and the clustered sampling by village. The LCS model has several advantages over conventional difference-in-difference modeling, particularly in how it can draw on the flexibility of a structural equation modeling (SEM) framework.

The LCS model uses certain constraints on the baseline and endline values of the indicator to directly estimate a parameter for the difference. A sample LCS model is shown in Figure 4, where *Y1* is the baseline value of variable *Y* and *Y2* is the endline value of the same variable, and covariates are summarized in a single circle for ease of presentation. *Y1* and the "phantom" latent variable *DY* are regressed on the baseline covariates. The residual (disturbance) variance and intercept of *Y2* are fixed to 0 *a priori* (indicated by the lack of a curved arrow and the lack of a path from the triangle, respectively). This, combined with the regression coefficients from *Y1* and *DY* being fixed at 1, identifies the latent difference *DY.* With these constraints, *Y2* becomes the exact sum of *Y1* and the phantom variable *DY*. Because *Y2* is observed, *DY* then becomes *Y2* minus *Y1*. Both *Y1* and *DY* are regressed on the standard covariate set.

Figure 4. Example Latent Change Score (LCS) Model



Legend:

- *A straight arrow represents a regression path; the paths with a value of 1.0 are fixed; others are freely estimated.*
- *A path from the triangle indicates the mean or intercept of the variable is freely estimated.*
- *A curved arrow linking a variable to itself indicates a freely estimated variance or residual variance. The curved arrow between **Y1** and **DY** represents a free residual covariance (the variables are allowed to covary beyond having common causes).*

A one-group LCS model is conceptually equivalent to a paired *t*-test (Coman et al., 2013), but the SEM framework allows estimation of not only the *mean* change within treatment condition, but also its *variance* (which may differ between conditions). The difference between the mean of *DY* in the intervention group and the mean of *DY* in the control group, when estimating the model with clustering and analytic weights, is the estimated effect of the intervention. Difference tests used bias-corrected bootstrapped confidence intervals.

A traditional ANCOVA-style framework was applied to interval outcomes that were measured for most respondents only at endline, though following the SEM strategy of bootstrapped confidence intervals for consistency. In these models, *Y2* was regressed on the standard covariate set and an indicator for intervention condition. The regression coefficient for intervention was the estimate of interest[6].

Dose-Response Analysis

Dose-response analyses linking dose received of the IMAGINE intervention (among treatment participants) to time-to-birth and the primary family planning outcomes serve to supplement the intent-to-treat analysis of program effectiveness. Dose measurement was based on participant self-report in the endline survey. In each country, respondents were asked whether they had participated in IMAGINE "Girls' Groups" during implementation, and if so, how many. The predictor in these models was a 5-point indicator with response options ranging from "none" to "all."[7] These analyses otherwise followed the same analysis strategy and statistical methods as for program effectiveness.

Type I Error Control

IMAGINE was predicted to affect numerous outcomes beyond the primary outcome of delay of first birth after marriage. As appropriate, the Benjamini-Hochberg adjustment was applied to *a priori*-defined families of outcomes, which are noted in the *Results* section, to control the False Discovery Rate (FDR) within family to a nominal α of .05 (Benjamini & Hochberg, 1995).

[6] The initial plan to adjust measures in the LCS models for estimated scale reliability was found to be computationally intractable.

[7] Additional measures of dose focused on the participation of respondents' husbands in IMAGINE programming, either husband-only events in Niger or couples' events in Bangladesh. However, interpretation would have been complicated by the fact that availability of these programs to the respondents was confounded with marital status and the timing of marriage. Given that participation in the husband-only and couples' programs was broadly low, the dose measure was restricted to Girls' Groups, which were available to all participants.

RESULTS

Endline Sample

Attrition and Survey Mode

Attrition rates were low, with 2,611 (93.3%) of the original 2,629 participants in Bangladesh and 2,413 (97.8%) of 2,467 in Niger having endline data.[8] The endline interview was administered face-to-face for 80.7% (*n* = 2,107) of the retained sample in Bangladesh, with 17.7% (*n* = 462) surveyed by phone and 1.6% (*n* = 42) asked of a knowledgeable proxy. In Niger, 95.4% (*n* = 2,302) of endline surveys were administered face-to-face, 2.0% (*n* = 48) by phone, and 2.6% (*n* = 63) via proxy.

Descriptive Statistics

Given the high retention rates, the endline samples remained demographically similar to the baseline samples. All weighted descriptive statistics are presented using the baseline weights.

Tables 6 and 7 show weighted endline demographic statistics for Bangladesh and Niger after removing cases of apparent pregnancy before the start of intervention. In each country, baseline differences in demographics were assessed using *t*-tests for continuous variables and χ^2 tests of independence for discrete variables, adjusting for baseline weights and clustering by village[9]. The FDR was held to .05 across demographics within each country. In Bangladesh (Table 6), a lower proportion of treatment participants were married at baseline relative to control.[10] The treatment group also had a higher proportion of respondents practicing Hinduism vs. Islam.

[8] These comparison sample sizes are of the full baseline sample before left-censoring, as censoring due to early pregnancy could not be assessed for respondents lost to follow-up.

[9] Several authors (Altman, 1985; Mutz, Pemantle and Pham, 2019) have argued against use of statistical tests of balance for randomized trials. Most of these criticisms focus on use of tests when randomization was correctly performed. We note here that due to ethical and practical issues, true randomization could not occur for this study. The statistical tests of demographics are used here only as a convenient heuristic for considering any endline differences arising after exclusions and attrition, for which there was purposive selection of regions at baseline. The results of these tests are not used to adjust covariates in the subsequent models, another key criticism in the employment of balance tests. A measured and brief perspective on this practice may be found here: https://blogs.worldbank.org/impactevaluations/should-we-require-balance-t-tests-baseline-observables-randomized-experiments.

[10] This difference was not evident at baseline. Applying the same comparison to baseline marriage rates using the endline sample, 27.5% of treatment-group respondents were married at baseline versus 30.1% of control-group respondents, unadjusted *p* = .360.

Table 6. Bangladesh sample descriptive statistics by condition

	Treatment			Control		
Variable	*n*	$\bar{x}$ *or %*	*SD*	*n*	$\bar{x}$ *or %*	*SD*
Endline response mode	1289			1208		
Face-to-face		81.6%			79.5%	
Proxy		1.4%			2.0%	
Phone		17.0%			18.6%	
Age at endline	1289	19.3	1.5	1208	19.4	1.4
Married at baseline	1289	27.5%		1208	31.1%	
Married at endline[1]*	1283	65.6%		1201	72.2%	
Age at marriage	844	17.0	2.0	859	17.1	1.9
Highest education	1053			960		
None/primary		4.9%			5.5%	
Secondary		43.4%			43.8%	
Higher		51.7%			50.6%	
Bengali ethnicity	1289	100.0%		1208	100.0%	
Religious affiliation[2]*	1053			960		
Muslim		91.4%			94.5%	
Hindu		8.6%			5.5%	

Note: Demographic statistics are shown after removing cases of apparent pregnancy before the start of intervention (left-censored cases) and are calculated using baseline sample weights; n = sample size; $\bar{x}$ = *mean; SD = standard deviation for continuous measures; *significantly different from control at p<.05 after controlling for multiple comparisons; All variables included in tables were considered one family of comparisons;*

1: Sample size for marital status differs from total due to 13 missing/unknown values
2: Endline education and religious affiliation variables not available for phone and proxy respondents

In Niger (Table 7), the only statistically significant difference between treatment and control after FDR adjustment was for ethnicity (treated as Hausa versus other), with a lower proportion of Hausa respondents in the treatment group.

Table 7. Niger weighted primary sample demographics – by treatment & control

	Treatment			Control		
Variables	*N*	$\bar{x}$ *or %*	*SD*	*N*	$\bar{x}$ *or %*	*SD*
Endline Response Mode	1089			1034		
Face-to-face		96.1%			94.5%	
Proxy		2.0%			3.1%	
Phone		2.0%			2.4%	
Age at endline	1089	18.5	1.7	1034	18.5	1.8
Married at baseline	1089	26.8%		1034	27.2%	
Married at endline	1089	64.2%		1034	63.5%	
Age at marriage	706	16.6	1.9	693	16.7	2.0
Highest education[1]	1049			978		
No Schooling		31.4%			28.2%	
Primary		24.5%			27.1%	
Secondary/Higher		44.1%			44.7%	
Ethnicity*	1089			1034		
Hausa		89.6%			95.3%	
Other ethnicity		10.4%			4.7%	

Note: Demographic statistics are shown after removing cases of apparent pregnancy before the start of intervention (left-censored cases) and are calculated using baseline sample weights; n = sample size; $\bar{x}$ = *mean; SD = standard deviation for continuous measures; *significantly different from control at p<.05 after controlling for multiple comparisons; All variables included in tables were considered one family of comparisons;*

1: Endline education not available for phone and proxy respondents

Analysis Samples

The IMAGINE delay-of-birth outcome analysis and other primary and secondary outcome analyses necessarily used different sample compositions. Figures 5 (Bangladesh) and 6 (Niger) show the flow of the ultimate analysis samples from baseline to endline with unweighted percentages for key groups within the samples by treatment (top) and control (bottom). At the end (furthest to the right) of each panel within country, two analytic groups are shown: respondents *included* for the delay-of-birth analysis (top right portion of each panel) and the respondents *excluded* from the delay-of-birth analysis. These two groups were combined into the 'full analytic sample' for secondary analyses, such as health services utilization.

In both figures below, the left most bar represents the full baseline sample. Next, the baseline sample is divided by four groups. Two of these groups were excluded from all endline analyses: 1) endline non-respondents and 2) those who completed an endline survey but reported a first child date of birth less than 266 days after the intervention started. The remaining eligible respondents are then divided by marital status and whether they had a birth by the endline survey. Respondents who were married at least 266 days prior to the endline survey were considered eligible for the delay-of-birth analysis, among other primary and secondary outcome analyses. Unmarried respondents and those who married less than 266 before the endline survey were not eligible for the delay-of-birth analysis but are included in other IMAGINE outcome analyses.

Figure 5. Bangladesh flow chart from baseline sample to endline analysis samples.

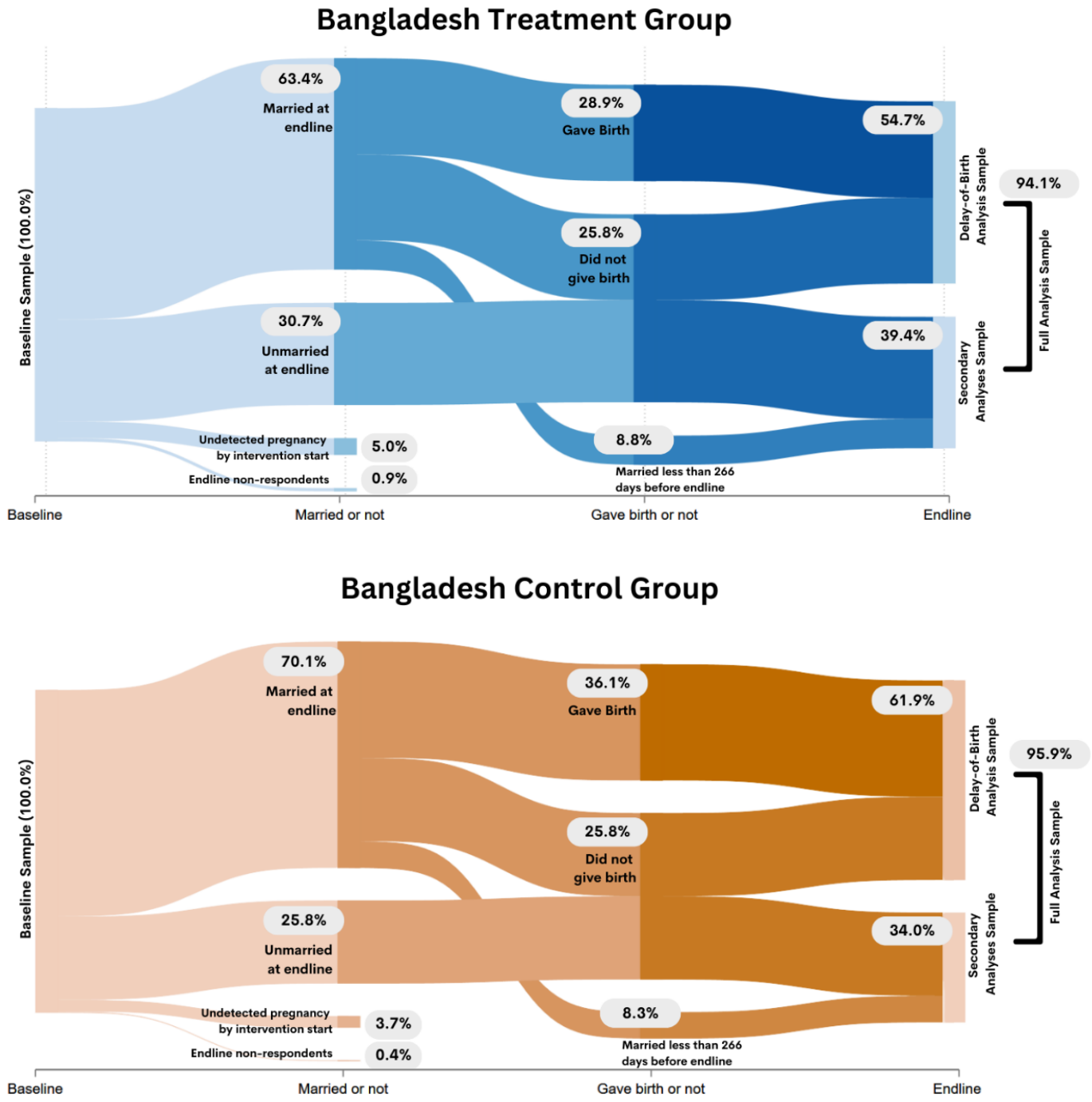
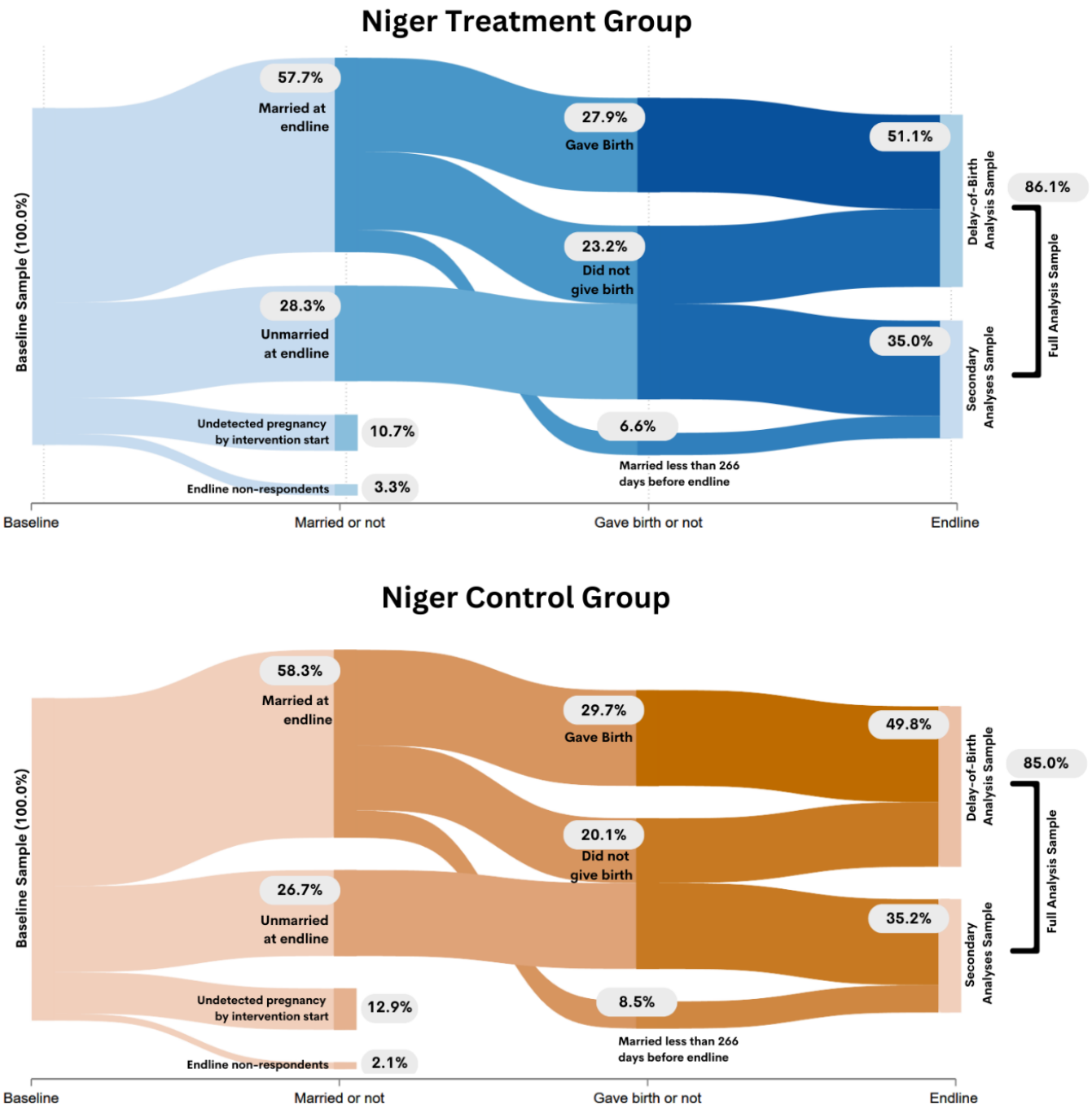


Figure 6. Niger flow chart from baseline sample to endline analysis samples.



Delay-of-Birth Outcome

The exclusion criterion for the delay-of-birth outcome based on date of marriage introduces the potential of a selection bias in the analysis, with the risk of a confound with treatment assignment. Differences in marriage during the study, total rates of marriage, birth rates among married respondents (which could also correspond to a reduced hazard for those in the treatment group), and birth rates in the full sample were all evaluated as potentially contributing to differences in the risk set. These results were not statistically incorporated into the analysis but may be considered caveats for interpretation of the time-to-birth results.

Bangladesh

The time-to-event models in Bangladesh showed no statistically significant treatment difference in time to first birth after marriage. Preliminary analyses, though, showed an intervention difference in marriage during the study: Respondents in the control condition were significantly more likely to marry during the study (before the cutoff) than respondents in the treatment condition. The sequential Cox model strategy partially mitigates this confound at a coarse level, yet this difference complicates the interpretation of the main finding.

Preliminary Analyses

During the study (i.e., between the baseline survey and the endline cutoff), significantly more control area respondents got married (49.3%) than the treatment group (41.7%; b = -0.33, SE = 0.14, t[89] = -2.33, p = .022). Table 6, above, shows this evolving difference in marriage at endline, with no significant difference in marriage proportion at baseline. The result suggests a possible impact of the intervention on timing of marriage, which led to differences in the "risk groups" of being married before the cutoff and not yet giving birth. Following from this, Table 8 shows that fewer respondents in the treatment group (30.8%) than in the control group (37.7%) gave birth over the course of the study, though this difference was not statistically significant when conditioned on marriage.

Table 8. Bangladesh marriage and birth descriptive statistics by condition

	Treatment		Control				
	n	**%**	**n**	**%**	**OR**	**95% CI**	***p***
Married before cutoff[a]	1289	**57.7**	1208	**64.5**	0.78	[0.57, 1.06]	.104
Gave birth by endline[b]	1289	**30.8**	1208	**37.7**	0.77	[0.60, 0.99]	.045*

*n = unweighted sample size; OR = odds ratio; CI = confidence interval; * Significantly different, p<.05*
Note: Unweighted analyses were most appropriate for Table 8 outcomes due to timing of marriage being incorporated in the endline weighting procedure.
a: married at least 266 days prior to endline survey (cutoff), among the full endline sample
b: among full endline sample

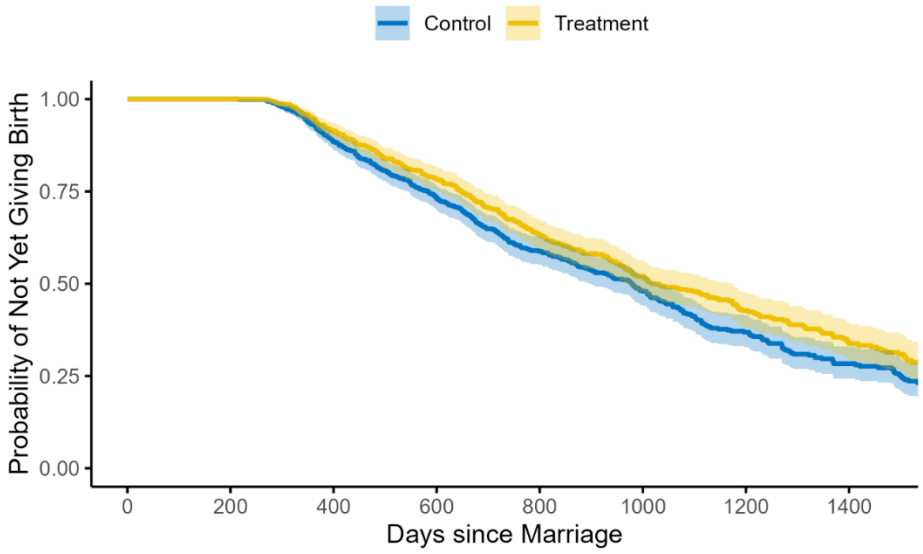
The sequential Cox proportional hazards regression model, using each respondent's date of marriage as her individual "day zero," showed no significant effect of intervention on the number of days from marriage to first birth, b = -0.002, SE = 0.100, t (356) = -0.02, p = .984. The estimated hazard ratio was 1.00, 95% CI [0.82, 1.21], indicating no discernible difference

between conditions in the likelihood of giving birth at any given point after marriage. The median days to first birth were 667 days in the treatment group (25th percentile= 443 days; 75th percentile=964 days) and 646 days in the control group (25th percentile= 441 days; 75th percentile=970 days).

In a Cox regression, this result reflects extrapolation from censored data (respondents who did not give birth before endline) to a scenario where every participant eventually gives birth, though it does not compensate for differences between respondents who enter the risk set (by being married before the cutoff) and those who don't, except insofar as that difference is reflected in weighting.

Cumulative survival curves in Figure 7 are based on the original data, before the transformations for the sequential Cox regression, with the caveat that these plots are descriptive rather than corresponding directly to the inferential tests. The curves suggest a modest difference in the direction of a delay attributable to the intervention, but with overlapping confidence intervals at most points along the curves.

Figure 7. Survival curve for delay-of-birth by treatment group in Bangladesh



Niger

The time-to-event models in Niger showed a modest point estimate for an increase (delay) in time to first birth for respondents in the treatment versus control conditions, but this difference was not statistically significant.

Preliminary Analyses

There was no statistically significant confounding of marriage with the delay-of-birth outcome. As shown in Table 9, the treatment and control groups did not significantly differ in the percent of respondents who were married or who had given birth after baseline.

Table 9. Niger marriage and birth descriptive statistics by condition

	Treatment		Control				
	n	**%**	**n**	**%**	**OR**	**95% CI**	**p**
Married before cutoff[a]	1089	**59.5**	1034	**56.7**	1.10	[0.72, 1.69]	.646
Gave birth by endline[b]	1089	**34.7**	1034	**36.5**	0.91	[0.69, 1.22]	.534

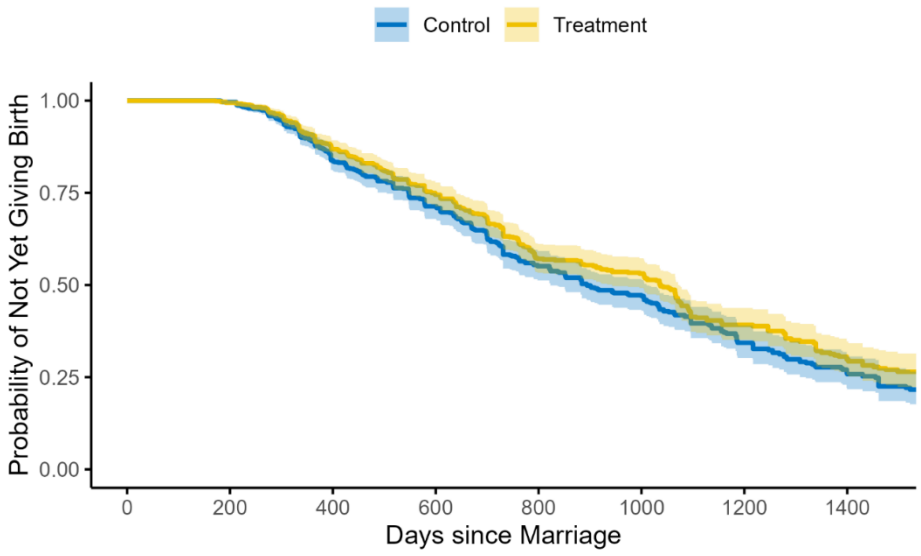
*n = unweighted sample size; OR = odds ratio; CI = confidence interval; * Significantly different, p<.05;*
Note: Unweighted analyses were most appropriate for Table 9 outcomes due to timing of marriage being incorporated in the endline weighting procedure.

a: married at least 266 days prior to endline survey (cutoff), among the full endline sample
b: among full endline sample

The Cox regression model showed no significant effect of intervention on the number of days from marriage to first birth, b = -0.113, SE = 0.112, t (328) = -1.01, p = .312. The estimated hazard ratio was 0.89, 95% CI [0.72, 1.11], suggesting that the probability of a treatment-group respondent having given birth given a particular duration of marriage was approximately 89% that of comparable control-group respondent. The median days to first birth were 657 days [IQR 397-1011 days] in the treatment group and 642 days [IQR 394-1000 days] in the control group.

Cumulative survival curves in Figure 8 are based on the original data, before the transformations for the sequential Cox regression, with the caveat that these plots are descriptive rather than corresponding directly to the inferential tests. The curves suggest a modest difference in the direction of a delay attributable to the intervention, but with overlapping confidence intervals at most points along the curves (not statistically significant).

Figure 8. Survival curve for delay-of-birth by treatment group in Niger



Primary Family Planning Outcomes

Bangladesh

Table 10 shows descriptive statistics and model results by treatment and previous birth subgroup (gave birth v. no birth) for the primary family planning use outcomes in Bangladesh. No significant intervention effects were observed for any of the four primary family planning indicators. To capture change during the study period, analyses of current contraception use, modern contraception use, and unmet need for contraception included only married, non-pregnant respondents; only married respondents who had never used contraception at baseline were included in analyses of ever-use (i.e., lifetime use) of contraception.

Across treatment and control groups, relatively high percentages of respondents reported ever or current use of contraception (80.8% and 61.6%, respectively, in the treatment group and 77.0% and 59.1%, respectively, in the control group). Moreover, over half of respondents in both the treatment group (54.9%) and the control group (54.4%) reported currently using a modern method of contraception. No significant differences between treatment and control group were observed in subgroup analyses by previous birth. Higher rates of contraceptive use (ever, current and modern method use) were observed among respondents who gave birth compared to those with no birth.

Table 10. Bangladesh primary family planning outcomes – model results and weighted descriptive statistics for overall sample and within subgroups by birth status

	Treatment		Control				
	n	**%**	**n**	**%**	**OR**	**95% CI**	***p*** [a]
Ever use of contraception[b]	**765**	**80.8**	**704**	**77.0**	**1.33**	**[0.75, 2.35]**	**.646**
Gave birth	337	92.1	369	87.3	1.89	[0.86, 4.15]	.111
No birth	428	71.8	335	65.6	1.40	[0.70, 2.80]	.336
Current contraceptive use[c]	**830**	**61.6**	**814**	**59.1**	**1.13**	**[0.80, 1.60]**	**.655**
Gave birth	438	73.5	502	71.4	1.13	[0.73, 1.72]	.586
No birth	392	48.3	312	39.3	1.37	[0.78, 2.41]	.276
Modern contraceptive use[c]	**830**	**54.9**	**814**	**54.4**	**1.08**	**[0.76, 1.52]**	**.666**
Gave birth	438	67.8	502	68.2	1.02	[0.67, 1.56]	.912
No birth	392	40.4	312	32.1	1.44	[0.81, 2.58]	.215
Unmet need[c]	**830**	**15.6**	**814**	**21.2**	**0.70**	**[0.50, 0.97]**	**.128**
Gave birth	438	22.1	502	26.4	0.79	[0.56, 1.12]	.184
No birth	392	8.4	312	12.7	0.63	[0.36, 1.08]	.094

*n = weighted sample size; % = weighted prevalence rate; OR = odds ratio for effect of treatment; CI = confidence interval for odds ratio; *significantly different from control at p<.05 after controlling for multiple comparisons; All overall (non-subgroup) analyses were considered one family of comparisons; subgroup analyses were not adjusted for multiple comparisons.*

a: p values reported for overall (non-subgroup) analyses are Benjamini-Hochberg adjusted values after controlling for multiple comparisons;
b: among married respondents, who had never used contraception at baseline (previous use of contraception was not used in the analysis model)
c: among married, non-pregnant respondents

Table 11 shows weighted rates of current contraceptive usage by method and treatment group in Bangladesh. Among current contraceptive users, the most common methods of contraception included the pill, male condoms, and injectables. However, there was no significant effect of the intervention on type of method used.

Table 11. Bangladesh current contraceptive use by method - weighted descriptive statistics

	Treatment (n=511)	Control (n=481)
	%	**%**
Female sterilization	**0**	**0**
Male sterilization	**0**	**0**
IUD	**0**	**0**
Injectable	**15.9**	**16.1**
Implant	**0.5**	**0.9**
Pill	**50.8**	**58.5**
Male condom	**25.8**	**18.8**
Female condom	**0**	**0**
Emergency contraception	**0.4**	**0.4**
Standard days	**0.2**	**0.2**
Rhythm	**6.4**	**2.7**
Withdrawal	**5.8**	**6.8**
Lactational Amenorrhea	**0.3**	**0**
Other modern method	**0**	**0.2**
Traditional	**0.2**	**0.2**

n = weighted sample size;

Endline face-to-face and phone respondents only. Rates shown are among current contraceptive users;

Niger

Table 12 shows descriptive statistics and model results by treatment and previous birth subgroup (gave birth v. no birth) for the primary family planning use outcomes in Niger. In contrast to Bangladesh, there was a statistically significant effect of the intervention on ever-use of contraception, current use of contraception, and current use of a modern contraceptive method.

A greater percentage of respondents in the treatment group (34.1%) reported ever using contraception than in the control group (19.1%). Likewise, more respondents in the treatment group reported currently using contraception (29.4% treatment v. 17.7% control) and using a modern method of contraception (24.7% treatment v. 13.6% control). Unmet need for contraception did not significantly differ based on treatment group (27.1% treatment v. 30.4% control).

Significant differences in family planning use outcomes were found between treatment and control among both the previous birth (gave birth) and no previous birth (no birth) subgroups. Compared to control, treatment respondents with a previous birth had higher rates of ever use of contraception (45.8% treatment v. 28.0% control) and modern method use (33.1% treatment v. 19.6% control). Among the no previous birth subgroup, treatment respondents had higher rates of ever use of contraception (20.7% treatment v. 7.8% control), current contraceptive use, (15.6% treatment v. 4.8% control), and modern method use (12.8% treatment v. 3.9% control), compared to control. Higher rates of contraceptive use were observed among respondents with previous births compared to those who did not give birth, similar to results in Bangladesh.

Table 12. Niger primary family planning outcomes – model results and weighted descriptive statistics for overall sample and within subgroups by birth status

	Treatment		Control				
	n	**%**	**n**	**%**	**OR**	**95% CI**	***p*** [a]
Ever use of contraception[b]	**738**	**34.1**	**734**	**19.1**	**2.28**	**[1.30, 4.00]**	**.020***
Gave birth	394	45.8	411	28.0	2.21	[1.04, 4.68]	.039*
No birth	344	20.7	323	7.8	3.30	[1.59, 6.85]	.002*
Current contraceptive use[c]	**634**	**29.4**	**628**	**17.7**	**1.95**	**[1.09, 3.48]**	**.032***
Gave birth	373	39.0	387	25.7	1.81	[0.87, 3.75]	.112
No birth	261	15.6	241	4.8	3.78	[1.58, 9.09]	.003*
Modern contraceptive use[c]	**634**	**24.7**	**628**	**13.6**	**2.12**	**[1.20, 3.74]**	**.020***
Gave birth	373	33.1	387	19.6	2.02	[1.02, 3.98]	.043*
No birth	261	12.8	241	3.9	3.58	[1.42, 9.04]	.008*
Unmet need[c]	**634**	**27.1**	**628**	**30.4**	**0.81**	**[0.52, 1.27]**	**.355**
Gave birth	373	30.7	387	33.1	0.84	[0.49, 1.43]	.519
No birth	261	21.8	241	26.0	0.79	[0.39, 1.58]	.498

*n = weighted sample size; % = weighted prevalence rate; OR = odds ratio for effect of treatment; CI = confidence interval for odds ratio; *significantly different from control at p<.05 after controlling for multiple comparisons; All overall (non-subgroup) analyses were considered one family of comparisons; subgroup analyses were not adjusted for multiple comparisons.*

a: p values reported for overall (non-subgroup) analyses are Benjamini-Hochberg adjusted values after controlling for multiple comparisons;
b: among married respondents, who had never used contraception at baseline (previous use of contraception was not used in the analysis model)
c: among married, non-pregnant respondents

Table 13 shows weighted rates of current contraceptive usage by method and treatment group in Niger. Among those currently using contraception, injectables, the pill, traditional methods, and implants were the most common forms of contraception. No statistically significant differences in rates of method use were observed between treatment and control groups.

Table 13. Niger current contraceptive use by method - weighted descriptive statistics

	Treatment (n=186)	Control (n=111)
	%	%
Female sterilization	**0**	**1.5**
Male sterilization	**0.5**	**0**
IUD	**0.9**	**0**
Injectable	**53.7**	**38.9**
Implant	**6.5**	**7.4**
Pill	**25.3**	**31.1**
Male condom	**0**	**0**
Female condom	**0**	**0**
Emergency contraception	**0**	**0**
Standard days	**0**	**0**
Rhythm	**0**	**0**
Withdrawal	**0**	**0.8**
Other modern method	**0.5**	**0**
Traditional	**16.2**	**24.7**

n = weighted sample size;

Endline face-to-face and phone respondents only. Rates shown are among current contraceptive users; Note: Lactational amenorrhea was not a response option for contraceptive methods in Niger.

Secondary Outcomes

Bangladesh

Table 14 shows weighted rates of contraceptive methods ever used by method and treatment in Bangladesh. Among respondents who had ever used contraception, the pill, male condoms, withdrawal, injectables, and the rhythm method were most frequently endorsed forms of contraception. Respondents in the treatment group (58.8%) were more likely than respondents in the control group (46.6%) to have ever used male condoms; the former were also more likely to have used rhythm methods of contraception (14.1% v. 7.7%).

Table 14. Bangladesh contraceptive methods ever used

	Treatment (n=782)	Control (n=726)
	%	**%**
Female sterilization	**0**	**0**
Male sterilization	**0**	**0**
IUD	**0**	**0**
Injectable	**12.8**	**15.9**
Implant	**0.5**	**1.5**
Pill	**74.9**	**80.5**
Male condom	**58.8**	**46.6 ***
Female condom	**0**	**0**
Emergency contraception	**1.8**	**2.8**
Standard days	**0.4**	**0.1**
Rhythm	**14.1**	**7.7 ***
Withdrawal	**16.5**	**19.3**
Lactational Amenorrhea	**0.4**	**0**
Other modern method	**0.1**	**0.1**
Traditional	**0.1**	**0.1**

*n = weighted sample size; *significantly different from control at p<.05 after controlling for multiple comparisons; All indicators included in table were considered one family of comparisons.*

Endline face-to-face and phone respondents only. Rates shown are among current contraceptive users.

Table 15 shows descriptive statistics and model results for binary secondary outcomes in Bangladesh. Some intervention effects were observed for health and economic indicators. A larger percentage of respondents reported participating in vocational training in the treatment group (51.8%) compared to the control group (6.3%). Respondents in the treatment group (25.5%) were also more likely to have participated in income generating activities within the past year compared to respondents in the control group (13.9%).

Table 15. Bangladesh health service utilization and livelihood outcomes

	Treatment		Control				
	n	**%**	**n**	**%**	**OR**	**95% CI**	***p*** [a]
Health service utilization							
Knows of a place to access FP[†]	1044	**98.5**	940	**96.7**	2.57	[1.03, 6.42]	.143
Visited by CHW in past 6 months[†]	1275	**8.5**	1176	**8.2**	1.09	[0.78, 1.51]	.809
Discussed FP with a CHW[†]	1275	**3.4**	1176	**3.0**	1.14	[0.67, 1.94]	.896
Visited health facility in past 6mo.[†]	1276	**35.2**	1182	**36.5**	1.01	[0.79, 1.29]	.998
Discussed FP at health facility[†]	1276	**6.1**	1182	**4.8**	1.31	[0.79, 2.17]	.537
Satisfaction with SRHR service	436	**49.1**	404	**56.8**	0.69	[0.44, 1.07]	.250
Recommend health service	436	**97.8**	404	**97.5**	1.17	[0.43, 3.15]	.896
Livelihoods							
Has savings[†]	1044	**62.2**	940	**62.2**	0.99	[0.64, 1.54]	.970
Has capital assets[†]	1044	**36.4**	940	**33.9**	1.24	[0.86, 1.78]	.557
Participation in VSLA[†]	1044	**9.4**	940	**7.3**	1.29	[0.80, 2.08]	.484
Participation in vocational training[†]	1044	**51.8**	940	**6.3**	17.87	[11.27, 28.33]	.013*
Income generation in past week[†]	1276	**13.1**	1182	**8.8**	1.71	[1.02, 2.85]	.182
Income generation in past 12mo.[†]	1276	**25.5**	1182	**13.9**	2.42	[1.62, 3.60]	.007*

*n = weighted sample size; % = weighted prevalence rate; OR = odds ratio for effect of treatment; CI = confidence interval for odds ratio; *significantly different from control at p<.05 after controlling for multiple comparisons; All indicators included in table were considered one family of comparisons; Endline face-to-face and phone respondents only.† signifies indicators where the baseline measure of the outcome was used as a covariate in the analysis model.*

a: p values reported are Benjamini-Hochberg adjusted values after controlling for multiple comparisons

Tables 16-18 show model results and weighted rates of ordinal (table 16a-f) and nominal family planning related outcomes (table 17a-d), and ordinal socio-economic outcomes (table 18a-d) in Bangladesh. The intervention was associated with changes in ideal timing of birth after marriage, with treatment respondents wanting longer delays between marriage and first birth compared to control. No intervention effects were found for constructs such as ideal age for first child, months of continuous contraceptive use, or preferred number of children. While the mean ideal age for having a first child was about 20 years old (treatment=20.1 and control=20.2) the mean actual age of first birth (among the subsample who gave birth during the study) was 18.2 years old (median=18, SD=1.6) in the treatment group and 18.4 years old (median=18, SD=1.6) in the control group. There was, however, a significant intervention effect on mobility. Respondents in the treatment group (39.5%) reported a higher level of social and economic mobility than respondents in the control group (27.5%), though they did not significantly differ in reported monthly income, savings, or assets.

Tables 16a-f. Bangladesh family planning related outcomes

	Treatment % (n=1031, $\bar{x}$ =25.7 , SD=9.0)			Control % (n=912, $\bar{x}$ =24.2 , SD=9.6)					
a: Ideal timing of birth after marriage (mos.)†	**≤ 23**	**24-35**	**≥ 36**	**≤ 23**	**24-35**	**≥ 36**	**OR**	**95% CI**	***p*** [a]
	14.0	63.6	22.4	23.7	57.4	18.9	1.67	[1.23, 2.28]	.006

	Treatment % (n=1030, $\bar{x}$ =20.1, SD=1.3)			Control % (n=921, $\bar{x}$ =20.2, SD=1.2)					
b: Ideal age for first child†	**<=19**	**20**	**>=21**	**<=19**	**20**	**>=21**	**OR**	**95% CI**	***p***
	17.0	57.5	25.5	14.1	62.9	23.0	0.99	[0.75, 1.32]	.949

	Treatment % (n=521, $\bar{x}$ =5.9, SD=7.6)					Control % (n=486, $\bar{x}$ =6.4, SD=7.2)							
c: Months of continuous contraceptive use#	**0-3**	**4-7**	**8-11**	**12-15**	**16+**	**0-3**	**4-7**	**8-11**	**12-15**	**16+**	**OR**	**95% CI**	***p*** [a]
	53.3	18.8	8.3	9.6	10.2	47.8	22.1	10.1	9.8	10.2	0.87	[0.63, 1.21]	.835

	Treatment % (n=1256, $\bar{x}$ =2.1, SD=0.6)					Control % (n=1166, $\bar{x}$ =2.0, SD=0.5)							
d: Preferred number of children in lifetime†	**0**	**1**	**2**	**3**	**4+**	**0**	**1**	**2**	**3**	**4+**	**OR**	**95% CI**	***p*** [a]
	2.1	4.0	81.3	10.7	1.9	2.8	2.6	86.1	7.7	0.8	1.11	[0.77, 1.61]	.835

	Treatment % (n=704)			Control % (n=687)					
e: Husband's preferred number of children	**Fewer**	**Same**	**More**	**Fewer**	**Same**	**More**	**OR**	**95% CI**	***p*** [a]
	1.1	96.1	2.8	2.4	93.0	4.6	0.86	[0.44, 1.70]	.835

	Treatment % (n=521)		Control % (n=486)				
f: Rights-based family planning#	**≤ 1**	**≥ 2**	**≤ 1**	**≥ 2**	**OR**	**95% CI**	***p*** [a]
	82.7	17.3	81.5	18.5	0.92	[0.59, 1.43]	.835

*n = weighted sample size; OR = odds ratio for effect of treatment; CI = confidence interval for odds ratio; *significantly different from control at p<.05 after controlling for multiple comparisons; All indicators included in table were considered one family of comparisons; † signifies indicators where the baseline measure of the outcome was used as a covariate in the analysis model;* [#]among respondents who indicated previous use of contraception. *a: p values reported are Benjamini-Hochberg adjusted values after controlling for multiple comparisons*

Tables 17a-d. Bangladesh family planning related outcomes

a: Access point to initiate recent/current FP method	Treatment % (n=521)			Control % (n=486)			*F*	df	*p* [a]
	Public	Private/NGO	Other	Public	Private/NGO	Other			
	7.4	76.7	15.9	8.5	73.5	17.9	0.20	2, 88	.999

b: Most recent access point for FP method	Treatment % (n=521)			Control % (n=486)			*F*	df	*p* [a]
	Public	Private/NGO	Other	Public	Private/NGO	Other			
	10.6	73.6	15.8	12.7	68.8	18.5	0.40	2, 88	.999

c: Husband's involvement in FP use decision[†]	Treatment % (n=377)			Control % (n=355)			*F*	df	*p* [a]
	Her decision	His decision	Joint decision	Her decision	His decision	Joint decision			
	3.0	2.5	94.5	2.6	6.6	90.9	2.84	2, 87	.256

d: Reason for discontinuation of FP method[#]	Treatment % (n=190)			Control % (n=164)			*F*	df	*p* [a]
	Wanted pregnancy	Infrequent sex	Other	Wanted pregnancy	Infrequent sex	Other			
	51.2	29.5	19.3	51.6	28.5	19.9	0.04	2, 82	.964

*n = weighted sample size; df = Denominator degrees of freedom, numerator degrees of freedom for F test; *significantly different from control at p<.05 after controlling for multiple comparisons; All indicators included in table were considered one family of comparisons; due to small sample sizes in some categories of nominal outcomes, a reduced set of covariates was used in the analysis model for these outcomes including: age, education, and marital status;* [†] *Among respondents who previously used contraception and their husband was aware of the contraceptive use; The infrequent sex category for reason for discontinuation of FP method also include the 'husband away' response option;*

[#]Among respondents who indicated previous use of contraception and discontinued use.

a: p values reported are Benjamini-Hochberg adjusted values after controlling for multiple comparisons

Tables 18a-d. Bangladesh non-family planning outcomes

	Treatment % (n=1044)		Control % (n=940)				
	≤ 2	3	≤ 2	3	OR	95% CI	*p* [a]
a: Mobility score[†]	60.5	39.5	72.5	27.5	1.81	[1.24, 2.64]	.012*

	Treatment % (n=1044)				Control % (n=940)						
	0	≤1000	1001-2500	≥3000	0	≤1000	1001-2500	≥3000	OR	95% CI	*p* [a]
b: Reported monthly income	91.5	3.9	2.1	2.5	95.0	1.8	1.1	2.1	1.68	[0.99, 2.85]	.106

	Treatment % (n=1044)				Control % (n=940)						
	0	≤300	301-1200	≥1201	0	≤300	301-1200	≥1201	OR	95% CI	*p* [a]
c: Reported savings[†]	37.8	19.3	19.6	23.3	37.8	23.5	19.6	19.1	1.12	[0.81, 1.56]	.490

	Treatment % (n=1044)				Control % (n=940)						
	0	1	2	3+	0	1	2	3+	OR	95% CI	*p* [a]
d: Ownership of household assets & resources Index [†]	24.4	47.2	24.4	4.0	27.1	46.2	24.2	2.5	1.19	[0.88, 1.62]	.339

*n = weighted sample size; OR = odds ratio for effect of treatment; CI = confidence interval for odds ratio; *significantly different from control at p<.05 after controlling for multiple comparisons; All indicators included in table were considered one family of comparisons.*

a: p values reported are Benjamini-Hochberg adjusted values after controlling for multiple comparisons

Scale Outcome Analyses

Multiple positive impacts of the intervention were observed in the latent change score models for secondary outcome scales in Bangladesh. Compared to the control group, the treatment group showed significantly greater improvements in knowledge about pregnancy risk and more positive expectations about family planning use, delaying childbirth, and income generation (see Table 19). Respondents in the treatment group also showed a significant decline in their belief of family planning myths relative to the control group, as well as increases in social cohesion, collective efficacy, equitable normative expectations about girls' roles, and self-efficacy to engage in economic activities.

Endline-only analyses included measurements of secondary outcomes that were available at endline only, given that large portions of the sample were excluded from baseline measurement on these scales (e.g., due to only married individuals being asked to contribute responses at baseline). Table 20 shows the results of the endline-only regression analyses. These analyses suggested no significant differences between the treatment and control group on indicators of self-efficacy, interspousal communication, or household and financial decision-making.

Table 19. Bangladesh difference-in-differences analyses on normative, knowledge, and self-efficacy outcome scales

	Treatment (n = 1044)			Control (n = 940)				
	Baseline Mean	**Endline Mean**	**Change**	**Baseline Mean**	**Endline Mean**	**Change**	**DiD**	**CI**
Early pregnancy risk knowledge	2.57	3.28	0.69	2.82	3.11	0.31	**0.38***	[0.18, 0.57]
Expectations about FP use	3.90	4.04	0.14	4.08	4.01	-0.06	**0.20***	[0.05, 0.35]
Expectations about delaying childbirth	3.83	4.01	0.17	4.09	3.96	-0.11	**0.28***	[0.08, 0.48]
Expectations about income generation	3.82	3.89	0.09	4.02	3.78	-0.23	**0.32***	[0.14, 0.50]
Belief in FP myths	2.66	2.22	-0.45	2.42	2.58	0.16	**-0.61***	[-0.79, -0.43]
Self-efficacy to visit a health facility	3.03	3.47	0.43	3.37	3.49	0.19	0.24	[0.00, 0.49]
Social cohesion	3.74	3.86	0.12	4.03	3.96	-0.05	**0.18***	[0.05, 0.30]
Collective efficacy	3.59	3.89	0.28	3.99	3.77	-0.20	**0.48***	[0.17, 0.79]
Normative expectations about girls' roles	3.38	3.41	0.04	3.68	3.32	-0.35	**0.39***	[0.20, 0.58]
Self-efficacy to engage in economic activities	3.76	4.26	0.49	4.06	4.04	0.00	**0.49***	[0.33, 0.65]

*Change = covariate-adjusted change from baseline to endline; DiD = difference-in-differences; *significantly different from control at p<.05 after controlling for multiple comparisons; All indicators included in table were considered one family of comparisons*

Note: baseline and endline means are not adjusted for covariates, whereas change scores are. This explains small discrepancies between the two. The following baseline covariates were included in the analysis model for these outcomes: age, education, attending a health visit in past 6 months, knowledge of a place to obtain family planning, income generation in the past 12 months, possession of savings, possession of assets, marital status, and religious affiliation.

Table 20. Bangladesh endline-only analyses on normative, knowledge, and self-efficacy outcome scales

	Treatment (n = 739) $\bar{x}$	Control (n = 706) $\bar{x}$	Diff	CI
Self-efficacy to use and discuss FP[a]	4.27	4.14	0.13	[0.00, 0.26]
Self-efficacy to refuse sex[a]	3.62	3.50	0.12	[-0.13, 0.38]
FP provider quality[b]	8.71	8.98	-0.27	[-0.71, 0.18]
Frequency of interspousal communication[a]	2.53	2.53	0.00	[-0.09, 0.10]
Household decision-making[c]	1.62	1.59	0.03	[-0.01, 0.06]
Financial decision-making[c]	1.38	1.30	0.08	[0.03, 0.14]

Diff = difference between treatment and control endline means.

[a] 5=high 1=low

[b] 13=high 0=low

[c] 2=high 1=low

Niger – Secondary outcomes

Table 21 shows weighted rates of contraceptive methods ever used by method and treatment in Niger. Among respondents who had ever used contraception, the most common methods of contraception included injectables, the pill, and traditional methods. The treatment and control groups differed only in their use of injectables, with a larger percent of respondents in the treatment group (57.2%) reporting using injectables than in the comparison group (37.2%).

Table 21. Niger contraceptive methods ever used - weighted descriptive statistics by treatment

	Treatment (n=186)	Control (n=111)
	%	**%**
Female sterilization	**0**	**1.1**
Male sterilization	**0**	**0**
IUD	**0.4**	**0**
Injectable	**57.2**	**37.2 ***
Implant	**5.8**	**6.0**
Pill	**33.9**	**35.7**
Male condom	**0**	**0**
Female condom	**0**	**0**
Emergency contraception	**0**	**0**
Standard days	**1.2**	**0**
Rhythm	**0**	**0**
Withdrawal	**0**	**2.5**
Other modern method	**0**	**0**
Traditional	**17.2**	**26.7**
Female sterilization	**0**	**1.1**

*n = weighted sample size; * Significantly different between treatment and control at p<.05; Endline face-to-face and phone respondents only. Rates shown are among current contraceptive users.*

Table 22 shows descriptive statistics and model results for binary secondary outcomes in Bangladesh. Findings related to health and economic indicators were largely in the direction of showing a positive impact of the intervention. A larger percentage of respondents in the treatment group (93.3%) reported knowing of a place to access family planning services than in the control group (80.7%). A larger portion of the treatment group had also been visited by a community health worker in the past 6 months (20.0% treatment v. 8.6% control) or had visited a health facility (36.7% treatment v. 26.6% control) over that same period. Among respondents who had received such services, those in the treatment group more often endorsed discussing family planning with a provider (15.7% for community health visits and 14.2% for health facility visits) than those in the control group (4.4% for community health visits).

Income generation and resource indicators also statistically differed between treatment and control groups. Compared to the control group, a larger percentage of the treatment group reported having savings (49.0% treatment v. 27.3% control) and capital assets (63.0% treatment v. 42.5% control). A larger portion of respondents in the treatment group had also participated in VSLA's (40% treatment v. 11.0% control) and in vocational training (67.6% treatment v. 2.9% control). Finally, there was an impact of the intervention on income generation, with proportionally more respondents in the treatment group reporting income generating activity in the past week and the past year (28.5% and 56.4%, respectively) relative to respondents in the control group (19.0% and 44.2%, respectively).

Table 22. Niger health service utilization and livelihood outcomes

	Treatment		Control				
	n	**%**	**n**	**%**	**OR**	**95% CI**	***p*** [a]
Health service utilization							
Knows of a place to access FP[†]	1029	**93.3**	985	**80.7**	3.41	[2.05, 5.65]	.002*
Visited by CHW in past 6 months[†]	1046	**20.0**	1009	**8.6**	2.64	[1.27, 5.49]	.007*
Discussed FP with a CHW[†]	1046	**15.7**	1009	**4.4**	4.05	[1.77, 9.26]	.004*
Visited health facility in past 6mo.[†]	1046	**36.7**	1009	**26.6**	1.58	[1.16, 2.14]	.005*
Discussed FP at health facility[†]	1046	**14.2**	1009	**6.6**	2.50	[1.52, 4.11]	.003*
Satisfaction with SRHR service	376	**38.3**	259	**33.6**	1.27	[0.77, 2.09]	.380
Recommend health service	376	**99.2**	259	**98.6**	1.79	[0.32, 9.94]	.501
Livelihoods							
Has savings[†]	1029	**49.0**	985	**27.3**	2.62	[1.76, 3.90]	.002*
Has capital assets[†]	1029	**63.0**	985	**42.5**	2.37	[1.43, 3.94]	.013*
Participation in VSLA[†]	1029	**40.0**	985	**11.0**	5.88	[3.16, 10.94]	.002*
Participation in vocational training[†]	1029	**67.6**	985	**2.9**	70.44	[31.75, 156.25]	.003*
Income generation in past week[†]	1046	**28.5**	1009	**19.0**	1.77	[1.21, 2.61]	.005*
Income generation in past 12mo.[†]	1046	**56.4**	1009	**44.2**	1.75	[1.24, 2.47]	.003*

*n = weighted sample size; % = weighted prevalence rate; OR = odds ratio for effect of treatment; CI = confidence interval for odds ratio; *significantly different from control at p<.05 after controlling for multiple comparisons; All indicators included in table were considered one family of comparisons; Endline face-to-face and phone respondents only.† signifies indicators where the baseline measure of the outcome was used as a covariate in the analysis model. a: p values reported are Benjamini-Hochberg adjusted values after controlling for multiple comparisons*

Tables 23-25 show model results and weighted rates of ordinal (Table 23a-f) and nominal family planning related outcomes (table 26a-d), and ordinal socio-economic outcomes (table 27a-c) in Niger. Similar to the findings in Bangladesh, there was little impact of the intervention on constructs associated with family planning in Niger, including ideal timing of birth, ideal age for first child, and preferred number of children. The mean ideal age for having a first child was 17.8 years old for the treatment group and 17.5 in control. The mean actual age of first birth (among the subsample who gave birth during the study) was 17.4 (median=17, SD=2.0) in the treatment group and 17.5 (median=17, SD=2.0) in the control group, quite close to the idealized age for

first birth in both groups. There was, however, a significant difference in assets and resources between the treatment and control group. Respondents in the treatment group reported a higher level of ownership over household assets/resources than respondents in the control group (i.e., scores of 2 or greater).

Tables 23a-f. Niger family planning related outcomes

	Treatment % (n=648, $\bar{x}$ =17.8, SD=8.2)				Control % (n=575, $\bar{x}$ =16.4, SD=7.7)						
a: Ideal timing of birth after marriage (mos.) †	**≤11**	**12-23**	**24-35**	**≥36**	**≤11**	**12-23**	**24-35**	**≥36**	**OR**	**95% CI**	***p*** [a]
	16.0	43.6	31.1	9.3	20.5	46.8	26.1	6.5	1.45	[0.78,2.69]	.284

	Treatment % (n=741, $\bar{x}$ =17.8, SD=2.0)				Control % (n=682, $\bar{x}$ =17.5, SD=2.2)						
	≤15	**16-17**	**18-19**	**≥20**	**≤15**	**16-17**	**18-19**	**≥20**	**OR**	**95% CI**	***p*** [a]
b: Ideal age for first child†	14.2	25.6	33.9	26.3	18.4	28.9	31.5	21.3	1.40	[0.91, 2.16]	.381

	Treatment % (n=181, $\bar{x}$ =5.5, SD=8.2)					Control % (n=72, $\bar{x}$ =8.1, SD=10.1)							
c: Months of continuous contraceptive use#	**0-3**	**4-7**	**8-11**	**12-15**	**16+**	**0-3**	**4-7**	**8-11**	**12-15**	**16+**	**OR**	**95% CI**	***p*** [a]
	55.5	24.0	10.9	6.8	2.7	48.5	11.9	15.0	12.3	12.2	0.56	[0.24, 1.27]	.245

	Treatment % (n=843, $\bar{x}$ =7.2 , SD=3.2)						Control % (n=755, $\bar{x}$ =7.8, SD=3.2)								
d: Preferred number of children in lifetime†	**0**	**1-3**	**4-6**	**7-9**	**10-12**	**13+**	**0**	**1-3**	**4-6**	**7-9**	**10-12**	**13+**	**OR**	**95% CI**	***p*** [a]
	2.6	4.3	41.1	22.0	27.3	2.7	2.8	2.7	31.4	25.6	33.8	3.8	0.71	[0.45, 1.11]	.268

	Treatment % (n=364)			Control % (n=310)					
e: Husband's preferred number of children	**Fewer**	**Same**	**More**	**Fewer**	**Same**	**More**	**OR**	**95% CI**	***p*** [a]
	2.2	41.9	55.9	2.2	42.5	55.3	1.07	[0.62, 1.85]	.815

	Treatment % (n=181)		Control % (n=72)				
f: Rights-based family planning#	**≤ 4**	**5**	**≤ 4**	**5**	**OR**	**95% CI**	***p*** [a]
	46.2	53.8	60.5	39.5	1.91	[0.94, 3.87]	.438

*n = weighted sample size; OR = odds ratio for effect of treatment; CI = confidence interval for odds ratio; *significantly different from control at p<.05 after controlling for multiple comparisons; All indicators included in table were considered one family of comparisons; † signifies indicators where the baseline measure of the outcome was used as a*

covariate in the analysis model; # among respondents who indicated previous use of contraception. *a: p values reported are Benjamini-Hochberg adjusted values after controlling for multiple comparisons*

Tables 24a-d. Niger family planning related outcomes

	Treatment % (n=181)			Control % (n=72)					
a: Access point to initiate recent/current FP method	**Public**	**Private**	**Other**	**Public**	**Private/NGO**	**Other**	***F***	**df**	***p*** [a]
	83.2	3.5	13.2	65.6	6.0	28.4	1.85	2, 70	.328

	Treatment % (n=181)			Control % (n=72)					
b: Most recent access point for FP method	**Public**	**Private**	**Other**	**Public**	**Private/NGO**	**Other**	***F***	**df**	***p*** [a]
	83.8	2.3	14.0	65.6	6.0	28.4	2.20	2, 70	.472

	Treatment % (n=159)			Control % (n=89)					
c: Husband's involvement in FP use decision[†]	**Her decision**	**His decision**	**Joint decision**	**Her decision**	**His decision**	**Joint decision**	***F***	**df**	***p*** [a]
	14.3	10.0	75.8	12.0	12.2	75.8	0.22	2, 54	.807

	Treatment % (n=68)			Control % (n=30)					
d: Reason for discontinuation of FP method[#]	**Wanted pregnancy**	**Infrequent sex**	**Other**	**Wanted pregnancy**	**Infrequent sex**	**Other**	***F***	**df**	***p*** [a]
	50.6	18.7	30.7	60.9	17.2	22.0	0.34	2, 41	.948

*n = weighted sample size; df = Denominator degrees of freedom, numerator degrees of freedom for F test; *significantly different from control at p<.05 after controlling for multiple comparisons; All indicators included in table were considered one family of comparisons; due to small sample sizes in some categories of nominal outcomes, a reduced set of covariates was used in the analysis model for these outcomes including: age, education, and marital status;* [†] *Among respondents who previously used contraception and their husband was aware of the contraceptive use; The infrequent sex category for reason for discontinuation of FP method also include the 'husband away' response option;*
[#]Among respondents who indicated previous use of contraception and discontinued use.

a: p values reported are Benjamini-Hochberg adjusted values after controlling for multiple comparisons

Tables 25a-c. Niger non-family planning outcomes

	Treatment % (n=1029)		Control % (n=985)				
	≤ 2	3	≤ 2	3	OR	95% CI	*p* [a]
a: Mobility score[†]	51.7	48.3	53.5	46.5	1.09	[0.66, 1.80]	.999

	Treatment % (n=1029)				Control % (n=985)						
	0	≤1999	2000-3999	≥4000	0	≤1999	2000-3999	≥4000	OR	95% CI	*p* [a]
b: Reported monthly income[†]	20.0	19.9	26.5	33.6	12.0	32.1	26.2	29.6	1.06	[0.66, 1.71]	.799

	Treatment % (n=1029)					Control % (n=985)							
	0	1	2	3	4	0	1	2	3	4	OR	95% CI	*p* [a]
c: Ownership of household assets & resources Index [†]	19.7	24.7	25.3	18.5	11.8	37.7	28.5	17.6	9.2	7.0	2.45	[1.57, 3.82]	.003*

*n = weighted sample size; OR = odds ratio for effect of treatment; CI = confidence interval for odds ratio; *significantly different from control at p<.05 after controlling for multiple comparisons; All indicators included in table were considered one family of comparisons.*

a: p values reported are Benjamini-Hochberg adjusted values after controlling for multiple comparisons

Scale Outcome Analyses

The following section report on IMAGINE priority scale results. As mentioned above, the composition of the scales can be found in Appendix E. Additionally, reliability analyses were conducted on 12 of the IMAGINE priority scales (reliability analysis results can be found in Appendix D). Overall, the analyzed scales were found to have good reliability, nearly all having coefficient alpha statistics of .70 or higher (in both countries).

Table 26 shows results of the latent change score analyses in Niger. Difference-in-differences estimation in Niger showed a positive impact of the intervention on several secondary outcomes. Compared to respondents in the control group, respondents in the treatment group showed a greater decline in beliefs in family planning myths. Respondents in the treatment group also endorsed more equitable normative expectations about girls' roles and greater increases in self-efficacy to visit a health facility and to engage in economic activities. Although respondents in both the treatment and control groups showed a decline over time in self-efficacy to use and discuss family planning, this decline was less severe in the treatment group, again suggestive of a positive impact of the intervention.

Endline-only analyses (see Table 27) also showed several significant differences between groups in favor of an intervention effect. These differences included: greater self-efficacy to refuse sex ($\bar{x}$ difference = 0.28), greater involvement in household decision-making ($\bar{x}$ difference = 0.11), and greater involvement in financial decision-making ($\bar{x}$ difference = 0.08) among respondents in the treatment group versus control group.

Table 26. Niger Difference-in-differences analyses on selected secondary outcome scales

	Treatment (n = 1029)			Control (n = 985)				
	Baseline Mean	**Endline Mean**	**Change**	**Baseline Mean**	**Endline Mean**	**Change**	**DiD**	**CI**
Early pregnancy risk knowledge	2.08	2.92	0.83	2.16	2.83	0.64	0.19	[-0.07, 0.45]
Expectations about FP use	2.67	3.31	0.63	2.58	3.02	0.47	0.16	[-0.14, 0.47]
Expectations about delaying childbirth	3.01	3.41	0.41	3.04	3.33	0.29	0.12	[-0.16, 0.39]
Expectations about income generation	3.14	3.65	0.52	3.23	3.58	0.36	0.16	[-0.18, 0.50]
Self-efficacy to use and discuss FP	3.48	3.15	-0.33	3.50	2.82	-0.66	**0.33***	[0.08, 0.57]
Belief in FP myths	3.05	2.48	-0.58	3.03	2.92	-0.14	**-0.44***	[-0.60, -0.28]
Self-efficacy to visit a health facility	2.57	3.16	0.58	2.71	2.82	0.09	**0.49***	[0.18, 0.81]
Social cohesion	3.66	3.68	0.01	3.70	3.71	0.00	0.01	[-0.15, 0.18]
Collective efficacy	3.10	3.35	0.26	3.14	3.00	-0.14	0.40	[0.00, 0.80]
Normative expectations about girls' roles	2.56	2.82	0.26	2.60	2.63	0.05	**0.21***	[0.09, 0.34]
Self-efficacy to engage in economic activities	3.20	3.88	0.67	3.27	3.70	0.41	**0.26***	[0.05, 0.47]

*Change = covariate-adjusted change from baseline to endline; DiD = difference-in-differences; *significantly different from control at p<.05 after controlling for multiple comparisons; All indicators included in table were considered one family of comparisons;*

Note: baseline and endline means are not adjusted for covariates, whereas change scores are. This explains small discrepancies between the two. The following baseline covariates were included in the analysis model for these outcomes: age, education, attending a health visit in past 6 months, knowledge of a place to obtain family planning, income generation in the past 12 months, possession of savings, possession of assets, marital status, and ethnicity.

Table 27. Niger Post-only regression analyses on secondary outcome scales

	Treatment (n = 751) $\bar{x}$	Control (n = 734) $\bar{x}$	Diff	CI
Self-efficacy to refuse sex[a]	1.99	1.71	**0.28***	[0.06, 0.50]
FP provider quality[b]	8.66	8.17	0.49	[-0.26, 1.24]
Frequency of interspousal communication[a]	2.78	2.82	-0.04	[-0.22, 0.13]
Household decision-making[c]	1.45	1.34	**0.11***	[0.05, 0.17]
Financial decision-making[c]	1.35	1.28	**0.08***	[0.03, 0.11]

*Diff = difference between treatment and control endline means; *significantly different from control at p<.05 after controlling for multiple comparisons; All indicators included in table were considered one family of comparisons*

[a] *5=high 1=low*

[b] *13=high 0=low*

[c] *2=high 1=low*

Dose-Response Analysis

The dose measure was participation in IMAGINE Girls' Groups as self-reported in the endline survey. Response options on the 5-point indicator included "did not participate," "only a few," "some (less than half)," "most (more than half)," and "all." The analysis sample was restricted to the treatment group and this variable used as the predictor in lieu of the binary treatment indicator.

Bangladesh

Table 28 shows the unweighted rates of the IMAGINE Girls' Groups dose measure in Bangladesh. The time-to-first-birth models in Bangladesh showed a statistically significant negative (beneficial) association with participation in Girls' Groups, with a hazard ratio per point difference on the dose measure scale of 0.89, 95% CI [0.82, 0.96], p = .003. This result can be interpreted in the following way: In any given amount of time since marriage, a respondent with one-category higher participation had a model-implied probability of giving birth that was 11% lower than that of a respondent with one category lower participation. Extending across the range of the dose measure, the probability of giving birth by a respondent attending all groups was 39% lower than that of a respondent not attending any groups.[11]

Table 28. Bangladesh descriptive statistics for IMAGINE dose measure

Response category	%
Did not participate	**17.9**
Attended only a few sessions	**13.5**
Attended some sessions (less than half)	**12.0**
Attended most sessions (more than half)	**24.1**
Attended all sessions	**32.5**

% = unweighted percentage among treatment group only.

Table 29 shows model results from dose for the primary family planning use outcomes in Bangladesh. No significant dose effects were observed for any of the four primary family planning indicators. As in the primary intervention analyses, modern contraception use, and unmet need for contraception included only married, non-pregnant respondents; analysis of ever-use (i.e., lifetime use) of contraception included only married respondents who had never used contraception at baseline.

[11] These are model-implied differences adjusted for covariates and assuming a linear effect of dose on the log-hazard.

Table 29. Bangladesh primary family planning outcomes – dose-response model results

	n	OR	95% CI	*p* [a]	OR (SD scale)[b]
Ever use of contraception[c]	765	0.98	[0.85, 1.12]	.717	0.96
Current contraceptive use[d]	830	1.08	[0.95, 1.24]	.872	1.14
Modern contraceptive use[d]	830	1.05	[0.92, 1.21]	.886	1.09
Unmet need[d]	830	0.94	[0.76, 1.16]	.718	0.90

*n = weighted sample size; OR = odds ratio for effect of dose; CI = confidence interval for odds ratio; *significant at p<.05 after controlling for multiple comparisons; All indicators included in table were considered one family of comparisons*

a: p values reported are Benjamini-Hochberg adjusted values after controlling for multiple comparisons

b: odds ratio per standard deviation difference in dose

c: among married respondents, who had never used contraception at baseline (previous use of contraception was not used in the analysis model)

d: among married, non-pregnant respondents

Niger

Table 30 shows the unweighted rates of the IMAGINE Girls' Groups dose measure in Niger. The time-to-first-birth models in Niger showed no statistically significant association with participation in Girls' Groups, with a hazard ratio per point difference on the dose item of 0.99, 95% CI [0.89, 1.09], p = .793.

Table 30. Niger descriptive statistics for IMAGINE dose measure

Response category	%
Did not participate	**13.8**
Attended only a few sessions	**10.2**
Attended some sessions (less than half)	**15.6**
Attended most sessions (more than half)	**35.3**
Attended all sessions	**25.1**

% = unweighted percentage among treatment group only.

Table 31 shows model results from dose for the primary family planning use outcomes in Niger. Dose did not significantly impact the primary family planning outcomes in Niger after FDR correction. As in Bangladesh, analyses of current contraception use, modern contraception use, and unmet need for contraception included only married, non-pregnant respondents; analysis of ever-use (i.e., lifetime use) of contraception included only married respondents who had never used contraception at baseline.

Table 31. Niger primary family planning outcomes – dose-response model results

	n[a]	OR	95% CI	*p* [b]	OR (SD scale)[c]
Ever use of contraception[d]	729	1.27	[1.04, 1.53]	.076	1.39
Current contraceptive use[e]	626	1.23	[1.01, 1.50]	.088	1.34
Modern contraceptive use[e]	626	1.21	[0.99, 1.47]	.081	1.30
Unmet need[e]	626	1.16	[0.86, 1.58]	.324	1.24

*n = weighted sample size; OR = odds ratio for effect of dose; CI = confidence interval for odds ratio; *significant at p<.05 after controlling for multiple comparisons; All indicators included in table were considered one family of comparisons*
a: treatment group sample sizes differ from table 12 due to a small amount of 'do not know' responses to dose variables
b: p values reported are Benjamini-Hochberg adjusted values after controlling for multiple comparisons
c: odds ratio per standard deviation difference in dose
d: among married respondents, who had never used contraception at baseline (previous use of contraception was not used in the analysis model)
e: among married, non-pregnant respondents

SUMMARY OF FINDINGS

There were no significant differences by treatment condition on delay of time to first birth in either Bangladesh or Niger. However, findings in Bangladesh were complicated by a treatment difference in rates of marriage during the study, a difference that was not present at baseline. Among respondents unmarried at baseline, treatment-group participants were significantly *less* likely to marry between baseline and endline than control-group participants. This potential impact of the intervention on marriage timing should be considered when interpreting the delay-of-birth findings in Bangladesh. Furthermore, the dose-response analyses within the treatment areas showed a significant association between greater participation in Girls' Groups and a greater delay of time to first birth.

Despite the absence of a causal intervention impact (as intent-to-treat) based on the primary hypothesis in either country, intervention differences were apparent in other outcomes.

In Bangladesh, significant impacts of the IMAGINE intervention were observed for many of the secondary outcomes. Respondents in the treatment group showed greater improvements in family planning perceptions, reproductive health knowledge, and psychosocial outcomes (e.g., social cohesion, collective efficacy) than respondents in the control group. However, there were no intervention differences in primary family planning use outcomes, including rates of current or lifetime use of contraception, and few differences in health and economic indicators.

In Niger, primary family planning use outcomes significantly differed between the treatment and control groups, with a larger percentage of respondents in the treatment group reporting lifetime use and current use of contraception and current use of a modern method of contraception. There were fewer impacts of the intervention on secondary scale outcomes (e.g., personal agency and social expectations constructs) in Niger than in Bangladesh. However,

respondents in the treatment group in Niger reported greater health service utilization, social and economic mobility, and engagement in income-generating activities than respondents in the control group.

A limitation of the study design was that many of the respondents in the analysis were married at the time of the baseline interview but were excluded from the analysis if they gave birth before or within a few months after the start of the IMAGINE program. Thus, participants married at baseline were more likely to be excluded from the analysis if they were relatively likely to become pregnant soon after, introducing heterogeneity in the sample that was systematically linked to marital status. This is of particular concern for the model of delay of first birth, since participants who were married at the baseline interview but included in the analysis had a substantially longer minimum time between marriage and childbirth than participants who married later during the study. An additional issue for this model is that, at least in Bangladesh, participants in the treatment condition were less likely to get married during the period of the study than participants in the control condition. This suggests a systematic difference in who entered the "risk set" for first birth in Bangladesh.

A further consideration is that the study design was based on statistical power analyses for the primary delay-of-birth outcome. Other outcomes were not the primary focus for power and sample size estimation.

Lastly for the dose-response analyses, there appeared to be benefits to greater levels of participation in IMAGINE Girls' Groups in Bangladesh. Among respondents assigned to the treatment group, greater participation in groups was associated with later time to first birth. One caveat to this analysis is that data on the *timing* of the respondent's participation was not collected, and so an individuals' participation relative to their dates of marriage and/or pregnancy could not be examined.

DISCUSSION

Bangladesh

Results overall show a mixed effect in relationships to the program theory of change—demonstrating a positive impact on mobility and across a range of factors related to norms, knowledge, and self-efficacy. Additionally, there was a positive impact on income generation and mobility, which is indicative of greater economic empowerment. A significant impact of treatment was also found on ideal timing of birth after marriage, where treatment respondents desired a longer delay between marriage and first birth. However, there was no impact on ideal family size, actual delay of birth after marriage or contraceptive use (ever or current). Despite this finding, there was a trend of greater ever and current use among program participations who have not had a birth compared to non-program participants. While this effect was not statistically significant, the trend for both indicators is in the same direction.

The positive impacts on girl's agency, mobility, and economic empowerment and on beliefs related to contraceptive use and economic inclusion are indicative of shifting gender dynamics and possibly also community norms, despite no impact of the timing of birth after marriage or contraceptive use. The unexpected positive impact on the timing of marriage is also significant, in terms of being indicative of shifting community norms. The results of the qualitative evaluation also indicate shifts in gender dynamics and specifically norms related to the timing of childbearing. Men and women reported being more comfortable talking about childbearing decisions together and confronting family and community pressure to have children quickly after marriage. There was also some evidence of mothers-in-law being supportive of delayed childbearing, but this perspective was mixed, showing continued ambivalence towards delayed childbearing.

The negative impact of COVID on access to and utilization of contraception was significant in Bangladesh and could have diminished the program's impact. While some girls may have wanted to use contraception, they may not have been able to access it at all or access the more highly effective methods. Results indicate that less than 10% of respondents were visited by a community health worker in the past six months, less than one-third visited a health facility, and most contraceptive users reported using either the pill or condoms. Finally, the pandemic drove social and economic insecurity that increased rates of child marriage across the country (Hossain, 2021).

In the context of Bangladesh, results demonstrate impact on several factors that have a significant influence on timing of first birth—including reducing girls' and couples' fears and myths related to contraceptive use and attenuating familial pressure to have children quickly after marriage. However, the pandemic restricted access to contraception and increased rates of child marriage. While the results indicate shifts or disruptions in community norms related to timing of birth after marriage, it is clear that a tipping point has yet to be reached in the communities where IMAGINE was implemented. Ideally, programs like IMAGINE should be implemented over a longer period in order to continue to increase girl's agency, improve couple's communication, and address contextual issues that pressure girls and couples to have children soon after marriage (Samandari, 2020).

Niger

As in Bangladesh, results demonstrate that the program had mixed effect in relationship to the program theory of change—demonstrating positive impact on contraceptive use, positive impact on agency among and on family planning related myths, increased engagement in income generating activities limited impact on attitudes and beliefs that are highly influenced by prevailing social norms, and no impact on the timing of first birth. In terms of contraceptive method, IMAGINE participants were less likely to use traditional methods and more likely to use hormonal contraception.

Despite mixed effects overall and no impact on the primary outcome of delayed first birth, the results demonstrate that the program has promise as an approach for delaying first birth and improving birth spacing, in that program participants have more agency to make decisions about contraceptive and are more likely to use contraception. While the evaluation showed limited impact attitudes and beliefs related to family size and timing of first birth, this is not surprising. Shifts in social norms occur slowly, and therefore, a program running for two years is unlikely to show immediate impact on social norms and the behaviors that are highly influenced by those norms, especially when the expected shift is quite radical, in the context of Niger. These changes could continue emerge over time. The fact that results show impact on aspects more likely to change in the short term is potentially indicative of change that could emerge related to other 'downstream' or more 'radical' shifts, like family size preference or timing and spacing of births. That said, the results may reflect the limited potential of short-term programs with limited reach relevant to population size.

Niger remains a context where total fertility is among the highest in the world, where typically both men and women desire more children than they have, and where, among women who use contraception, traditional methods remain prevalent (Samandari, 2019; ICF, 2023). Furthermore, both husbands (Spindler, 2018) and religious leaders (Cannon, 2022) have substantial influence on whether and when men and women use contraception and thus limit and space pregnancies. Additionally, childbearing decisions are very much grounded in preferences for large families (Spindler, 2018). Results of past research on contraceptive dynamics highlights several prevailing attitudes that drive women's motivations to use contraception including to improve their health and to provide a better life for their existing children and women with these motivations are more likely to use modern, highly effective contraception (Camber Collective, 2015). Finally, results of the IMAGINE qualitative evaluation reveal that women's participation in economic activities is not seen as a reason to delay first birth.

CONCLUSION

In this context, programs like IMAGINE that are gender-synchronized and holistic (Breakthrough Action, 2020)—focused on empowering women, improving women's financial inclusion, transforming relations among women and men and engaging men and religious leaders—hold the most promise in supporting women (and men) to have healthier and more productive lives and more promising futures for their children. Ideally, to be more effective, programs should be implemented and monitored over longer periods of time, especially those targeting religious leaders and men. Finally, previous research in Niger suggests the possible utility of engaging not just girls by also boys prior to marriage and in the pre-marital courtship period (Spindler, 2018).

Future programs will benefit from the knowledge and wisdom that:

1. **Change takes time**. After two years evidence of norms and behavior change is present, but changes are different in both contexts and behaviors of delayed childbearing after marriage

have yet to emerge. Therefore, different sequencing of interventions over more time—including using VSLA as a platform—could potentially multiply the impact of future programs. Nonetheless, we should not downplay the value of incremental change – which quite clearly has happened based on the totality of the qualitative and quantitative evidence.

2. **Change requires engaging multiple actors at different levels of the socio-cultural ecology.** Active engagement with critical reflection and collective action has greater potential for change – evidence from two RCTs shows a gender light approach to have less impact (Abdiboru in Ethiopia targeting young unmarried adolescents and Win-Win in Burundi targeting small-scale farmers). Therefore, including gender synchronized household intervention in Niger and community-level collective action in Bangladesh may drive greater impact.

3. **Changes in the ecosystem strengthen and support individual-, household- and community-level changes**. In the context of the IMAGINE program, this means that the ecosystem supports and enables evolution of social norms, which in turn, generally drive behavior changes like delaying first birth or using contraception. Social norms change tends to come before behavior change, but not always – reminding us that change is complex and non-linear. Thus, behavioral changes may emerge unpredictably based on what is happening in the ecosystem and how the intervention has impacted that ecosystem. In Niger, this means that girls may be using FP more frequently and engaging in IGAs but they still feel pressure to have children early on in marriage, and thus explains why results show no evidence of impact on delay. In Bangladesh, this means that girls may have more self-efficacy and there are signs of norms shifting again these have yet to translate to greater economic empowerment, more contraceptive use, or delayed childbirth.

REFERENCES

Altman, D. G. (1985). Comparability of randomised groups. *Journal of the Royal Statistical Society, Series D, 34,* 125-136. https://doi.org/10.2307/2987510

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B, 57,* 289-300.

Camber Collective (2015). Increasing contraceptive use in Niger Final Report. Available at: Increasing Contraceptive Use in Niger: Final Report 2015 - The Compass for SBC, last accessed March 10, 2023.

Cannon AC, Mandal M, McGuire C, Calhoun LM, Mumuni T, Speizer IS. A vignette-based approach to understanding social norms around family planning in three Nigerian cities. Glob Public Health. 2022 Jul;17(7):1379-1391. doi: 10.1080/17441692.2021.1928261.

Coman, E. N., Picho, K., McArdle, J. J., Villagra, V., Dierker, L., & Iordache, E. (2013). The paired *t*-test as a simple latent change score model. *Frontiers in Psychology, 4,* 738.

Gran, J. M., Røysland, K., Wolbers, M., Didelez, V., Sterne, J. A., Ledergerber, B., Furrer, H., von Wyl, V., & Aalen, O. O. (2010). A sequential Cox approach for estimating the causal effect of treatment in the presence of time-dependent confounding applied to data from the Swiss HIV Cohort Study. *Statistics in Medicine*, *29*, 2757–2768.

Hsieh, F. Y., Lavori, P. W., Cohen, H.J., & Feussner, J. R. (2003). An overview of variance inflation factors for sample-size calculation. *Evaluation & the Health Professions*, 26(3), 239-257. doi:10.1177/0163278703255230

Hossain MJ, Soma MA, Bari MS, Emran TB, Islam MR. COVID-19 and child marriage in Bangladesh: emergency call to action. BMJ Paediatr Open. 2021 Nov 23;5(1):e001328. doi: 10.1136/bmjpo-2021-001328.

Institut National de la Statistique (INS) et ICF. 2023. Enquête sur les Indicateurs du Paludisme au Niger 2021 : Rapport final. Rockville, Maryland, USA : INS et ICF.

Little, R.J.A. & Rubin, D.B. (1987) Statistical Analysis with Missing Data. John Wiley & Sons, New York.

Mansournia, M. A., & Altman, D. G. (2016). Inverse probability weighting. *BMJ (Clinical research ed.)*, *352*, i189.

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology, 60,* 577-605.

Mutz, D. C., Pemantle, R. & Pham, P. (2019). The perils of balance testing in experimental design: Messy analyses of clean data. *The American Statistician, 73*(1), 32-42. 10.1080/00031305.2017.1322143

Rao, J. N. K., Wu, C. F. J., & Yue, K. (1992). Some recent work on resampling methods for complex surveys. *Survey Methodology, 18,* 209–217.

Robins, J. M., Hernán, M. Á., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology, 11,* 550-560.

Samandari G, Grant C, Brent L, Gullo S. "It is a thing that depends on God": barriers to delaying first birth and pursuing alternative futures among newly married adolescent girls in Niger. Reprod Health. 2019 Jul 18;16(1):109. doi: 10.1186/s12978-019-0757-y

Samandari, G., Sarker, B.K., Grant, C. et al. Understanding individual, family and community perspectives on delaying early birth among adolescent girls: findings from a formative evaluation in rural Bangladesh. BMC Women's Health 20, 169 (2020). https://doi.org/10.1186/s12905-020-01044-z

Santos, J. R. A. (1999). Cronbach's alpha: A tool for assessing the reliability of scales. *Journal of Extension, 37*(2), 1-5.

Seaman, S. R., & White, I. R. (2013). Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research*, *22*(3), 278-295.

Spindler, E., Perlman, D., Chaibou, S., Silverman, J., Carter, N., Boyce, S., Levtov, R., Vlahovicova, K., & Lauro, G. (2018). Child marriage, fertility, and family planning in Niger: Results from a study inspired by the International Men and Gender Equality Survey (IMAGES): Executive summary. Washington, DC: Promundo-US

Tsuchiya, A. (2021). Introduction to time-dependent confounders and marginal structural models. *Annals of Clinical Epidemiology, 3,* 37-45.

United Nations Children's Fund (UNICEF), Ending Child Marriage: Progress and prospects, UNICEF, New York, 2014.

West, B. T., Berglund, P., & Heeringa, S. G. (2008). A closer examination of subpopulation analysis of complex-sample survey data. *The Stata Journal, 8,* 520-531.

West, B. T., Sakshaug, J. W., & Aurelien, G. A. S. (2018). Accounting for complex sampling in survey estimation: A review of current software tools. *Journal of Official Statistics, 34,* 721-752.

APPENDIX A: MISSING DATA PATTERNS

Attrition

Retention of adolescents from baseline to endline was extremely high. In Niger, 54 (2.19%) adolescents were lost to follow-up and only 18 (0.68%) in Bangladesh. Given these small numbers, no special steps were taken to accommodate attrition, as any benefit would be likely outweighed by the added complexity.

Missing Data by Mode

About 4.6% (n=111) of respondents in Niger and 19.3% (n=504) of respondents in Bangladesh were surveyed by phone or proxy survey at endline. These surveys did not contain items for certain control variables and outcome variables that were planned for endline secondary analyses. These surveys were shortened by the field teams in conjunction with CARE to prioritize collecting the necessary variables for the primary outcomes. The teams were encouraged to prioritize face-to-face interviews to minimize data loss.

Weighting for analyses of primary outcomes is unaffected, as it is based on variables that were collected in all modes. As the study was powered for the primary outcome, this missingness does not affect power for the primary analyses of interest. ***Overall***, the current concern is therefore more about risk of bias in secondary outcomes than in power.

The consequences of missing data and available solutions depend on the mechanism by which datapoints are missing. If data are missing completely at random (MCAR; the likelihood of a datapoint being missing is unrelated to its true value), common missing-data strategies will not introduce bias, but may affect statistical power. When data are missing at random (MAR), auxiliary variables, when adjusted for, eliminate any residual association between the probability of missingness and the true value. Auxiliary variables can be included either in the model with a full-information estimator (all variables with maximum likelihood [ML], or exogenous variables with scaled weighted least squares [WLS, aka WLSMV]) or included in the imputation frame with multiple imputation [MI]. If MCAR does not hold, and variables to establish MAR are not available (or not included), then data are Missing Not At Random (MNAR), which could lead to biased results.

To assess the degree of complexity added by the missing data patterns, preliminary analyses were designed to show whether the face-to-face respondents systematically differed from proxy and phone surveys on a set of baseline characteristics. The key analyses were multinomial logistic regression models predicting endline survey mode from key demographics and baseline values of selected outcomes, using baseline weights. Predictors included intervention status, age (integer) marital status, education (ordinal), healthcare visit within previous 6 months (yes/no), family planning knowledge (continuous), income-generating activity within the previous year (yes/no), personal savings (yes/no), and household assets (yes/no), and all two-way interactions. Clustering was not used at this time due to computational issues.

The set of predictors and interactions did not significantly predict survey mode in either country; Niger: Wald χ^2 (108) = 123.58, p = .145; Bangladesh: Wald χ^2 (108) = 124.35, p = .134, suggesting that these predictors are not needed as auxiliary variables, and the data missing by mode are plausibly MCAR.

Given these results, the data are plausibly MCAR, though the possibility of data being MNAR can never be known with complete certainty. The use of full-information estimators (ML or WLSMV) in inferential models will retain the full sample and maximize power, but listwise deletion can be used for descriptive statistics without introducing bias. Conducting multiple imputation would be resource-intensive and provide minimal benefits for this type of missingness.

APPENDIX B: WEIGHTING

Even if intervention assignment itself were not empirically observed to be associated with time of marriage, this would still be a critical consideration because of potential confounds between intervention and higher-order terms involving time of marriage (e.g., if assignment is associated with the conjunction of time of marriage and age at study entry). *This issue was central to the weighting strategy given the impact of marriage on an array of SRHR outcomes.*

Weighting Strategy

An adolescent's inverse probability weight for marriage was derived by modeling the probability of marriage iteratively across windows of time from baseline to endline and taking the product of the inverses of the probabilities (Tsuchiya, 2021).

The goal was to derive analytic weights that would approximately balance respondents who, by endline, reported comparable times of being married. This strategy was optimized for balanced time of marriage rather than maintaining balance between intervention and comparison groups because time of marriage is a critical factor in predicting date of first birth (adolescents married earlier had a longer window of opportunity to give birth). This balance may also be preferable for other sexual and reproductive health (SRH) outcomes because of their relation to marriage, though if intervention assignment is not associated with time of marriage, the result for single-point endline outcomes would be approximately the same if balanced for intervention.

The products of the successive IPWs for marriage and the baseline weights at the individual level became the analytic weights for use in marginal structural models (MSMs). At the end of the weighting process, respondent-level analytic weights were normalized to a mean of 1 within country to reflect the actual sample size.[12]

Predictors

Regression predictors were selected with some slight differences between countries reflecting the different demographic data collected within country. All predictors were from baseline measurement. In both countries, the baseline components of the weighting equation included:

- Intervention assignment;
- Age at baseline (in years, modeled as continuous);
- Schooling (based on cell sizes, categorized in Niger as none, primary, and secondary or higher; categorized in Bangladesh as less than higher education vs. higher education); and
- Whether the adolescent reported any income-generating activity in the past year.

[12] Two respondents in Bangladesh who were not married by baseline did not respond as to whether they were married at endline, and therefore were not asked a date of marriage. For these two respondents, their (normalized) baseline weights were carried forward as analytic weights.

The models for Niger also included:

- Ethnicity (Hausa vs. other); and
- Whether the adolescent reported the household having any capital assets.

The models for Bangladesh also included:

- Religion (Muslim vs. Hindu); and
- Whether the adolescent reported having any personal savings.

In both countries, all two-way interactions of these predictors, including intervention assignment, were included in the propensity models.

Determination of period length

The first window was defined as "by baseline" using reports of being married on the baseline interview. Subsequent windows depended on dates reported in the endline interview. The second window was bounded by the date of the baseline interview and the initiation of intervention (March 1, 2019, in Niger and January 15, 2019, in Bangladesh) to separate marriages occurring before intervention start from those potentially affected by intervention on timing of marriage (though a study goal was to avoid such effects). Successive windows using a series of durations (30, 60, 90, 120, 180, and 365 days) were assessed for balance between using a precise timing of marriage and avoiding extreme weights (and therefore large design effects).

In practice, windows of shorter than 180 days resulted in extreme individual weights, some exceeding 100. The raw data indicated this could have been related to small cells of the number of marriages in a window when crossed with one of the first-order predictors in the logistic regression.

Using a window of 180 days, 5 of the 18 cells in Niger resulting from crossing window with ethnicity had *n*s < 10, with 3 of those yielding *n*s <= 5. Marriage in each window crossed with income-generating activity and household assets also resulted in small cells. Similar results were obtained for religion and income-generating activity in Bangladesh. When these variables and their interactions were included in the logistic regressions, the resulting analytic weights led to large individual weights and, consequently, large design effects.

Given a window of 365 days, only one cross-classification with window in Niger (of 66) and three in Bangladesh (of 60) yielded cells with fewer than 10 adolescents, all involving marrying in the shortened window between the baseline interview and the beginning of intervention. In Niger, this included 9 non-Hausa participants. In Bangladesh, this included 3 Hindu adolescents; 9 with higher education; and 5 with income-generating activities. Given the conceptual importance of retaining this boundary, 365-day windows were applied.

This resulted in 6 marriage windows in each country: married at baseline, married at country-specific intervention start, married within 1 year of intervention start, married between 1 and 2

years after intervention start, married more than 2 years after intervention start, and unmarried at endline). In practice, this was consolidated to 5 windows, as very few respondents were married more than 2 years after intervention start but at least 266 days before endline, and so that window was collapsed with that for 1 to 2 years after intervention start. The respondent-level model-implied probabilities of being married in each time window were derived from the coefficients of the corresponding logistic-linear regression.

Weighting Results

Bangladesh

In Bangladesh, weights were calculated for 2,497 adolescents. This reflected all eligible respondents who completed interviews at both baseline and endline. For some statistics below, 13 adolescents[13] who did not have a valid value for whether they were married at endline were omitted.

Summary

Analytic weights for Bangladesh, normalized to a mean of 1, ranged from 0.64 to 15.87 with a median of 0.94. Comparing estimates between intervention and comparison respondents for select variables (the weighting components, additional measures collected at baseline related to SRH and household finances, marriage at endline and age at marriage, and mode of endline data collection), imbalance was generally minimal even before utilizing weights, with the exception of the financial variables for which substantially more comparison than intervention respondents reported having personal savings and household assets at baseline. Further, marriage rates were somewhat higher among respondents in the control group. Neither baseline nor analytic weights made substantial differences in these imbalances of or balance in aggregate. Comparing estimates between respondents married and unmarried at endline showed that married respondents were much more likely to have completed the endline survey by phone versus in person. Again, neither baseline nor analytic weights yielded substantial differences on these variables. Analytic weights, being derived largely from post-baseline events, are expected to have greater impact on balance in endline measures.

Details

Table B1 shows descriptive statistics for the analytic weights, in aggregate, separately by comparison and intervention groups, and separately by marital status at endline.

[13] In addition to the two adolescents from footnote 5, eleven other respondents who reported being married at baseline but had no response at endline are not included in balance checks based on marriage at endline.

Table B1. Bangladesh Summary of Analytic Weights

Group	*N*	$\bar{x}$	*SD*	Min	Q1	Med	Q3	Max
Full sample	2,497	1.00	0.66	0.64	0.78	0.94	1.06	15.87
Control	1,208	1.00	0.51	0.71	0.80	0.96	1.06	9.39
Treatment	1,289	1.00	0.78	0.64	0.76	0.91	1.07	15.87
Unmarried	771	0.77	0.07	0.64	0.73	0.75	0.78	1.14
Married	1,713	1.11	0.78	0.73	0.92	1.01	1.10	15.87

$\bar{x}$ = mean; SD = standard deviation for continuous measures; Min= minimum weight; Q1= 25th percentile weight; Med= median weight; Q3= 75th percentile weight; Max= maximum weight

Table B2 shows summary statistics of the weighting components as well as additional measures collected at baseline related to SRH, including economic measures, and variables related to marriage and endline data collection. The first set of columns shows the unweighted descriptive statistics; the second set the statistics using weights determined at baseline; and the third the statistics using analytic weights. Only minimal differences are apparent between unweighted estimates and either set of weights with the exception of the indicator for marital status at endline. The creation of the analytic weights tended to weight individual married respondents more heavily than unmarried, as seen in Table B1, resulting in a higher estimate of marriage prevalence. Marital status and timing were among the most important considerations in the balancing scheme, to reduce the impact of differential incidence of marriage evolving during the course of the intervention. Descriptive tables in this report show simpler, sample-weighted rates of marriage and other demographics. The change in analytic-weighted marital proportion is an artifact of analytic weighting only and should not affect inferential statistics for analyses of program effects.

Table B2. Select Variables in Bangladesh by Weighting Strategy

Characteristic	Unweighted	Baseline Weights	Analytic Weights
Baseline Age	16.4	16.4	16.4
Condition: Treatment	.52	.52	.52
Survey Mode			
Face-to-face	.81	.81	.80
Proxy	.02	.02	.02
Phone	.18	.18	.19
Married at Endline	.69	.69	.76
Age at Marriage	17.0	17.0	17.0
Baseline Highest Education Level			
None/primary	.07	.07	.07
Secondary	.67	.67	.67
Higher	.26	.26	.26
Health Visit Past 6-Months	.12	.12	.12
Family Planning Knowledge	.95	.95	.95
Past-Year Income Activity	.03	.03	.03
Has Savings	.59	.59	.59
Has Assets	.14	.14	.15
Religion			
Muslim	.92	.92	.92
Hindu	.08	.08	.08

Table B3 shows the same values separately for respondents who were in intervention vs. comparison areas at baseline. The greatest differences between estimates for treatment and control were in economic variables, with substantially more treatment respondents reporting having personal savings and household assets. This was not notably altered by either weighting scheme.

Table B3. Select Variables in Bangladesh by Intervention and Weighting Strategy

	Unweighted		Baseline Weights		Analytic Weights	
Characteristic	Treatment	Control	Treatment	Control	Treatment	Control
Baseline Age	16.4	16.4	16.4	16.4	16.4	16.4
Survey Mode						
Face-to-face	.82	.80	.82	.80	.81	.78
Proxy	.01	.02	.01	.02	.01	.02
Phone	.17	.18	.17	.19	.18	.20
Married at Endline	.66	.72	.66	.72	.74	.78
Age at Marriage	17.0	17.0	17.0	17.0	17.0	17.0
Baseline Education Level						
None/primary	.07	.07	.07	.07	.06	.07
Secondary	.66	.68	.66	.68	.66	.68
Higher	.27	.25	.27	.25	.27	.25
Health Visit Past 6-Months	.14	.10	.14	.10	.14	.11
Family Planning Knowledge	.93	.97	.93	.97	.93	.97
Past-Year Income Activity	.04	.03	.04	.03	.04	.03
Has Savings	.51	.66	.52	.67	.51	.67
Has Assets	.08	.20	.08	.20	.09	.21
Religion						
Muslim	.90	.93	.90	.93	.91	.93
Hindu	.10	.07	.10	.07	.09	.07

Table B4 makes analogous comparisons for respondents who were unmarried or married at endline. The largest difference was in the proportion of respondents completing the endline survey by phone: 22% of married respondents vs. 8% of unmarried in the unweighted estimates (essentially unchanged with either set of weights). Other differences included a greater proportion of treatment participants in the unmarried group, higher levels of education among the unmarried respondents, and higher rates of asset ownership among the unmarried respondents. None of these differences was notably affected by weighting.

Table B4. Select Variables in Bangladesh by Endline Marital Status and Weighting Strategy

	Unweighted		Baseline Weights		Analytic Weights	
Characteristic	Married	Unmarried	Married	Unmarried	Married	Unmarried
Baseline Age	16.4	16.2	16.4	16.4	16.4	16.3
Condition: Treatment	.49	.56	.49	.57	.50	.58
Survey Mode						
Face-to-face	.76	.92	.76	.92	.76	.92
Proxy	.01	.01	.01	.01	.01	.01
Phone	.22	.08	.23	.08	.22	.08
Age at Marriage	17.0	N/A	17.0	17.0	17.0	N/A
Baseline Education Level						
None/primary	.08	.05	.08	.04	.07	.05
Secondary	.68	.64	.68	.64	.68	.65
Higher	.24	.31	.24	.31	.25	.30
Health Visit Past 6-Months	.11	.13	.11	.14	.12	.13
Family Planning Knowledge	.94	.96	.94	.96	.95	.96
Past-Year Income Activity	.03	.04	.03	.03	.03	.04
Has Savings	.58	.61	.58	.62	.58	.60
Has Assets	.16	.10	.16	.10	.16	.10
Religion						
Muslim	.93	.89	.93	.89	.93	.89
Hindu	.07	.11	.07	.11	.07	.11

Niger

In Niger, weights were calculated for 2,123 adolescents. This reflected all eligible respondents who completed interviews at both baseline and endline.

Summary

Analytic weights for Niger, normalized to a mean of 1, ranged from 0.64 to 7.89 with a median of 0.89. Comparing estimates between intervention and comparison respondents for select variables (the weighting components, additional measures collected at baseline related to SRH and household finances, marriage at endline and age at marriage, and mode of endline data collection), imbalance before applying weights was low in most variables being compared *other* than education, personal income before baseline, and ethnicity. Baseline weights compensated for the difference in education, but neither set of weights substantially mitigated the other

differences. Comparing estimates between adolescents married and unmarried at endline, married respondents were much more likely to have completed the endline survey by phone rather than face-to-face. Neither set of weights notably affected this result. Analytic weights, being derived largely from post-baseline events, are expected to have greater impact on balance in endline measures

Details

Table B5 shows descriptive statistics for the analytic weights, in aggregate, separately by comparison and intervention groups, and separately by marital status at endline.

Table B5. Niger Summary of Analytic Weights

Group	*N*	$\bar{x}$	*SD*	Min	Q1	Med	Q3	Max
Full sample	2,123	1.00	0.51	0.64	0.75	0.89	1.07	7.89
Control	1,034	1.01	0.56	0.64	0.76	0.88	1.07	7.89
Treatment	1,089	0.99	0.46	0.65	0.74	0.92	1.08	5.80
Unmarried	724	0.76	0.13	0.64	0.68	0.72	0.78	2.44
Married	1,399	1.13	0.59	0.68	0.88	0.99	1.16	7.89

$\bar{x}$ = mean; SD = standard deviation for continuous measures; Min= minimum weight; Q1= 25th percentile weight; Med= median weight; Q3= 75th percentile weight; Max= maximum weight

Table B6 shows summary statistics of the weighting components as well as additional measures collected at baseline related to SRH, including economic measures. The first set of columns shows the unweighted descriptive statistics; the second set the statistics using weights determined at baseline; and the third the statistics using analytic weights. Only minimal differences are apparent between unweighted estimates and either set of weights with the exception of the indicator for marital status at endline. The creation of the analytic weights tended to weight married respondents more heavily than unmarried, as shown in Table B5, resulting in a higher estimate of marriage prevalence. This artifact of the weighting is unlikely to affect inferential statistics for analyses of program effects.

Table B6. Select Variables in Niger by Weighting Strategy

Characteristic	Unweighted	Baseline Weights	Analytic Weights
Baseline Age	15.7	15.7	15.7
Condition: Treatment	.51	.50	.51
Survey Mode			
Face-to-face	.96	.95	.95
Proxy	.03	.02	.03
Phone	.02	.02	.02
Married at Endline	.66	.64	.74
Age at Marriage	16.6	16.6	16.7
Baseline Highest Education Level			
No Schooling	.38	.38	.36
Primary	.24	.24	.24
Secondary/Higher	.38	.39	.40
Health Visit Past 6-Months	.09	.10	.09
Family Planning Knowledge	.64	.64	.63
Past-Year Income Activity	.28	.28	.27
Has Savings	.09	.08	.09
Has Assets	.19	.18	.18
Ethnicity			
Hausa	.92	.92	.92
Other ethnicity	.08	.08	.08

Table B7 shows the same values separately for respondents who were in intervention vs. comparison areas at baseline. The greatest differences between estimates for treatment and control were in education, income-generating activity, and ethnicity, where participants in the treatment group were substantially less likely to have had an income-generating activity in the year before baseline, likely to have higher educational attainment, and more likely to have been of an ethnicity other than Hausa. Baseline weights resulted in balance on educational attainment, but analytic weights did not and neither set compensated substantially for the other differences.

Table B7. Select Variables in Niger by Intervention and Weighting Strategy

	Unweighted		Baseline Weights		Analytic Weights	
Characteristic	Treatment	Control	Treatment	Control	Treatment	Control
Baseline Age	15.7	15.7	15.7	15.7	15.7	15.7
Survey Mode						
Face-to-face	.96	.95	.96	.95	.96	.94
Proxy	.02	.03	.02	.03	.03	.04
Phone	.02	.02	.02	.02	.02	.02
Married at Endline	.65	.67	.64	.64	.73	.75
Age at Marriage	16.6	16.6	16.7	16.6	16.7	16.7
Baseline Education Level						
No Schooling	.36	.41	.38	.38	.34	.38
Primary	.24	.24	.24	.23	.24	.23
Secondary/Higher	.40	.35	.38	.39	.41	.38
Health Visit Past 6-Months	.12	.07	.12	.08	.12	.07
Family Planning Knowledge	.64	.63	.64	.64	.64	.62
Past-Year Income Activity	.23	.33	.25	.32	.24	.31
Has Savings	.09	.08	.08	.09	.09	.08
Has Assets	.20	.17	.20	.16	.19	.17
Ethnicity						
Hausa	.89	.95	.90	.95	.90	.95
Other	.11	.05	.10	.05	.10	.05

Table B8 makes analogous comparisons for respondents who were unmarried or married at endline. The largest difference was in the proportion of respondents completing the endline survey by phone: 22% of married respondents vs. 8% of unmarried in the unweighted estimates and essentially unchanged with either set of weights. Other differences included more treatment participants in the unmarried group, higher levels of education among the unmarried respondents, and higher rates of asset ownership among the unmarried respondents. None of these differences was notably affected by weighting.

Table B8. Select Variables in Niger by Endline Marital Status and Weighting Strategy

	Unweighted		Baseline Weights		Analytic Weights	
Characteristic	Married	Unmarried	Married	Unmarried	Married	Unmarried
Baseline Age	15.8	15.6	15.8	15.6	15.7	15.6
Condition: Treatment	.51	.53	.51	.50	.50	.52
Survey Mode						
Face-to-face	.95	.97	.95	.96	.94	.97
Proxy	.03	.01	.03	.01	.04	.01
Phone	.02	.02	.02	.02	.02	.02
Age at Marriage	16.6	N/A	16.6	N/A	16.7	N/A
Baseline Education Level						
No Schooling	.45	.26	.45	.25	.39	.28
Primary	.24	.24	.24	.24	.24	.24
Secondary/Higher	.31	.51	.32	.51	.37	.48
Health Visit Past 6-Months	.10	.08	.11	.09	.10	.08
Family Planning Knowledge	.62	.67	.63	.66	.62	.66
Past-Year Income Activity	.29	.25	.30	.25	.28	.26
Has Savings	.09	.07	.09	.07	.09	.07
Has Assets	.21	.13	.21	.13	.20	.14
Ethnicity						
Hausa	.91	.93	.92	.94	.92	.93
Other	.09	.07	.08	.06	.08	.07

APPENDIX C: IMPUTATION OF DATES

Dates of marriage were algorithmically imputed for use in the analyses of time to first birth. These dates were required to assess entry into the "risk set" for giving birth.

- If a respondent was married at baseline but no date of marriage was provided, the date of marriage for purposes of the analysis was set to the median date of marriage across other respondents reporting married at baseline, separately by age at baseline.
- If a respondent married during the study (i.e., was unmarried at baseline and married at endline), date of marriage was imputed to a random date, depending on whether a date of childbirth was reported:
    - If the respondent had given birth by endline, a random date was drawn from a uniform distribution between their date of baseline interview and 266 days (38 weeks) before the reported date of birth.
    - If the respondent had not given birth, a random date was drawn from a uniform distribution between their date of baseline interview and their date of endline interview. The following SAS DATA-step code implements the imputation. Italicized statements are comments.

```
*Initializes the random number generator for replicability;
  call streaminit(6229);

* Imputing date of marriage;
* SAS dates (e.g., "21837") are calculated as the number of days since January 1, 1960,
so that they can be handled numerically;

**** Niger;

*If married at baseline (after consistency checks) and does not have a date of marriage
at endline, complete the following do loop;
  if b_married_imp_ng_t1 = 1 and has_dom_ng_t2 = 0 then do;

*Dates generated from median of other respondents by baseline age;
    if b_age_ng_t1 = 15 then b_date_of_marriage_imp_ng_t2 = 21837;
    if b_age_ng_t1 = 16 then b_date_of_marriage_imp_ng_t2 = 21564;
    if b_age_ng_t1 = 17 then b_date_of_marriage_imp_ng_t2 = 21533;
    if b_age_ng_t1 = 18 then b_date_of_marriage_imp_ng_t2 = 21472;
    if b_age_ng_t1 = 19 then b_date_of_marriage_imp_ng_t2 = 21363;
    end;

*If date of marriage was not missing, complete the variable with the original value;
        else b_date_of_marriage_imp_ng_t2 = b_date_of_marriage_ng_t2;

*If not married at baseline (after consistency checks) but married at endline (married
during the study), and does not have a date of marriage, complete the following do loops;

*If did not give birth (has_dob = 0 [false]);
  if b_married_imp_ng_t1 = 0 and b_married_imp_ng_t2 = 1 and has_dom_ng_t2 = 0
    and has_dob_ng_t2 = 0 then do;

*Establish length of time between baseline and endline;
    studylength_ng_t2 = a_interview_date_ng_t2 - a_interview_date_imp_ng_t1;

*Generate a random number between 1 and that length;
    marrieddate_ng_t2 = rand('uniform', 1, studylength_ng_t2);

*Add that random number to the date of baseline to impute a date of marriage;
    b_date_of_marriage_imp_ng_t2 = a_interview_date_imp_ng_t1 + marrieddate_ng_t2;
```

```
    end;

*If did give birth (has_dob = 1 [true]);

  if b_married_imp_ng_t1 = 0 and b_married_imp_ng_t2 = 1 and has_dom_ng_t2 = 0
    and has_dob_ng_t2 = 1 then do;

*Establish length of time between baseline and likely date of conception;
    studylength_ng_t2 = (b_child_dob_ng_t2 - 266) - a_interview_date_imp_ng_t1;

*Generate a random number between 1 and that length;
    marrieddate_ng_t2 = rand('uniform', 1, studylength_ng_t2);

*Add that random number to the date of baseline to impute a date of marriage;
    b_date_of_marriage_imp_ng_t2 = a_interview_date_imp_ng_t1 + marrieddate_ng_t2;
    end;

**** Bangladesh;

*If married at baseline (after consistency checks) and does not have a date of marriage
at endline, complete the following do loop;
  if b_married_imp_bgd_t1 = 1 and has_dom_bgd_t2 = 0 then do;

*Dates generated from median of other respondents by baseline age;
    if b_age_bgd_t1 = 15 then b_date_of_marriage_imp_bgd_t2 = 21797;
    if b_age_bgd_t1 = 16 then b_date_of_marriage_imp_bgd_t2 = 21716;
    if b_age_bgd_t1 = 17 then b_date_of_marriage_imp_bgd_t2 = 21621;
    if b_age_bgd_t1 = 18 then b_date_of_marriage_imp_bgd_t2 = 21451;
    if b_age_bgd_t1 = 19 then b_date_of_marriage_imp_bgd_t2 = 21568;
    end;

*If date of marriage was not missing, complete the variable with the original value;
        else b_date_of_marriage_imp_bgd_t2 = b_date_of_marriage_bgd_t2;

*If not married at baseline (after consistency checks) but married at endline (married
during the study), and does not have a date of marriage, complete the following do loops;

*If did not give birth (has_dob = 0 [false]);
  if b_married_imp_bgd_t1 = 0 and b_married_imp_bgd_t2 = 1 and has_dom_bgd_t2 = 0
    and has_dob_bgd_t2 = 0 then do;

*Establish length of time between baseline and endline;
    studylength_bgd_t2 = a_interview_date_bgd_t2 - a_interview_date_imp_bgd_t1;

*Generate a random number between 1 and that length;
    marrieddate_bgd_t2 = rand('uniform', 1, studylength_bgd_t2);

*Add that random number to the date of baseline to impute a date of marriage;
    b_date_of_marriage_imp_bgd_t2 = a_interview_date_imp_bgd_t1 + marrieddate_bgd_t2;
    end;

*If did give birth (has_dob = 1 [true]);

  if b_married_imp_bgd_t1 = 0 and b_married_imp_bgd_t2 = 1 and has_dom_bgd_t2 = 0
    and has_dob_bgd_t2 = 1 then do;

*Establish length of time between baseline and likely date of conception;
    studylength_bgd_t2 = (b_child_dob_bgd_t2 - 266) - a_interview_date_imp_bgd_t1;

*Generate a random number between 1 and that length;
    marrieddate_bgd_t2 = rand('uniform', 1, studylength_bgd_t2);

*Add that random number to the date of baseline to impute a date of marriage;
    b_date_of_marriage_imp_bgd_t2 = a_interview_date_imp_bgd_t1 + marrieddate_bgd_t2;
    end;
```

APPENDIX D: SUPPLEMENTARY TABLES

Scale Reliability Analysis

Twelve selected priority scales and indices were selected for reliability analysis[14]. All scales were constructed per CARE guidelines (see Appendix E). Alphas of 0.70 or higher suggest good internal consistency scale reliability (Santos, 1999).

Bangladesh

Table D1 shows scale weighted means and coefficient alpha for selected scales in Bangladesh. Generally, analyses revealed high reliability among the scales. Coefficient alpha statistics of 0.70 or higher were observed in eight analyzed scales. Coefficient alpha values of less than 0.70 were only observed in the self-efficacy to discuss family planning and non-financial household decisions scales, both of which were only administered to married respondents in Bangladesh. Table D2 shows scale breakdown by treatment group. Overall, scale reliability did not differ between treatment and control.

Table D1. Bangladesh Primary Sample Weighted Means and coefficient alphas of Priority Scales – Combined Treatment & Control Endline Data

	Primary sample			
Variables	*N*	$\bar{x}$ *or %*	*SD*	*α*
Priority scales full sample	1985			
Early pregnancy risk knowledge[a]		3.20	0.88	
Belief in family planning myths[b]		2.39	0.74	0.85
SE to go to a health facility[b]		3.48	1.03	0.79
SE to engage in economic activity[b]		4.15	0.74	0.73
Social Cohesion[b]		3.91	0.59	0.82
Collective efficacy[b]		3.83	0.97	0.90
Mobility[c]		2.71	0.33	0.87
Total assets[a]		1.06	0.80	
Priority Scales for married only	1445			
SE to discuss family planning[b]		4.20	0.73	0.51
SE to refuse sex[b]		3.53	1.23	0.86
Non-financial household decisions[d]		1.61	0.26	0.57
Financial household decisions[d]		1.35	0.35	0.85

$\bar{x}$ = mean; SD = standard deviation;
SE=Self-efficacy α = coefficient alpha
[a] 4=high 0=none
[b] 5=high 1=low
[c] 3=high 1=low
[d] 2=high 1=low

[14] Reliability is not reported for the early pregnancy risk knowledge and mobility measures. These measures are considered indices, consisting of multiple items that cumulatively measure a construct, yet the individual items are not expected to be highly correlated.

Table D2. Bangladesh Primary Sample Weighted Means and coefficient alphas of Priority Scales and Indexes – by Treatment & Control Endline Data

	Treatment				Control			
Variables	*N*	$\bar{x}$ or %	*SD*	α	*N*	$\bar{x}$ or %	*SD*	α
Priority scales full sample	1044				940			
Early pregnancy risk knowledge[a]		3.28	0.80			3.11	0.95	
Belief in family planning myths[b]		2.22	0.70	0.83		2.58	0.73	0.84
SE to go to a health facility[b]		3.47	1.03	0.81		3.49	1.03	0.79
SE to engage in economic activity[b]		4.26	0.70	0.73		4.04	0.77	0.73
Social Cohesion[b]		3.86	0.62	0.85		3.96	0.55	0.78
Collective efficacy[b]		3.89	0.93	0.90		3.77	1.00	0.91
Mobility[c]		2.75	0.32	0.88		2.66	0.35	0.86
Total assets[a]		1.09	0.82			1.02	0.78	
Priority Scales for married only	739				706			
SE to discuss family planning[b]		4.27	0.75	0.52		4.14	0.71	0.49
SE to refuse sex[b]		3.62	1.22	0.86		3.50	1.22	0.87
Non-financial household decisions[d]		1.62	0.27	0.59		1.59	0.25	0.54
Financial household decisions[d]		1.38	0.37	0.86		1.30	0.32	0.84

**Significantly different from control at p<.05; $\bar{x}$ = mean; CI = confidence interval*
SE=Self-efficacy α = coefficient alpha
[a] 4=high 0=none
[b] 5=high 1=low
[c] 3=high 1=low
[d] 2=high 1=low

Niger

Scales were constructed in Niger in the same way as in Bangladesh. Tables D3 and D4 show Niger overall scale summary and reliability statistics and comparisons by treatment group, respectively. Endline scale reliability in Niger was comparable to Bangladesh for most scales. All overall reliability analyses shown in Table D3 indicate relatively reliable scales with coefficient alpha statistics of 0.70 or higher.

Table D3. Niger Primary Sample Weighted Means and coefficient alphas of Priority Scales – Combined Treatment & Control Endline Data

	Primary sample			
Variables	*N*	$\bar{x}$ or %	*SD*	α
Priority scales full sample	2014			
Early pregnancy risk knowledge[a]		2.88	1.14	
Belief in family planning myths[b]		2.69	0.83	0.91
SE to go to a health facility[b]		3.00	1.10	0.80
SE to engage in economic activity[b]		3.79	0.78	0.72
Social Cohesion[b]		3.69	0.69	0.82
Collective efficacy[b]		3.18	1.20	0.92
Mobility[c]		2.74	0.35	0.78
SE to discuss family planning[b]		2.99	1.16	0.76
SE to refuse sex[b]		1.86	1.06	0.90
Total assets[a]		1.49	1.29	
Priority Scales for married only	1485			
Non-financial household decisions[d]		1.40	0.36	0.75
Financial household decisions[d]		1.32	0.32	0.83

**Significantly different from control at p<.05; $\bar{x}$ = mean; CI = confidence interval*
SE=Self-efficacy α = coefficient alpha
[a] 4=high 0=none
[b] 5=high 1=low
[c] 3=high 1=low
[d] 2=high 1=low

Table D4. Niger Primary Sample Weighted Means and coefficient alphas of Priority Scales and Indexes – by Treatment & Control Endline Data

	Treatment				Control			
Variables	*N*	$\bar{x}$ or %	*SD*	α	*N*	$\bar{x}$ or %	*SD*	α
Priority scales full sample	1029				985			
Early pregnancy risk knowledge[a]		2.92	1.11			2.83	1.17	
Belief in family planning myths[b]		2.48	0.84	0.91		2.92	0.76	0.90
SE to go to a health facility[b]		3.16	1.08	0.79		2.82	1.11	0.80
SE to engage in economic activity[b]		3.88	0.78	0.72		3.70	0.77	0.66
Social Cohesion[b]		3.68	0.67	0.81		3.71	0.71	0.84
Collective efficacy[b]		3.35	1.11	0.89		3.00	1.27	0.93
Mobility[c]		2.77	0.34	0.78		2.72	0.37	0.78
SE to discuss family planning[b]		3.15	1.12	0.75		2.82	1.17	0.77
SE to refuse sex[b]		1.99	1.11	0.89		1.71	0.98	0.90
Total assets[a]		1.78	1.27			1.19	1.24	
Priority Scales for married only	751				734			
Non-financial household decisions[d]		1.45	0.35	0.73		1.34	0.36	0.76
Financial household decisions[d]		1.35	0.32	0.82		1.28	0.31	0.83

**Significantly different from control at p<.05; $\bar{x}$ = mean; CI = confidence interval*
SE=Self-efficacy α = coefficient alpha
[a] 4=high 0=none
[b] 5=high 1=low
[c] 3=high 1=low
[d] 2=high 1=low

Demographic Analyses

Table D5. Bangladesh Unweighted Sample Descriptive Statistics by Condition

	Treatment (N=1289)		Control (N=1208)	
Variable	$\bar{x}$ or %	SD	$\bar{x}$ or %	SD
Endline Response Mode				
Face-to-face	81.7%		79.5%	
Proxy	1.4%		2.0%	
Phone	16.9%		18.5%	
Age	19.27	1.46	19.35	1.37
Married	66.0%		72.1%	
Age at marriage	16.96[1]	1.98	17.04[2]	1.86
Highest Education				
None/Primary	5.0%[3]		5.7%[4]	
Secondary	43.7%		43.8%	
Higher	51.3%		50.5%	
Bengali Ethnicity	100%		100%	
Religious affiliation (Endline)				
Muslim	91.2%[3]		94.7%[4]	
Hindu	8.8%		5.3%	

$\bar{x}$ = mean; SD = standard deviation for continuous measures; 1 treatment group age at marriage N=844; 2 control group age at marriage N=859; 3 treatment group endline education and religious affiliation N=1053; 4 control group endline education and religious affiliation N=960;

Table D6. Niger Unweighted Sample Descriptive Statistics by Condition

	Treatment (N=1089)		Control (N=1034)	
Variable	$\bar{x}$ *or %*	SD	$\bar{x}$ *or %*	SD
Endline Response Mode				
Face-to-face	96.3%		94.6%	
Proxy	1.9%		3.3%	
Phone	1.7%		2.1%	
Age	18.6	1.48	18.5	1.58
Married	64.8%		67.0%	
Age at marriage	16.6[1]	1.61	16.6[2]	1.73
Highest Education				
No School	39.4%[3]		37.5%[4]	
Primary	17.4%		24.3%	
Secondary/Higher	43.2%		38.1%	
Ethnicity				
Hausa	89.2%*		95.0%	
Other Ethnicity	10.8%		5.0%	

**Significantly different from control at p<.05; $\bar{x}$ = mean; SD = standard deviation for continuous measures; 1 treatment group age at marriage N=706; 2 control group age at marriage N=693; 3 treatment group endline education N=1049; 4 control group endline education N=978;*

APPENDIX E: SCALES AND MEASURES

Scale Title	Number of Items	Source Adapted From	Analyzed at Baseline	Notes
Early Pregnancy Risk Knowledge	4		Yes	
Belief in FP myths and misconceptions	10	These came from: https://www.guttmacher.org/journals/ipsrh/2015/12/belief-family-planning-myths-individual-and-community-levels-and-modern; https://www.guttmacher.org/journals/ipsrh/2001/06/why-nigerian-adolescents-seek-abortion-rather-contraception-evidence-focus; https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-015-1483-1	Yes	
Self-efficacy to discuss and use FP	4	https://careinternational-my.sharepoint.com/:b:/g/personal/anne_laterra_care_org/EXF77BNPh4xCjGJuBDq5WOsBK2s5isn-3zVFHTnpHzquGg?e=Heea8H	Yes	
Self-efficacy to refuse sex	5	https://careinternational-my.sharepoint.com/:b:/g/personal/anne_laterra_care_org/EXF77BNPh4xCjGJuBDq5WOsBK2s5isn-3zVFHTnpHzquGg?e=Heea8H	Yes	
Self-efficacy to go to health facility	5	https://careinternational-my.sharepoint.com/:b:/g/personal/anne_laterra_care_org/EXF77BNPh4xCjGJuBDq5WOsBK2s5isn-3zVFHTnpHzquGg?e=Heea8H	Yes	
Self-efficacy to engage in economic activities	6	https://careinternational-my.sharepoint.com/:b:/g/personal/anne_laterra_care_org/EXF77BNPh4xCjGJuBDq5WOsBK2s5isn-3zVFHTnpHzquGg?e=Heea8H	Yes	

Social Cohesion	**8**	https://careinternational-my.sharepoint.com/:b:/g/personal/anne_laterra_care_org/EXF77BNPh4xCjGJuBDq5WOsBK2s5isn-3zVFHTnpHzquGg?e=Heea8H	Yes	
Collective Efficacy	**5**	https://careinternational-my.sharepoint.com/:b:/g/personal/anne_laterra_care_org/EXF77BNPh4xCjGJuBDq5WOsBK2s5isn-3zVFHTnpHzquGg?e=Heea8H	Yes	
Participation in General Household Decision Making scale	**4**	https://careinternational-my.sharepoint.com/:b:/g/personal/anne_laterra_care_org/EXF77BNPh4xCjGJuBDq5WOsBK2s5isn-3zVFHTnpHzquGg?e=Heea8H [adapted to only non-financial decision making questions]	Yes	
Mobility scale	**9**	Adapted from https://careinternational-my.sharepoint.com/:b:/g/personal/anne_laterra_care_org/EXF77BNPh4xCjGJuBDq5WOsBK2s5isn-3zVFHTnpHzquGg?e=Heea8H	Yes	
Ownership of Household Assets / Resources	**4**	https://careinternational-my.sharepoint.com/:b:/g/personal/anne_laterra_care_org/EXF77BNPh4xCjGJuBDq5WOsBK2s5isn-3zVFHTnpHzquGg?e=Heea8H	Yes	
Participation in household financial decision-making scale	**8**	https://careinternational-my.sharepoint.com/:b:/g/personal/anne_laterra_care_org/EXF77BNPh4xCjGJuBDq5WOsBK2s5isn-3zVFHTnpHzquGg?e=Heea8H [adapted to only financial decision making questions]	Yes	
Rights Based FP	**6**	Adapted from Boydell et al. Rights-based family planning service delivery index instruments: https://www.dropbox.com/s/fls6r2znzlthbi3/RBFP%20SD%20I	No	*applies only to current FP users at time of data collection

		ndex%20Instrument%20Overview%20Final.pdf?dl=0		
Quality of interaction with FP service Provider	**13**	Adapted from Boydell et al. Rights-based family planning service delivery index instruments: https://www.dropbox.com/s/fls6r2znzlthbi3/RBFP%20SD%20Index%20Instrument%20Overview%20Final.pdf?dl=0	No	*applies only to those who reported visiting a health facility for care for themselves in 6 months proceeding the survey
Frequency of Interspousal Communication	**6**		No	
Normative expectations about girls' role	**10**	Adapted from Tipping Point	No	
Outcome expectations about FP use	**2**	Developed from formative research	No	
Outcome expectations about delaying childbirth	**4**	Developed from formative research	No	
Outcome expectations Engaging in income generating activities	**3**	Developed from formative research	No	

Early Pregnancy Risk Knowledge Index

1. A woman can get pregnant on the very first time she has sexual intercourse
2. A woman is most likely to get pregnant if she has sexual intercourse half-way between her period
3. From one menstrual period to the next, are there certain days when a woman is more likely to become pregnant?
4. After the birth of a child, a woman can become pregnant before her menstrual period has returned

Item response options: True, False or Don't Know, where 'True' = 1 and 'Don't Know' or 'False' = 0

Index constructed by summing the item scores. The Index score range is 0-4. A higher index score indicates a higher knowledge.

Family Planning Myths and Misconceptions Scale

1. Contraceptives cause disabilities/birth defects in children
2. Using contraceptives before first birth can cause infertility
3. Using contraceptives at any point can cause infertility
4. An IUD/Copper-T can travel inside a young woman's body to her heart or her brain
5. Contraceptives diminish sex drive
6. Contraceptives can harm your womb
7. People who use contraceptives end up with health problems
8. Contraceptives are dangerous to a women's health
9. Using contraceptives causes dirty blood to stay inside the body
10. Using contraceptives will make a woman bleed all the time

Item response options: 5 point likert scale, where strongly agree = 5, agree = 4, neither agree / disagree = 3, disagree = 2, strongly disagree =1

Scale constructed by summing the item scores and dividing by the number of items. The scale score range is 1-5, a higher scale score indicates a higher belief in common myths and misconceptions related to family planning use.

Self-efficacy to Discuss and Use Family Planning Scale

1. How sure are you that you could bring up the topic of family planning with your husband?
2. How sure are you that you could tell your husband that you wanted to use family planning?
3. How sure are you that you could use family planning, even if your husband did not want to?
4. How sure are you that you could use family planning, even if your mother-in-law did not want to?

Item response options: 5-point Likert scale, where Completely Sure = 5, Somewhat Sure = 4, Neither Sure/Unsure = 3, Somewhat Unsure = 2, and Not at all Sure = 1

The scale was constructed by summing the item scores and dividing by the number of items.

The scale score range is 1-5, and a higher scale score indicates higher self-efficacy to discuss and use family planning.

Self-efficacy to Refuse Sex Scale

1. How sure are you that you could refuse to have sex with your husband when you don't want to have sex but he does?
2. How sure are you that you could refuse to have sex with your husband if you were feeling tired?
3. How sure are you that you could refuse to have sex with your husband if he gets angry with you if you don't have sex?
4. How sure are you that you could refuse to have sex with your husband if he threatens to hurt you if you won't have sex?
5. How sure are you that you could refuse to have sex with your husband if he threatens to have sex with other women if you don't have sex with him?

Item response options: 5-point Likert scale, where Completely Sure = 5, Somewhat Sure = 4, Neither Sure/Unsure = 3, Somewhat Unsure = 2, and Not at all Sure = 1
The scale was constructed by summing the item scores and dividing by the number of items. The scale score range is 1-5, and a higher scale score indicates higher self-efficacy to refuse sex.

Self-efficacy to Go to the Health Facility Scale

1 How sure are you that you could go to the health facility if you were worried that the staff would treat you badly?
2. How sure are you that you could go to the health facility if your husband objected to your going?
3. How sure are you that you could go to the health facility if you feel you have some work to do at home?
4. How sure are you that you could go to the health facility if your mother-in-law objected to your going?
5. How sure are you that you could go to the health facility if your family thought you were neglecting your household duties?
Item response options: 5-point Likert scale, where Completely Sure = 5, Somewhat Sure = 4, Neither Sure/Unsure = 3, Somewhat Unsure = 2, and Not at all Sure = 1
The scale was constructed by summing the item scores and dividing by the number of items. The scale score range is 1-5, and a higher scale score indicates higher self-efficacy to go to the health facility.

Self-efficacy to Engage in Economic Activities Scale

1. How sure are you that you could participate in income generating activity if you wanted to?
2. How sure are you that you could participate in an income generating activity if your family/husband objected to you doing so?
3. How sure are you that you could participate in an income generating activity if your family/husband would not help with your other household duties so that you could do so?
4. How sure are you that you could work with your family / husband to create a household budget?
5. If you went to the market, how sure are you that you could negotiate fair prices for things you might want to buy ?
6. If you went to the market, how sure are you that you could negotiate fair prices for things you might want to sell ?

Item response options: 5-point Likert scale, where Completely Sure = 5, Somewhat Sure = 4, Neither Sure/Unsure = 3, Somewhat Unsure = 2, and Not at all Sure = 1
The scale was constructed by summing the item scores and dividing by the number of items. The scale score range is 1-5, and a higher scale score indicates higher self-efficacy to engage in economic activities

Social Cohesion Scale

1. The majority of people in this community can be trusted
2. The majority of people in this community generally get along with each other
3. I feel that I am really a part of this community
4. I can rely on people in my community if I need to borrow money.
5. I can rely on people in my community if I need to talk about my problems.
6. I can rely on people in my community to help deal with a violent or difficult family member
7. I can rely on people in my community to help take care of my household if I need to go to the doctor or hospital.
8. The people in my community are an integrated group

Item response options: 5-point Likert scale, where Strongly Agree = 5, Agree = 4, Neither Agree Nor Disagree = 3, Disagree = 2, and Strongly Disagree = 1

The scale was constructed by summing the item scores and dividing by the number of items. The scale score range is 1-5, and a higher scale score indicates higher social cohesion.

Collective Efficacy Scale

1. How sure are you that girls in your community could work together to prevent each other from being beaten or injured by family members?
2. How sure are you that girls in your community could work together to improve how adolescents are treated at the health facility?
3. How sure are you that girls in your community could work together to prevent each other from being married too young?
4. How sure are you that girls in your community could work together to get government services you need?
5. How sure are you that girls in your community could work together to improve the health and well-being of girls in your community?

Item response options: 5-point Likert scale, where Completely Sure = 5, Somewhat Sure = 4, Neither Sure/Unsure = 3, Somewhat Unsure = 2, and Not at all Sure = 1

The scale was constructed by summing the item scores and dividing by the number of items. The scale score range is 1-5, and a higher scale score indicates higher collective efficacy.

General (non-financial) Household Decision Making Scale

1. Which member of your household usually makes decisions about your healthcare?
2. Which member of your household usually makes decisions about when you will visit family/ relatives / friends?
3. Which member of your household usually makes decisions about when your whole household will visit family/relatives/friends?
4. Which member of your household usually makes decisions about when you and your husband have sex?
5. Which member of your household usually makes decisions about whether you and your husband use family planning?

Item response options: Her, Her husband, Her and her husband together, Mother-in-law, Father-in-law, Mother / Father, someone else, where Her or her and husband together =2 and all other responses =1.

The scale is constructed by summing the item scores and dividing by the number of items. The scale score range is 1-2, and a higher scale score indicates more equitable decision-making in the household.

Mobility Scale

"Please tell me whether you are permitted to go to the following places on your own, only if someone accompanies you, or not at all."

1. To go to the market to buy or sell things
2. To go fetch water
3. To go to training courses (ex. Literary courses, VSLA training, etc.)
4. To go to the health facility
5. To go to a community meeting
6. To go to friends in same village
7. To go to friends outside the village / in another village
8. To go to mosque
9. To go to a girls group meeting

Item response options: Permitted on her own, permitted if accompanied, not permitted at all, where on her own = 3, if accompanied = 2, and not permitted = 1.

The scale is constructed by summing the item scores and dividing by the number of items.

The scale score ranges is 1-3, and a higher score indicates a higher level of mobility.

Ownership of Household Assets / Resources Index

1. Aside from your household chores and work, did you do any work outside the home in the past 6 months for which you received money?
2. Aside from your household chores and work, did you do any work outside the home in the past 6 months for which you were paid in goods?
3. Do you have any money/ cash savings of your own, including in a VSLA group?
4. Do you own any assets that could help you generate income? I mean assets (ex: goats, chickens, grain) that you could sell if you needed money.

Item response options: Yes or No, where Yes = 1 and No = 0

The index was constructed by summing the item scores.

The index score range is 0-5, and a higher index score indicates more control over assets.

Participation in Household Financial Decision-Making Scale

1. Which member of your household usually makes decisions about making large household purchases?
2. Which member of your household usually makes decisions about making household purchases for daily needs?
3. Which member of your household usually makes decisions about how to use the money that you bring into the household?
4. Which member of your household usually makes decisions about how to use the money your husband brings into the household?
5. Which member of your household usually makes decisions about what to do with large assets (like a cow) or financial savings?
6. Which member of your household usually makes decisions about when your family will sell a small asset (like a chicken)?
7. Which member of your household usually makes decisions about taking out a loan or opening a savings account?
8. Which member of your household usually makes decisions about whether you can work to earn money?

Item response options: Her, Her husband, Her and her husband together, Mother-in-law, Father-in-law, Mother / Father, someone else, where Her or her and husband together =2 and all other responses =1.

The scale is constructed by summing the item scores and dividing by the number of items.

The scale score range is 1-2, and a higher scale score indicates more equitable decision-making in the household.

Rights Based FP Index

This measure applies only to current FP users at time of data collection

1. Was the method you received the one that you wanted?
2. (If wanted method was not received) Because the method you wanted was not available did the provider refer you to another provider who could provide that method?
3. At that time, did the provider explain to you how to use the method that you did receive effectively?
4. At that time, were you told about side effects or problems you might have with the method?
5. Were you told what to do if you experienced side effects or problems?
6. At that time, were you told about other methods of family planning that you could use?

Item response options: Yes or No, where Yes = 1 and No = 0.

The index was constructed by summing the item scores. The index score range is 0-6, and a higher index score indicates more comprehensive family planning counseling.

Quality of interaction with FP service provider Index

applies only to those who reported visiting a health facility for care for themselves in 6 months proceeding the survey

1. I felt I could discuss any problems, questions or concerns with the provider without feeling embarrassed
2. I felt that the provider was knowledgeable about family planning
3. The provider gave me enough information to make a decision about if I should use family planning method and what method to use
4. The provider was courteous and polite to me
5. I am satisfied with the care and services I received from the provider
6. I felt that I had enough privacy when I was with the provider
7. The provider made me feel embarrassed or ashamed for asking for FP services*
8. The provider ignored by requests or preferences*
9. The information the provider gave me was easy to understand
10. The provider did not listen to what I was saying*
11. The provider helped me get the family planning method that was best for me
12. The provider strongly encouraged me to use one FP method even though I preferred a different method*
13. The provider scolded or insulted me during my visit*

Item response options: “Agree”, “Disagree” or “Not Applicable”. Where Agree = 1 and Disagree and Not Applicable = 0.
*Items 7,8,10,12 and 13 are reverse coded. Disagree = 1 and Agree and Not Applicable = 0
The index was constructed by summing the item scores. The index score range is 0-13, and a higher index score indicates better quality interaction with FP service provider

Frequency of Interspousal Communication Scale

1. How often do you and your husband discuss things that happen during the day?
2. How often do you and your husband discuss your worries or feelings?
3. How often do you and your husband discuss what to spend household money on?
4. How often do you and your husband discuss when to have children?
5. How often do you and your husband discuss whether to use family planning?
6. How often do you and your husband discuss whether you should work to earn money?

Item response options: 5-point scale, where Always = 5, Often = 4, Sometimes = 3, Seldom = 2, and Never= 1
The scale was constructed by summing the item scores and dividing by the number of items. The scale score range is 1-5, and a higher scale score indicates more frequent interspousal communication

Normative expectations about girls' role Scale

1. People in my village expect girls to be accompanied when going to any place.
2. People in my village think that girls should not work outside home.
3. People in our village expect that girls are not good at managing money/ do not have a mind for business
4. People in our village expect husbands / men to earn all the money needed to support a household
5. People in my village think a girl should marry as early as possible to protect her chastity.
6. People in our village expect a girl to marry before the age of 18.
7. People in our village expect a girl to discontinue her study after marriage.
8. People in my village expect married girls to stay at home.
9. People in our village expect girls to have their first child soon after marriage.
10. People in my village expect that a married couple will not use family planning until they have had at least 5 children.

Item response options: 5-point scale, where Strongly disagree = 5, Disagree= 4, Neither agree / disagree= 3, Agree = 2, and Strongly Agree= 1
The scale was constructed by summing the item scores and dividing by the number of items. The scale score ranges from 1-5. The higher the score the more equitable the normative expectations are about the role for girls.

Outcome expectations about FP use Bangladesh

1. If people in my village knew I was using family planning, they would believe that I am unfaithful to my husband
2. If people in my village knew I was using family planning, they would believe that I am not committed to my marriage / family

Item response options: 5-point scale, where Strongly disagree = 5, Disagree= 4, Neither agree / disagree= 3, Agree = 2, and Strongly Agree= 1
The scale was constructed by summing the item scores and dividing by the number of items. The scale score ranges from 1-5. The higher the score the more positive outcome expectations about FP use are.

Outcome expectations about FP use Niger

1. If people in my village knew I was using FP, they would think badly of me
2. If people in my village knew I was using family planning, they would criticize my husband/family for allowing it
3. If people in my village knew that I was using family planning, they would say that I was not following the divine prescriptions

Item response options: 5-point scale, where Strongly disagree = 5, Disagree= 4, Neither agree / disagree= 3, Agree = 2, and Strongly Agree= 1
The scale was constructed by summing the item scores and dividing by the number of items. The scale score ranges from 1-5. The higher the score the more positive outcome expectations about FP use are. In Niger, these three items were asked and Anne suggested using them though the two items asked in Bangladesh are the original scale items. Please see page 6 of CARE IMAGINE Data and Variable Questions_2022_04_27_CAREV1.

Outcome expectations about delaying childbirth

1. If I do not have a child soon after marriage, my husband would consider divorcing me
2. If I do not have a child soon after marriage, my mother-in-law/family would disown / reject me.
3. If I do not have a child soon after marriage, my husband will lose respect / status in the community
    a. The original scale includes this item, however, this is not captured in Niger. In Niger, we will employ a three-item version of this scale. Please see page five of the short brief "CARE IMAGINE Data and Variable Questions_2022_04_27_CAREV1."
4. If I did not have a child soon after marriage, people in my village would think I was "infertile" or "barren"

Item response options: 5-point scale, where Strongly disagree = 5, Disagree= 4, Neither agree / disagree= 3, Agree = 2, and Strongly Agree= 1

The scale was constructed by summing the item scores and dividing by the number of items. The scale score ranges from 1-5. The higher the score the more positive outcome expectations about delaying childbirth are.

Outcome expectations engaging in income generating activities

1. If I participate in income-generating activities, my husband would not like that I am ignoring my duties as a wife.
2. If I participate in income-generating activities, people in my village will respect me for earning money
3. If I participate in income-generating activities, people in my village will gossip about me and call me "spoiled" or "ruined"

Item response options: 5-point scale, where Strongly disagree = 5, Disagree= 4, Neither agree / disagree= 3, Agree = 2, and Strongly Agree= 1

The scale was constructed by summing the item scores and dividing by the number of items. The scale score ranges from 1-5. The higher the score the more positive outcome expectations about engaging in income generating activities are.