Courant Research Centre

'Poverty, Equity and Growth in Developing and Transition Countries: Statistical Methods and Empirical Analysis'

Georg-August-Universität Göttingen
(founded in 1737)

Discussion Papers

No. 287

Effects of a large-scale Participatory Learning and Action Programme in Women's Groups on Health, Nutrition, Water, Sanitation, and Hygiene: a Cluster-Randomized Controlled Trial in Bihar, India

Lisa Bogler, Christian Bommer, Cara Ebert, Abhijeet Kumar, SV Subramanian, Malavika A. Subramanyam, Sebastian Vollmer

First Version: December 2021
This Version: April 2023

Platz der Göttinger Sieben 5 · 37073 Goettingen · Germany
Phone: +49-(0)551-3921660 · Fax: +49-(0)551-3914059

Email: crc-peg@uni-goettingen.de Web: http://www.uni-goettingen.de/crc-peg

Effects of a Large-Scale Participatory Learning and Action Programme in Women's Groups on Health, Nutrition, Water, Sanitation, and Hygiene: a Cluster-Randomized Controlled Trial in Bihar, India

Lisa Bogler[1], Christian Bommer[1], Cara Ebert[2], Abhijeet Kumar[1], SV Subramanian[3], Malavika A. Subramanyam[4]*, Sebastian Vollmer[1]*

[1]University of Göttingen, Göttingen, Germany, [2]RWI – Leibniz Institute for Economic Research, Essen, Germany, [3]Harvard T.H. Chan School of Public Health, Boston, United States, [4]IIT Gandhinagar, Gandhinagar, India

Abstract

Evidence from small-scale randomized controlled trials suggests that interventions relying on community involvement through a participatory learning and action (PLA) approach can improve health outcomes in resource-poor settings. However, it is only poorly understood whether PLA-based interventions are effective after scale-up in a real world setting. In a cluster-randomized controlled trial in Bihar, India, we assessed whether the PLA approach improved health, nutrition, water, sanitation, and hygiene (HNWASH) outcomes in adults and children when implemented state-wide by a government-supported agency. In the intervention, trained female facilitators ran 20 structured participatory meetings about key HNWASH topics in state-supported women's groups. Unlike the strong results of small-scale trials, in the scaled-up government implemented intervention we do not observe systematic improvements in HNWASH knowledge, attitudes, practices or health outcomes. We discuss aspects of programme implementation that could explain these null effects. Our findings call for caution when promising public health interventions are transformed into large policy programmes.

Key words: participatory learning and action; women's groups; health; nutrition; HNWASH

Corresponding author: Lisa Bogler, Waldweg 26, 37083 Goettingen, Germany, 0049 551 39 28962, lisa.bogler@cemis.uni-goettingen.de

* joint principal investigators

I. Introduction

Poor maternal and child health and nutrition persist as a public health challenge in many low and middle-income countries. In 2016, child mortality was 64.6 deaths per 1,000 live births among the poorest, and 31.3 deaths among the richest households in low and middle-income countries (Chao et al., 2018). The maternal mortality rate was estimated at 157 per 100.000 live births in low- and middle-income countries (Bauserman et al., 2020). India is no exception: India's child mortality rate was 39.4 deaths per 1,000 live births, and the most recent estimate of the maternal mortality rate was 174 deaths per 100,000 live births (World Bank, 2019). Stunting, an indication of chronic undernutrition, affects more than one third of children in India (World Bank, 2019).

Available evidence suggests that community participation has great importance for combating health and nutrition problems (Atkinson et al., 2010; Gibson & Anderson, 2009; Kar, 2003; Whittaker & Smith, 2015; WHO et al., 2009). A community-based participatory learning and action (PLA) approach implemented through women's groups has been shown to improve maternal and newborn health in different countries of South Asia and Africa. The participatory approach is designed to empower and create a sense of agency among participants (Freire, 1973). This in turn should enable them to identify problems and design solutions tailored to the locally perceived needs of each community. Evidence from multiple randomized controlled trials in low-resource settings suggests that PLA approaches in women's groups reduce neonatal mortality by up to 20 percent and maternal mortality by up to 23 percent (Prost et al., 2013). A commentary (Paul, 2016) on a PLA trial targeting newborn health (Tripathy et al., 2016) advocated for the use of women's groups as a scalable approach.

Yet, there are very few rigorous evaluations of the effectiveness of community-based participatory interventions implemented at scale. To our knowledge, the largest randomized PLA trial took place in Bangladesh and found no impact on neonatal mortality (Azad et al., 2010) before the trial was adjusted to achieve a much higher coverage of self-help groups (Fottrell et al., 2013). Maintaining sufficient coverage of the intervention is, however, not the only challenge for successful scale-up. At large scale more coordination and oversight are needed. An important policy question is whether the benefits of PLA programmes in controlled environments of small-scale RCTs can be preserved in a state-level implementation by government agencies.

The challenges of scaling up successful health interventions are widely known (Hanson et al., 2003; Mangham & Hanson, 2010; Milat et al., 2015; Perla et al., 2013; Yamey, 2012). Constraints to scale-up include a lack of resources, qualified personnel for intervention implementation and management, and political will. These challenges have not been sufficiently studied in the PLA context.

This study rigorously evaluates the effect of a PLA programme, called *Gram Varta,* implemented in women's groups on individual's health, nutrition, water, sanitation, and hygiene (HNWASH) in the Indian state of Bihar. State-supported agencies implemented *Gram Varta* in more than half of the districts in the state. This geographic scope was unprecedented and relied on existing women's self-help groups, which sets it apart from previous interventions (Azad et al., 2010; Fottrell et al., 2013). The implementation of *Gram Varta* was accompanied by a randomized controlled trial (RCT) in one of the districts, Madhepura, where the implementing agency was the Bihar state rural livelihoods project (Jeevika).

The evaluation of *Gram Varta* did not reveal effects on HNWASH outcomes. Instead, it highlights the importance of several features of programme design, and the difficulties of scaling up PLA interventions.

II. Material and methods

Study design and data collection

Setting

Bihar is a populous state with 104,099,000 inhabitants in the north-east of India and ranks among the poorest states in India (Reserve Bank of India, 2016). *Gram Varta* was first piloted in 2011 in Bihar's Patna district and has since been extended to 13 other districts, covering 48,058 self-help groups. This makes it the largest PLA programme in the world.

Gram Varta was implemented via existing village-based women's self-help groups that were affiliated with Women's Development Corporation (WDC), Jeevika, and Mahila Samkhya. To evaluate the programme, we accompanied the implementation of *Gram Varta* by means of a randomized controlled trial in six of the thirteen blocks (i.e., sub-districts) of Madhepura district, between 2015 and 2016. In Madhepura, the implementing agency was Jeevika, a state-sponsored agency that commenced operations in Bihar in 2006. The agency mobilises rural women to set up self-help groups with a focus on microfinance activities, and has been shown to be successful in reducing debts of the beneficiaries and empowering women in various dimensions (Datta, 2015).

Intervention: Gram Varta

Gram Varta used a PLA approach to improve maternal and child nutrition and health. This approach encourages participants to identify problems and create solutions themselves, empowering them through participatory methods. At the core of *Gram Varta* was a cycle of 20 structured meetings delivered via existing village-based women's self-help groups, supported by the state government. One cycle took 9-12 months and consisted of four distinct phases, to: (1) identify, discuss and prioritise problems; (2) identify and prioritise strategies to address the identified problems; (3) implement the strategies; and (4) evaluate the progress.

Jeevika, delivering *Gram Varta* in Madhepura, trained some of its existing staff to facilitate the meetings. The facilitators used social mapping to identify the target population and meeting sites that were comfortable, well-known, and visible. They invited community members to participate in the meetings and guided the self-help group through the PLA cycle using participatory techniques. These techniques included games, picture cards, stories, demonstrations, and other activities to encourage participants to think critically and engage in discussions. The meetings were about women's agency, attitude towards working together, service utilisation, and HNWASH knowledge and practices. In each meeting one major topic was discussed, such as undernutrition in children or the importance of cleanliness (see Table S1 for the complete agenda). While meetings were ideally to be held fortnightly, the order and exact timing of the meetings varied between groups as groups determined their own meeting schedules.

During the meeting cycle, participants identified key HNWASH problems in their households and communities, and developed strategies to remedy them. Participants drew up action plans, implemented their strategies, and monitored progress. Importantly, while women's self-help groups provided the framework for meetings, the whole community was invited to participate, including men and government-funded community workers such as the Accredited Social Health Activists (ASHAs) and *Anganwadi* workers (AWWs). ASHAs and AWWs are frontline workers who engage with the community on health and nutrition topics on a daily basis. Participants were expected to share the key messages they learned with family members, neighbours, and friends. In two official community meetings, participants could communicate previously identified problems to authorities and health service providers and request changes. Figure 1 describes the theory of change of the

intervention in detail.

Experimental design

We evaluated *Gram Varta* using a cluster-randomized controlled trial in Madhepura district. Madhepura has 13 blocks (administrative regions) comprising 443 *gram panchayats* (village clusters). Jeevika, the implementing agency, purposively chose 6 blocks and their 68 *gram panchayats* for potential implementation. Jeevika's self-help-group structure is federated and the highest structural level, the cluster level federation (CLF), only forms after two years of regular meetings and no major gaps in activities. Jeevika chose clusters with such mature self-help groups, i.e. those with CLF, for *Gram Varta* implementation. However, the objective criteria for selecting clusters were not shared with the research team.[2]

Based on Jeevika's selection of clusters, we randomized the implementation of *Gram Varta* at the level of *gram panchayats* using Stata. Randomization was stratified by block because census data revealed substantial heterogeneity between blocks. Thirty-four *gram panchayats* with 90 villages were assigned to the treatment group. The remaining 34 *gram panchayats* with 90 villages were introduced to *Gram Varta* post evaluation and formed the control group. Figure 2 maps out the number of households in each group. Programme implementation was strictly aligned with treatment assignment. Our sampling design ensured that study and control communities were separated by sufficient physical distance. Additionally, the traditional gender norms and patriarchal system in the area prevented women from travelling frequently to other villages (Mehta & Sai, 2021; Mondal et al., 2020). This reduced the likelihood of contamination and spillovers across *gram panchayats*.

To identify the survey sample, enumerators visited each village in the sample and recorded the number of households in each hamlet, a subdivision of a village. We interviewed an average of 22 households per village using probability proportional to size of hamlet to determine the number of households in each hamlet. Households were sampled through a random walk of the enumerators through the village. Participants in the control group were not informed about the implementation of *Gram Varta*. Survey data collection and data entry were blinded. Details regarding the power calculations can be found in the supplementary materials and Table S2.

Survey tools

Data were collected in multiple survey rounds: the baseline survey was conducted before the intervention, in March/April 2015, and the endline survey took place about two months after completion of the PLA cycle, in November/December 2016.

Three survey tools were administered in each household. The household head responded to questions about demographic, socioeconomic, and nutrition information at the household level, as well as HNWASH practices and attitudes. The second survey tool directed questions to a 'woman respondent' who was defined as a woman aged 15 to 49 with the youngest child in the household. The enumerators consulted with the male household head to identify the woman respondent. The purpose of this selection was to capture women with children below five years for whom several of the *Gram Varta* topics might be of highest interest.

Unfortunately, the female household head, who was often older than 49 years and not the woman with the youngest child in the household, was frequently mistakenly selected as the woman respondent at baseline. We changed the selection methodology for the endline survey to increase the number of respondents actually matching the group of interest, i.e. women with children aged less than five years old. Our new methodology selected the wife of the oldest son of the household head: we identified this person as being most likely to match our group of interest, because younger

sons and their wives often form separate households. The woman respondent was not chosen randomly, conditional on matching the group of interest, and we deliberately used very clear selection rules which could subsequently be checked against the household roster, to avoid any subjective leeway in the selection of the woman respondent. This approach prevented the reoccurrence of the mistakes made during the baseline survey. If there was no female matching the group of interest in the household, any married woman aged 15 to 49 was selected.

One day after the interview with the household head, the woman respondent was asked about her HNWASH practices, related attitudes and beliefs, self-help group membership, *Gram Varta* participation (in treatment areas), social capital, and about her children aged below five years. We collected detailed information about her last-born child, and less detailed information about other children aged below five years. On the same day, we took anthropometric measurements of all consenting household members. If a respondent was not present, we visited the household once on another day.

In addition to the quantitative surveys, two qualitative studies were carried out between June 2015 (baseline) and January 2017 (endline). The qualitative study in Madhepura district covered two villages from the treatment group and one village from the control group. At baseline, interviews were conducted with 24 attendees of self-help group meetings, three facilitators, three ASHAs/AWWs, three pregnant women, and two men. At endline, five additional women were interviewed. Focus group discussions provided further input. A second qualitative study was conducted in 12 blocks of six districts outside Madhepura district, completing 25 in-depth interviews in total. In three of these districts, WDC was the implementing agency. We used the qualitative interviews to understand the context of the communities before *Gram Varta* was implemented and how *Gram Varta* may have affected the communities. In this study, we used results from the qualitative interviews to inform the discussion of quantitative results.

Ethical clearance and trial registration

Ethical clearance was granted by the Institutional Ethics Committee of the Indian Institute of Technology Gandhinagar (approval number IEC/2014-15/2/MS/006). The study was funded by the International Initiative for Impact Evaluation (3ie), a US-based not-for-profit organisation. The content of the manuscript is the sole responsibility of the authors and was not influenced by 3ie. This study was registered with and a pre-analysis plan was submitted to 3ie before intervention roll-out. In addition, this study was registered after programme completion in the AEA RCT Registry and the unique identifying number is: AEARCTR-0004700.

Empirical framework

Outcomes

The main focus of *Gram Varta* was the dissemination of health knowledge among self-help group members through the use of participatory learning methods. We expected households in treatment areas to have better understanding of basic issues in health, hygiene, and nutrition, and to implement their knowledge through proper practices. We therefore analysed the impact of *Gram Varta* on health, nutrition, water, sanitation and hygiene (HNWASH) outcomes. Health outcomes would only be affected if knowledge and practices improved. Table S3 contains a list of outcomes analysed.

Self-rated health of adults, parental reports of their children suffering from diarrhoea or acute respiratory infections (ARI), measured body weight (of adults), weight-for-height (of under-fives), mid-arm circumference (under-fives), oedema (under-fives), and haemoglobin levels (all family

members) were used as indicators of health and nutritional status. Questions on hand-washing, use of toilets, and the domestic storage and treatment of drinking water were used to assess impact on sanitation and hygiene. All indicators were self-reported, except anthropometric measurements, presence of stool piles, sewage water, and type and cleanliness of toilet.

With this selection of outcomes, we stayed as close as possible to the extensive set of indicators originally proposed in the pre-analysis plan (Bommer et al., 2015) which reflected all indicators that *Gram Varta* aimed to affect. This study includes the indicators related to HNWASH knowledge and practices, the key element of *Gram Varta*, as well as health outcomes, which constitute end points of possible effects conditional on changes in HNWASH knowledge and practices. We dropped outcomes from the analysis that showed no variation at baseline -- these were variables for which the great majority of households held a correct belief or engaged in recommended behaviour prior to *Gram Varta* implementation such that the program would not have any further impact -- or that could not be expected to be impacted, such as height of adult men and women.

We divided outcomes into three groups, in accordance with the theory of change, of (i) belief and knowledge updating leading to (ii) changes in practices and resulting in (iii) improved health outcomes. Each of these outcome groups comprises a variety of specific outcomes divided into subgroups (see Table S3 for details).

We created indices for outcome groups when possible. Each index is calculated at the household or woman level by adding the number of outcomes which are true for this household or woman, and dividing the sum by the number of outcomes included in the group. The indices therefore range from 0 to 1. Outcomes for which 1 had a negative meaning were recoded such that 1 carried a positive meaning for the ease of comparing outcomes across dimensions. Outcomes were omitted from an index if they referred only to a subgroup. For example, “Frequency of breastfeeding” was excluded from the nutrition index as it was relevant only to children who were breastfed. The outcomes included in each index are marked in Table S3.

Several outcomes are measured at the child level; a household may have multiple children. These outcomes were summarised at the household level before being included in the respective index. The outcome was measured as true if it was true for any child in the household and not true if it was true for no child in the household. For child health, we additionally report results for the Composite Index for Anthropometric Failure (CIAF) which is 1 if a child is either stunted, wasted, or underweight, and 0 otherwise. Haemoglobin was recoded as a dummy for anaemia with a cut-off of 11 g/dl for inclusion in the index of child health.

Statistical analysis

As the overarching goal of *Gram Varta* was to improve HNWASH practices and health outcomes at the community level, this study primarily focuses on the intention-to-treat effect on members of the community where *Gram Varta* was implemented regardless of their actual participation in the meeting cycle. Because treatment assignment was random, a comparison of post-intervention outcomes between treatment and control groups, on average provides the unbiased intention-to-treat effect. We calculated two-sided, two-sample t-tests at endline, with block-level fixed effects and standard errors clustered at the *gram panchayat* level.

As robustness checks, we modelled linear regressions for binary and non-binary outcomes in the following form:

$$Y_{ip} = \alpha + \beta T_i + X_{ip}\delta + \omega_p + \varepsilon_{ip}$$

where Y_{ip} is the outcome for unit *i* (i.e.., woman respondent, household, or child) in *gram*

panchayat p measured at endline and T_i is the indicator for being in the treatment group. The coefficient β presents the intention-to-treat effect. For binary outcomes, the coefficient of the treatment variable can be interpreted as the percentage point change in the probability that the indicator of interest holds true. Standard errors were clustered at the *gram panchayat* level. X_{ip} is a range of household and women-specific characteristics that we included as covariates in the analysis. These were household size, and whether the household owned land, livestock or any durable household goods (assets), taken from the baseline survey. Religion and caste of household head were taken from the endline survey as these questions had not been included in the baseline questionnaire but can be assumed to be time-invariant. Age and education of the women respondents were taken from the endline questionnaire due to the change in selection methodology. ω_p are block-level fixed effects.

We further exploited the panel structure of our data and estimated a difference-in-differences model, using the subset of subjects present in both the baseline and endline data to account for potential unobserved heterogeneity.

$$Y_{ipt} = \gamma_i + \lambda_t + \beta D_i + X_{ipt}\delta + \varepsilon_{ipt}$$

where Y_{ipt} is the outcome for unit *i* in *gram panchayat p* at time *t*, with *t = 0* at baseline and *t = 1* at endline. γ_i and λ_t are unit-level and time fixed effects, respectively, where $\lambda_0 = 0$. D_i is a dummy variable that equals 1 if the data comes from the endline and unit *i* was assigned to the treatment group. The corresponding coefficient β estimates the intention-to-treat effect at endline.

We assessed statistical significance at the 5 percent level. We further applied the Benjamini-Hochberg method to correct for multiple hypotheses testing. All statistical analysis was done using Stata 16 (StataCorp LP).

III. Results

Baseline characteristics

We examined balance in observable characteristics for the treatment and control groups in the baseline sample (columns 1 to 4 of Table 1) and estimation sample of households (columns 5 to 8 of Table 1). The estimation sample consists of observations included in model (1) for which outcome data was available at both baseline and endline. Differences between the treatment and control groups in the baseline and estimation samples were almost identical in terms of size and significance.

Households were slightly larger in treatment villages compared to control villages, owned slightly more types of assets, and had slightly more education, and women respondents were slightly younger. However, the differences in standard deviation units were negligible. None reached the cut-off of 0.25 above which linear regression models may be sensitive to specifications (Imbens & Wooldridge, 2009). No noteworthy differences were observed in any outcome variable (see Table S4).

Survey and attrition

At baseline (approximately 10 months before the first group meetings), we interviewed the household head and one woman in 3,953 households from the 68 selected *gram panchayats* taking part in the evaluation. Out of these, 3,577 household heads (90.4 percent) and 3,153 women (79.8 percent) completed the respective questionnaire at endline. If one of the respondents was not present, if no proxy respondent was available, or if the respondent refused to participate, only the respective other questionnaire was completed. In several households, a woman respondent was missing at follow-up when there was no woman matching the selection criteria.

We checked for selective attrition by comparing characteristics of households that continued in the study to those that dropped out, both for control and treatment groups, and found some small differences in socioeconomic household characteristics. Households that dropped from the control group, for example, were slightly smaller than those that did not drop out. While small in terms of standard deviations, these differences were statistically significant at the 5 percent level. We therefore controlled for these characteristics in a robustness check.

We also checked for selective attrition specifically among women respondents in two ways. First, we checked differences in baseline characteristics between households that were retained because any woman was interviewed at follow-up, and those that were lost because no woman was interviewed at follow-up. In households with a woman respondent at both baseline and endline, it was not always the case that the same woman respondent was interviewed, because the selection criteria of woman respondents had been modified to improve the targeting of the group of interest. Because of this, we next checked differences between households where the same woman was interviewed in both waves versus households where no woman or a different woman was interviewed at follow-up. Households with the same respondent across waves had statistically significantly more children below the age of six years, and this was reflected also in the larger household size. Respondents that dropped out were statistically significantly older and less educated than those that stayed the same across waves. These differences are direct results of the modified criteria and affected the treatment and control groups equally. This is also reflected in the very similar treatment and control group characteristics of the baseline and estimation samples (see Table 1). Households with the same respondent across waves in the control group had slightly more assets, a factor we control for in a robustness check.

Coverage of women's groups and participation

Monthly progress reports listed the number of attendees in each category (i.e., gender, caste categories) for each of the 20 meetings summed across all self-help groups in each block. In Madhepura, PLA meetings were held in 3,129 self-help groups across all six blocks (see Table 2). The total population of the six blocks was 1,071,348 (Directorate of Census Operations, Bihar, 2011). Coverage, defined by the population per self-help group, varied from one group per 277 people to one group per 540 people across blocks. Average participation per self-help group meeting varied between 12 and 21 individuals, more than the 10 to 15 regular members previously reported for Jeevika-led self-help groups (Hoffmann et al., 2020). This average participation per meeting was calculated by averaging the number of participants reported across all self-help groups per meeting in each block, and dividing this average total participation by the number of existing self-help groups in the block, as *Gram Varta* was intended to be implemented in all existing self-help groups.

The majority of participants were women, while the average proportion of participating men (averaged across all self-help groups and meetings by block) ranged between 7.1 percent and 12.5 percent of participants. Participation by community workers was negligible, with no participation at all in some blocks. The proportion of pregnant women in PLA meetings varied between 3.5 and 6.6 percent of all participants.

Results of endline comparisons

Tables 3 to 5 present indicator means in control and treatment groups at endline and the estimated intention-to-treat effects (difference in means at endline) along with two-sided t-tests. Model (1) is a simple endline comparison of means, Model (2) controls for covariates, and Model (3) is employs a difference-in-differences estimation using panel observations only. We report conventional confidence intervals and multiple hypotheses corrected p-values using the Benjamini-Hochberg method.

Table 3 presents the results for nutritional beliefs, hygiene beliefs and health knowledge. Gram Varta did not systematically affect belief or knowledge indicators. Effect sizes are small and negative in most cases, except for the belief that feeding thick breastmilk is important. This effect is positive, large (7.54 percentage points) and statistically significant in the difference-in-differences specification of Model (3) at 5 percent level but not statistically significant after Benjamini-Hochberg correction for multiple hypotheses testing.

Table 4 presents the results on practice indicators. Gram Varta's impact on intake of micronutrients among children under five years was negative, although mostly not statistically significant. Whether the last-born child received any iron-folic acid (IFA) tablet or syrup seemed to be statistically significantly reduced (-5.37 percentage points, Model (1)). This effect remains statistically significant in Model (2) after Benjamini-Hochberg correction. Effects on feeding practices regarding the last-born child were mixed, with some coefficients in the expected direction, and others in the opposite direction. The effects on whether the last-born child was breastfed within 24 hours after birth and the nutrition practice index regarding the last-born child were positive and statistically significant in Model (2), but not consistently statistically significant across model specifications and not after Benjamini-Hochberg correction. The finding was similarly mixed regarding practices related to preventing diseases among children under five. While indicators of childhood vaccinations and deworming treatment changed in the expected direction, we found a statistically significant, negative effect on the probability of insecticide-treated bednet use for last-born children (-5.92 percentage points, Model (1)) which is large considering that less than 20 percent of last-born children in our sample slept under an insecticide-treated bednet. However, the negative effect is only statistically significant for Model (1) and is not statistically significant after Benjamini-Hochberg correction. The index summarising these outcomes related to the prevention of diseases suggests no statistically significant impact.

Our results showed a positive effect on water storage and treatment. *Gram Varta* appeared to increase statistically significantly the probability that the household's water treatment was adequate (1.86 percentage points, Model (1)) which corresponds to a large relative effect size given low base levels (1.57%) and which remains statistically significant after Benjamini-Hochberg correction. *Gram Varta*'s impact on hygiene practices, including the use of soap, was negative or mixed at best, and the summary index suggests no impact. The probability that the household practiced open defecation despite a toilet being available was reduced (-5.59 percentage points), but this is statistically significant only in Model (1) and not after Benjamini-Hochberg correction. Similarly, the frequency of buying soap fell, but the effect is statistically significant only in Model (2) and not statistically significant after Benjamini-Hochberg correction.

Table 5 presents the results on health outcomes. The two self-reported indicators on women's health were not affected. Similarly, recent diseases of the last-born child and other children below five years were not affected by *Gram Varta*. Further, the analysis did not show an impact of *Gram Varta* on anthropometric measurements among women or men aged 20-49 years, or children below five years. Out of 16 indicators based on anthropometric measurements, nine showed changes in the expected direction in models (1) and (2), but none was statistically significantly different between treatment and control. The two indices capturing child health are negative and not statistically significant.

To assess impact heterogeneity by treatment intensity, we created a proxy of actual exposure to *Gram Varta* using the subgroup of respondents who self-reported to be members of Jeevika-led self-help groups and participated regularly in meetings. Of all women respondents, 43.1 percent were members of any self-help group, 92.1 percent of whom reported that at least one of their groups was led by Jeevika, and 86.2 percent indicated that they participated regularly in Jeevika-led groups. We repeated the main analysis for this subgroup of treated households. In alignment with the main analysis, this exercise did not reveal consistent effects on outcomes (see Tables S5, S6, S7

in the supplementary material).

IV. Discussion

We did not find that *Gram Varta* had the expected impact on HNWASH beliefs and knowledge, practices, and health outcomes in Madhepura, Bihar. While the qualitative work indicated some improvements in knowledge among younger women, this did not translate into a quantitatively measurable impact. It is possible that the observation period was too short to detect impacts on some of our health outcomes. However, as changes in knowledge and practices are necessary conditions for a change in health outcomes, the absence of programme impacts on the latter is not surprising.

We tested whether the lack of program impacts can be explained by concurrent programmes affecting the study population (Scott et al., 2021, 2022) by examining time trends in the outcome indices for the treatment and control groups separately (Figure S1 and Table S8). In the presence of concurrent programs, we would expect systematic changes that also extend to the control group. However, while there were improvements in some indicators in both groups, other indicators changed for the worse. Moreover, interviewed policy makers and the implementing agency were unable to identify concurrent programmes.

Previous trials and reviews have identified low coverage (i.e., a large population per self-help group) as a barrier to successful scale-up. (Fottrell et al., 2013; Prost et al., 2013; Seward et al., 2017). *Gram Varta*, however, was implemented with a coverage similar to that of successful PLA trials (Fottrell et al., 2013; Tripathy et al., 2010).

Building on the structure of existing self-help groups was one major distinction between this and previous successful PLA trials. However, Jeevika-led self-help groups were formed with a focus on microcredit activities. This is a marked difference to the approach in other PLA trials, in which self-help groups were formed around health topics (Azad et al., 2010; Colbourn et al., 2013).

Unlike other interventions, *Gram Varta* did not include a component of strengthening existing health services. Qualitative work showed a lack of responsiveness to demands made by the community among officials in charge of service delivery. One study in Malawi directly compared the PLA approach with a strengthening of health care facilities and found that both aspects are important for the intervention's success (Colbourn et al., 2013; Hanson et al., 2003).

Moreover, engaging health care providers in community-based interventions has been shown to contribute to their success (Morrow & Martin, 2003). In fact, other trials relied on dedicated health workers as facilitators of self-help group meetings (Tripathy et al., 2016). In the case of *Gram Varta*, PLA meetings were facilitated by Jeevika staff, who were recruited hastily to lead the meetings and only had to meet the requirements of being female and literate. One third of facilitators had not worked as a community mobilizer before the start of *Gram Varta*, revealing a lack of experience in this role. The motivation of a facilitator and her relationship with the community have been shown to strengthen community-based interventions (Barker et al., 2007; Glenton et al., 2013; Kar & Pasteur, 2005; Morrison et al., 2005) and these factors were presumably deficient in *Gram Varta*, where, according to qualitative interviews, some facilitators did not appear motivated to fulfil their tasks. Delayed honorarium payments and irregular schedules of recruitment and training of facilitators may explain some of the lack of motivation.

In addition, several meetings in the PLA cycle were delayed and irregular. This was due to funding delays, staffing delays, changes in the government officers in charge of the programme, local elections, and the co-opting of *Gram Varta* functionaries by other government programmes. The monthly progress reports reveal that a few meetings were held after long delays, followed by

several meetings in quick succession.

Gram Varta's poor implementation is also reflected in participants' programme awareness. In the treatment area, only 10.7 percent of woman respondents had heard of *Gram Varta* at endline and only 35 percent of those reported attendance in one to five *Gram Varta* meetings. While programme awareness and participation were also fairly low in other trials (Prost et al., 2013), *Gram Varta* ranks at the lower end in this regard.

Our study contributes novel and important evidence for the debate on scaling-up public health interventions. Nevertheless, the analysis suffers from some limitations. Among the potential factors that could explain *Gram Varta's* lack of impact, we cannot clearly identify which of these factors specifically explain the absence of treatment effects. However, it is likely that challenges during programme implementation, which seem to have been more serious than in other settings, arose to a significant degree because of the large scale of the programme.

A further limitation relates to the lack of detailed data on actual self-help group participation preventing us from conclusively evaluating the effect of *Gram Varta* conditional on participation. However, the treatment intensity analysis of respondents self-reporting regular participation in Jeevika-led self-help group meetings provides a reasonably good approximation, and equally showed no consistent treatment effects on outcomes.

Although our study did not find *Gram Varta* to be effective in improving HNWASH outcomes, it does not refute the proven effectiveness of the community-based PLA approach in general. It rather highlights the importance of several programme design aspects and difficulties in scaling up such interventions.

Endnotes

[1] See https://www.censusindia2011.com/bihar/madhepura-population.html (last accessed October 2, 2022).

[2] Further selection criteria might include the ease of implementation in and geographic location of clusters, but this is speculative.

Acknowledgements

We are very grateful to Sini Varghese and Sagar Atre for their qualitative work that contributed greatly to this study. We would also like to thank all enumerators for their reliable work during data collection.

References

Atkinson, J.-A. M., Fitzgerald, L., Toaliu, H., Taleo, G., Tynan, A., Whittaker, M., Riley, I., & Vallely, A. (2010). Community participation for malaria elimination in Tafea Province, Vanuatu: Part I. Maintaining motivation for prevention practices in the context of disappearing disease. *Malaria Journal*, *9*, 93. https://doi.org/10.1186/1475-2875-9-93

Azad, K., Barnett, S., Banerjee, B., Shaha, S., Khan, K., Rego, A. R., Barua, S., Flatman, D., Pagel, C., Prost, A., Ellis, M., & Costello, A. (2010). Effect of scaling up women's groups on birth outcomes in three rural districts in Bangladesh: A cluster-randomised controlled trial. *The Lancet*, *375*(9721), 1193–1202. https://doi.org/10.1016/S0140-6736(10)60142-0

Barker, G., Ricardo, C., & Nascimento, M. (2007). *Engaging men and boys in changing gender-based inequity in health: Evidence from programme interventions*. World Health Organization. http://apps.who.int/iris/bitstream/10665/43679/1/9789241595490_eng.pdf

Bauserman, M., Thorsten, V. R., Nolen, T. L., Patterson, J., Lokangaka, A., Tshefu, A., Patel, A. B., Hibberd, P. L., Garces, A. L., Figueroa, L., Krebs, N. F., Esamai, F., Nyongesa, P., Liechty, E. A., Carlo, W. A., Chomba, E., Goudar, S. S., Kavi, A., Derman, R. J., … Bose, C. (2020). Maternal mortality in six low and lower-middle income countries from 2010 to 2018: Risk factors and trends. *Reproductive Health*, *17*(3), 173. https://doi.org/10.1186/s12978-020-00990-z

Bommer, C., Subramanyam, M., & Vollmer, S. (2015). *Evaluation of Gram Varta in rural Bihar—A mixed methods study. Analysis plan* [Internal document].

Chao, F., You, D., Pedersen, J., Hug, L., & Alkema, L. (2018). National and regional under-5 mortality rate by economic status for low-income and middle-income countries: A systematic assessment. *The Lancet Global Health*, *6*(5), e535–e547. https://doi.org/10.1016/S2214-109X(18)30059-7

Colbourn, T., Nambiar, B., Bondo, A., Makwenda, C., Tsetekani, E., Makonda-Ridley, A., Msukwa, M., Barker, P., Kotagal, U., Williams, C., Davies, R., Webb, D., Flatman, D., Lewycka, S., Rosato, M., Kachale, F., Mwansambo, C., & Costello, A. (2013). Effects of quality improvement in health facilities and community mobilization through women's groups on maternal, neonatal and perinatal mortality in three districts of Malawi: MaiKhanda, a cluster randomized controlled effectiveness trial. *International Health*, *5*(3), 180–195. https://doi.org/10.1093/inthealth/iht011

Datta, U. (2015). Socio-Economic Impacts of JEEViKA: A Large-Scale Self-Help Group Project in Bihar, India. *World Development*, *68*, 1–18. https://doi.org/10.1016/j.worlddev.2014.11.013

Directorate of Census Operations, Bihar. (2011). *District Census Handbook – Madhepura*. http://www.censusindia.gov.in/2011census/dchb/1011_PART_B_DCHB_MADHEPURA.pdf

Fottrell, E., Azad, K., Kuddus, A., Younes, L., Shaha, S., Nahar, T., Aumon, B. H., Hossen, M., Beard, J., Hossain, T., Pulkki-Brannstrom, A.-M., Skordis-Worrall, J., Prost, A., Costello, A., & Houweling, T. A. J. (2013). The Effect of Increased Coverage of Participatory Women's Groups on Neonatal Mortality in Bangladesh: A Cluster Randomized Trial. *JAMA Pediatrics*, *167*(9), 816–825. https://doi.org/10.1001/jamapediatrics.2013.2534

Freire, P. (1973). *Education for Critical Consciousness*. Seabury Press.

Gibson, R. S., & Anderson, V. P. (2009). A Review of Interventions Based on Dietary Diversification or Modification Strategies with the Potential to Enhance Intakes of Total and Absorbable Zinc. *Food and Nutrition Bulletin*, *30*(1_suppl1), S108–S143. https://doi.org/10.1177/15648265090301S107

Glenton, C., Colvin, C., Carlsen, B., Swartz, A., Lewin, S., Noyes, J., & Rashidian, A. (2013). Barriers and facilitators to the implementation of lay health worker programmes to improve access to maternal and child health: Qualitative evidence synthesis. *Cochraine Database of Systematic Reviews*, *10*, CD010414.

Hanson, K., Ranson, M. K., Oliveira-Cruz, V., & Mills, A. (2003). Expanding access to priority health interventions: A framework for understanding the constraints to scaling-up. *Journal of International Development*, *15*(1), 1–14. https://doi.org/10.1002/jid.963

Hoffmann, V., Rao, V., Surendra, V., & Datta, U. (2020). Relief from usury: Impact of a self-help group lending program in rural India. *Journal of Development Economics*, *148*, 102567. https://doi.org/10.1016/j.jdeveco.2020.102567

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, *47*(1), 5–86.

Kar, K. (2003). *Subsidy or self-respect? : Community led total sanitation ; an update on recent developments* [IDS Working Paper No 184]. https://opendocs.ids.ac.uk/opendocs/handle/123456789/4052

Kar, K., & Pasteur, K. (2005). *Subsidy or self-respect? Participatory community-led total sanitation: an update on recent developments* [IDS Working Paper No 257]. https://opendocs.ids.ac.uk/opendocs/handle/123456789/4052

Mangham, L. J., & Hanson, K. (2010). Scaling up in international health: What are the key issues? *Health Policy and Planning*, *25*(2), 85–96. https://doi.org/10.1093/heapol/czp066

Mehta, V., & Sai, H. (2021). Freedom of Movement: Studying Women's Mobility in North India. *Urbanisation*, *6*(1_suppl), S77–S114. https://doi.org/10.1177/24557471211022566

Milat, A. J., Bauman, A., & Redman, S. (2015). Narrative review of models and success factors for scaling up public health interventions. *Implementation Science*, *10*(1), 113. https://doi.org/10.1186/s13012-015-0301-6

Mondal, D., Karmakar, S., & Banerjee, A. (2020). Women's autonomy and utilization of maternal healthcare in India: Evidence from a recent national survey. *PLoS ONE*, *15*(12), e0243553. https://doi.org/10.1371/journal.pone.0243553

Morrison, J., Tamang, S., Mesko, N., Osrin, D., Shrestha, B., Manandhar, M., Manandhar, D., Standing, H., & Costello, A. (2005). Women's health groups to improve perinatal care in rural Nepal. *BMC Pregnancy and Childbirth*, *5*, 6. https://doi.org/10.1186/1471-2393-5-6

Morrow, A., & Martin, L. (2003). *Community-based strategies for breastfeeding promotion and support in developing countries*. World Health Organization. http://apps.who.int/iris/bitstream/10665/42859/1/9241591218.pdf

Paul, V. K. (2016). Participatory women's groups: Time for integration into programmes. *The Lancet Global Health*, *4*(2), e74–e75. https://doi.org/10.1016/S2214-109X(16)00010-3

Perla, R. J., Bradbury, E., & Gunther-Murphy, C. (2013). Large-Scale Improvement Initiatives in Healthcare: A Scan of the Literature. *Journal for Healthcare Quality*, *35*(1), 30–40. https://doi.org/10.1111/j.1945-1474.2011.00164.x

Prost, A., Colbourn, T., Seward, N., Azad, K., Coomarasamy, A., Copas, A., Houweling, T. A. J., Fottrell, E., Kuddus, A., Lewycka, S., MacArthur, C., Manandhar, D., Morrison, J., Mwansambo, C., Nair, N., Nambiar, B., Osrin, D., Pagel, C., Phiri, T., … Costello, A. (2013). Women's groups practising participatory learning and action to improve maternal and newborn health in low-resource settings: A systematic review and meta-analysis. *The Lancet*, *381*(9879), 1736–1746. https://doi.org/10.1016/S0140-6736(13)60685-6

Reserve Bank of India. (2016). *Handbook of Statistics on Indian States*. Reserve Bank of India.

Scott, S., Gupta, S., Kumar, N., Raghunathan, K., Thai, G., Quisumbing, A., & Menon, P. (2021). A Women's Group-Based Nutrition Behavior Change Intervention in India Has Limited Impacts Amidst Implementation Barriers and a Concurrent National Behavior Change Campaign. *Current Developments in Nutrition*, *5*(Supplement_2), 179–179.

Scott, S., Gupta, S., Menon, P., Raghunathan, K., Thai, G., Quisumbing, A., Prasad, V., Hegde, A., Choudhury, A., & Khetan, M. (2022). A Quasi-Experimental Evaluation of a Nutrition Behavior Change Intervention Delivered Through Women's Self-Help Groups in Rural India: Impacts on Maternal and Young Child Diets, Anthropometry, and Intermediate Outcomes. *Current Developments in Nutrition*, *6*(6), nzac079.

Seward, N., Neuman, M., Colbourn, T., Osrin, D., Lewycka, S., Azad, K., Costello, A., Das, S., Fottrell, E., Kuddus, A., Manandhar, D., Nair, N., Nambiar, B., More, N. S., Phiri, T., Tripathy, P., & Prost, A. (2017). Effects of women's groups practising participatory learning and action on preventive and care-seeking behaviours to reduce neonatal mortality: A meta-analysis of cluster-randomised trials. *PLOS Medicine*, *14*(12), e1002467. https://doi.org/10.1371/journal.pmed.1002467

Tripathy, P., Nair, N., Barnett, S., Mahapatra, R., Borghi, J., Rath, S., Rath, S., Gope, R., Mahto, D., Sinha, R., Lakshminarayana, R., Patel, V., Pagel, C., Prost, A., & Costello, A. (2010). Effect of a participatory intervention with women's groups on birth outcomes and maternal depression in Jharkhand and Orissa, India: A cluster-randomised controlled trial. *The Lancet*, *375*(9721), 1182–1192. https://doi.org/10.1016/S0140-6736(09)62042-0

Tripathy, P., Nair, N., Sinha, R., Rath, S., Gope, R. K., Rath, S., Roy, S. S., Bajpai, A., Singh, V., Nath, V., Ali, S., Kundu, A. K., Choudhury, D., Ghosh, S. K., Kumar, S., Mahapatra, R., Costello, A., Fottrell, E., Houweling, T. A. J., & Prost, A. (2016). Effect of participatory women's groups facilitated by Accredited Social Health Activists on birth outcomes in rural eastern India: A cluster-randomised controlled trial. *The Lancet Global Health*, *4*(2), e119–e128. https://doi.org/10.1016/S2214-109X(15)00287-9

Whittaker, M., & Smith, C. (2015). Reimagining malaria: Five reasons to strengthen community engagement in the lead up to malaria elimination. *Malaria Journal*, *14*(1), 410. https://doi.org/10.1186/s12936-015-0931-9

WHO, UNICEF, & World Bank. (2009). *State of the World's Vaccines and Immunization*. World Health Organization.

World Bank. (2019). *World Development Indicators, Health Indicators, 2013-2018, India*. http://databank.worldbank.org/data/reports.aspx?source=2&country=IND

Yamey, G. (2012). What are the barriers to scaling up health interventions in low and middle income countries? A qualitative study of academic leaders in implementation science. *Globalization and Health*, *8*(1), 11. https://doi.org/10.1186/1744-8603-8-11

Table 1. Sample characteristics at baseline

	(a) Baseline sample				(b) Estimation sample			
Characteristic	**N**	**Mean (C)**	**Mean (T)**	**Difference in SD (P-Value)**	**N**	**Mean (C)**	**Mean (T)**	**Difference in SD (P-Value)**
Household composition								
Household size	3953	5.22	5.62	-0.19 (0.0000)	3577	5.26	5.61	-0.17 (0.0000)
Children below 6	3953	0.72	0.91	-0.19 (0.0000)	3577	0.73	0.89	-0.16 (0.0000)
Household assets								
Land ownership (yes/no)	3953	46.21%	48.05%	-0.04 (0.2475)	3577	46.59%	48.03%	-0.03 (0.3892)
Types of livestock owned	3953	0.91	0.91	0.00 (0.9367)	3577	0.93	0.92	0.01 (0.7953)
Number of cattle owned	2765	1.58	1.61	-0.01 (0.7545)	2533	1.60	1.63	-0.01 (0.8158)
Number of asset types owned	3953	4.32	4.54	-0.11 (0.0003)	3577	4.32	4.56	-0.12 (0.0003)
Highest education in household								
No education	3953	13.03%	9.93%	0.10 (0.0023)	3577	12.70%	9.47%	0.10 (0.0021)
Primary completed	3953	38.69%	41.51%	-0.06 (0.0702)	3577	38.71%	41.66%	-0.06 (0.0726)
Secondary or higher	3953	48.28%	48.56%	-0.01 (0.8639)	3577	48.59%	48.87%	-0.01 (0.8638)
Caste								
Scheduled castes	3953	11.36%	9.53%	0.06 (0.0593)	3577	12.48%	10.60%	0.06 (0.0783)
Scheduled tribes	3953	7.12%	6.84%	0.01 (0.7310)	3577	7.82%	7.61%	0.01 (0.8137)
Other backward classes	3953	39.80%	37.76%	0.04 (0.1886)	3577	43.70%	42.00%	0.03 (0.3018)
General caste category	3953	6.36%	8.21%	-0.07 (0.0255)	3577	6.99%	9.13%	-0.08 (0.0185)
Religion								
Hindu	3953	63.33%	62.80%	0.01 (0.7273)	3577	69.55%	69.84%	-0.01 (0.8497)
Muslim	3953	26.97%	25.34%	0.04 (0.2445)	3577	29.62%	28.18%	0.03 (0.3449)
Age of woman								
Age of woman	3875	37.99	36.76	0.10 (0.0016)	2307	33.19	32.12	0.12 (0.0037)
Share of women aged 18-29	3875	26.50%	31.30%	-0.11 (0.0010)	2307	35.36%	40.33%	-0.10 (0.0138)
Share of women aged 30-39	3875	30.37%	31.20%	-0.02 (0.5760)	2307	39.77%	40.50%	-0.01 (0.7214)
Share of women aged >40	3875	43.13%	37.49%	0.11 (0.0003)	2307	24.87%	19.17%	0.14 (0.0009)
Education of woman								
No education	3953	68.99%	64.37%	0.10 (0.0021)	2338	64.07%	58.85%	0.11 (0.0095)

Primary completed	3953	17.37%	21.54%	-0.11 (0.0009)	2338	19.50%	25.58%	-0.15 (0.0004)
Secondary or higher	3953	13.13%	13.68%	-0.02 (0.6097)	2338	15.91%	15.31%	0.02 (0.6902)

Note: The table presents the number of observations and means of covariates at baseline for control and treatment groups in (a) the baseline sample and (b) the estimation sample of model (1) for which baseline information is available; as well as the difference in standard deviations between control and treatment mean and the p-value of a t-test of the difference between the mean of the control and the treatment group in parentheses below the difference in standard deviations. The estimation sample with baseline information is considerably smaller for women characteristics due to a change in sampling within the household.

Table 2. Process indicators

Block: / Indicator:	Bihariganj	Gwalpara	Kumarkhand	Madhepura Sadar	Murliganj	Udakishun-ganj	Total
Population*	135534	126020	243629	191375	185631	189159	1071348
Number of SHGs	415	310	843	541	670	350	3129
Av. total participation	6829	4538	17737	8991	7877	5672	8607
Av. participation per SHG meeting	16	15	21	17	12	16	16
Av. number of ANMs	0	1	1	0	2	9	2
Av. number of ASHAs	18	3	35	1	7	74	23
Share of women	92.34%	91.73%	91.65%	92.86%	87.46%	92.66%	91.45%
Share of men	7.66%	8.27%	8.35%	7.14%	12.54%	7.34%	8.55%
Share of adolescent girls	9.94%	19.21%	14.82%	10.90%	19.49%	9.87%	14.04%
Share of lactating women	14.34%	24.76%	11.22%	12.68%	23.03%	16.18%	17.04%
Share of pregnant women	4.89%	5.82%	4.30%	3.46%	6.61%	4.85%	4.99%
Share of SHG members	67.94%	82.18%	55.71%	67.62%	83.68%	67.09%	70.70%
Share of scheduled tribes	1.30%	0.64%	1.50%	0.00%	17.06%	0.00%	3.42%
Share of scheduled caste	41.21%	38.97%	43.72%	35.44%	28.03%	27.24%	35.77%
Share of OBC	53.98%	57.28%	52.15%	59.03%	102.18%	62.63%	64.54%
Share of general caste	3.17%	2.96%	2.61%	2.71%	1.96%	10.09%	3.92%
Coverage**	327	407	289	354	277	540	342

Note: The table presents data from monthly progress reports (MPR) by block; MPRs listed for meetings 1 to 20 separately, how many participants of which category attended these meetings, summed across all SHGs in the respective block. Average total participation is calculated as the number of participants summed across all SHGs in a block, averaged across meetings 1 to 20; average participation per SHG meeting is the average total participation divided by SHGs in a block; the number of ANMs and ASHAs is the average number of these frontline workers attending a SHG meeting in each block; all shares are average shares (out of all meeting participants) of those participants belonging to a specific population group attending a SHG meeting in each block. The last column (Total) contains averages across blocks. SHG(s) = self-help group(s); Av. = Average; ANM(s) = auxiliary nurse midwife(-wives); ASHA(s) = accredited social health activist(s); OBC = other backward castes
* population in respective block according to 2011 population census
** number of population (according to 2011 population census) per self-help group

Table 3. Intention to treat effects for HNWASH outcomes of the *Gram Varta* trial – Belief and knowledge indicators

Measure/indicator	**Control**	**Treated**	**ITT effect**	**ITT effect**	**ITT effect**
	Mean (N)	Mean (N)	Model (1)	Model (2)	Model (3)
Nutritional beliefs					
Index of nutrition beliefs	91.73% (701)	90.68% (703)	-0.97 [-2.65,0.71]	-0.93 [-2.83,0.97]	0.82 [-5.89,7.52]
Balanced food important	97.94% (1748)	97.86% (1733)	-0.06 [-1.04,0.91]	-0.41 [-1.46,0.63]	0.93 [-1.63,3.48]
Feeding thick breastmilk important	93.62% (815)	95.23% (818)	1.98 [-0.03,3.98]	2.26 [-0.07,4.59]	7.54 [0.37,14.70]
Complementary feeding after 6m important	95.70% (837)	95.45% (835)	-0.36 [-2.28,1.55]	-0.52 [-2.91,1.87]	0.34 [-10.36,11.03]
Adding oil to meal of 6-59 months old important	79.49% (790)	74.45% (779)	-4.88 [-10.18,0.43]	-3.98 [-10.03,2.07]	-5.51 [-24.23,13.21]
Hygiene beliefs					
Index of hygiene beliefs	75.65% (1632)	76.25% (1632)	0.49 [-1.83,2.82]	-0.97 [-3.32,1.37]	-0.47 [-5.18,4.23]
Reprehends others for ODF	49.27% (1792)	49.01% (1763)	-0.58 [-5.78,4.62]	-2.95 [-8.55,2.66]	-2.05 [-11.75,7.64]
Believes ODF is health hazard	84.75% (1718)	86.00% (1700)	1.10 [-1.45,3.65]	-0.63 [-3.61,2.34]	3.71 [-1.64,9.07]
Toilet/covering excreta important	89.52% (1680)	90.66% (1670)	1.19 [-1.31,3.68]	0.73 [-1.19,2.65]	-2.31 [-8.15,3.52]
Health knowledge					
Index of health knowledge	42.43% (1334)	43.76% (1355)	1.58 [-1.50,4.67]	-2.08 [-5.71,1.56]	-5.50 [-17.47,6.47]
Heard of message: malaria/dengue	51.74% (1494)	55.15% (1494)	3.57 [-0.55,7.69]	-0.58 [-5.40,4.24]	-2.00 [-17.03,13.02]
Heard of message: diarrhoea	56.85% (1497)	55.50% (1499)	-1.01 [-4.96,2.94]	-3.22 [-7.70,1.26]	-8.19 [-23.25,6.87]
Heard of message: ARI	47.83% (1478)	50.72% (1467)	3.12 [-0.88,7.12]	0.40 [-4.55,5.34]	-2.62 [-15.77,10.54]
Heard of message: STI	17.47% (1374)	18.37% (1383)	0.97 [-1.99,3.92]	-1.64 [-5.32,2.03]	-4.72 [-14.89,5.46]
Knows danger signs of malaria	49.94% (1588)	50.61% (1565)	1.22 [-2.21,4.65]	-3.37 [-7.85,1.11]	-3.51 [-17.21,10.19]
Knows danger signs of ARI	39.29% (1588)	38.66% (1565)	-0.17 [-4.13,3.80]	-3.31 [-8.31,1.69]	-7.53 [-19.58,4.52]
Knows treatment of diarrhoea	36.15% (1588)	35.46% (1565)	-0.69 [-4.62,3.23]	-3.52 [-7.95,0.91]	-5.99 [-17.85,5.87]

The table presents intention-to-treat effect estimates with 95%-confidence intervals in square brackets as well as corresponding means of variables and sample sizes for control and treatment group members at endline. Model (1) is the simple endline comparison of means, Model (2) controls for covariates, Model (3) is the difference-in-differences model. Models (1) and (2) include block-level fixed effects. Model (3) could not be estimated for anthropometric indicators due to missing panel identification of other household members. Standard errors were clustered at the *gram panchayat* level. For binary indicators, means were expressed as percentages and intention-to-treat estimates as absolute marginal effects in terms of percentage points, while continuous and discrete measures were left untransformed. *Last born* refers to the last-born child of the woman respondent at the time of the respective survey, *other child* refers to all children under five years of the woman respondent except the last-born child, *any child* refers to all children under five years of the woman respondent. ITT = intention-to-treat; ODF = open defecation; ARI = acute respiratory infections; STI = sexually transmitted infections. Asterisks (*) indicate statistical significance after Benjamini-Hochberg correction.

Table 4. Intention to treat effects for HNWASH outcomes of the *Gram Varta* trial – Practice indicators

Measure/indicator	**Control**	**Treated**	**ITT effect**	**ITT effect**	**ITT effect**
	Mean (N)	Mean (N)	Model (1)	Model (2)	Model (3)
Intake of micronutrients					
Number of vitamin A doses (last-born child)	0.90 (654)	0.79 (670)	-0.13 [-0.29,0.04]	-0.20 [-0.40,-0.00]	-0.34 [-1.29,0.61]
IFA tablet/syrup (last born)	32.05% (780)	27.16% (777)	-5.37 [-9.71,-1.03]	-9.62 [-15.85,-3.38]*	-2.44 [-30.42,25.53]
Received vitamin A dose (any child)	47.17% (1183)	45.94% (1232)	-2.00 [-6.58,2.58]	-3.24 [-9.30,2.82]	-1.83 [-25.20,21.54]
Multivitamin tablet/syrup (other child)	49.63% (540)	48.50% (567)	-0.97 [-8.57,6.62]	3.96 [-5.30,13.22]	-3.64 [-45.07,37.79]
Nutritional beliefs and practice					
Index of nutrition practice	77.74% (831)	80.34% (834)	2.88 [-0.07,5.82]	3.99 [0.44,7.54]	8.22 [-3.55,20.00]
Breastfed child within 24h (last born)	85.83% (847)	87.84% (839)	2.23 [-0.85,5.32]	4.14 [0.14,8.14]	9.30 [-3.64,22.24]
Gave pre-lacteal feeding at first day (last born)	30.63% (839)	27.23% (841)	-3.76 [-7.90,0.38]	-4.03 [-9.04,0.99]	-7.11 [-26.29,12.07]
Frequency of breastfeeding (last born)	63.64% (484)	61.87% (493)	-0.68 [-6.75,5.39]	-3.77 [-12.48,4.94]	7.26 [-53.37,67.88]
Days of exclusive breastfeeding (last born)	275.92 (223)	315.51 (253)	43.39 [-19.25,106.03]	11.28 [-59.26,81.83]	-13.30 [-593.04,566.44]
Number of meals per day of child (last born)	2.96 (508)	3.00 (516)	0.04 [-0.15,0.23]	0.04 [-0.17,0.26]	0.20 [-0.72,1.11]
Prevention of diseases					
Index of disease prevention	46.69% (846)	46.72% (842)	0.13 [-1.18,1.44]	-0.32 [-1.66,1.02]	1.51 [-1.76,4.78]
All vaccinations (any child)	59.25% (643)	61.16% (708)	1.88 [-3.27,7.03]	4.47 [-1.71,10.65]	3.19 [-40.58,46.96]
Vaccination card (any child)	66.36% (1418)	67.43% (1437)	2.14 [-3.02,7.31]	2.83 [-2.71,8.37]	-9.78 [-32.06,12.50]
Child sleeps under bednet (last born)	97.91% (860)	97.31% (855)	-0.38 [-1.87,1.10]	-0.02 [-1.69,1.65]	2.52 [-7.49,12.54]
Treated bednet (last born)	18.41% (766)	12.66% (766)	-5.92 [-11.11,-0.74]	-5.27 [-10.87,0.33]	-6.37 [-27.55,14.81]
Deworming past 6m (any child)	48.71% (1400)	50.71% (1406)	1.16 [-3.43,5.75]	-0.75 [-6.06,4.56]	2.83 [-17.95,23.61]
Water treatment					
Treats drinking water	3.39% (1800)	4.74% (1772)	1.63 [0.34,2.92]*	1.04 [-0.35,2.42]	-0.36 [-3.26,2.54]
Water treatment is adequate	2.61% (1800)	4.18% (1772)	1.86 [0.72,3.00]*	1.64 [0.38,2.89]*	0.48 [-1.86,2.81]
Hygiene practices					
Index of hygiene practices	46.63% (1711)	47.06% (1676)	0.27 [-2.58,3.13]	-0.79 [-3.75,2.18]	-2.09 [-6.54,2.37]

Buys soap at least every 2 months	62.68% (1774)	59.24% (1737)	-3.63 [-8.28,1.02]	-5.19 [-9.77,-0.60]	-4.86 [-12.04,2.32]
Improved toilet at home	18.67% (1741)	19.74% (1707)	0.95 [-3.10,5.00]	-1.04 [-5.28,3.21]	-2.15 [-6.24,1.94]
Subjective cleanliness of toilet	69.88% (342)	69.32% (352)	-0.98 [-9.79,7.84]	-3.12 [-11.92,5.67]	-6.07 [-26.33,14.19]
Observed: stool piles	53.75% (1801)	51.38% (1773)	-2.68 [-6.25,0.88]	-2.19 [-6.85,2.46]	4.08 [-4.48,12.64]
Observed: Sewage water	24.94% (1800)	23.56% (1774)	-1.47 [-6.13,3.19]	-1.36 [-5.68,2.95]	2.79 [-5.81,11.40]
Disposes infant stool adequately	36.88% (320)	39.49% (314)	2.94 [-6.30,12.18]	4.19 [-5.79,14.16]	-1.94 [-28.57,24.69]
Household practices ODF but toilet available	17.23% (325)	11.57% (337)	-5.59 [-10.48,-0.70]	-5.02 [-10.55,0.51]	8.27 [-7.22,23.75]
Uses soap after toilet	54.13% (1803)	54.23% (1774)	-0.79 [-6.30,4.71]	-0.67 [-6.67,5.32]	0.72 [-9.30,10.74]
Uses soap before meal	23.18% (1803)	21.98% (1774)	-1.23 [-6.37,3.92]	-2.09 [-8.10,3.93]	-1.30 [-10.55,7.94]

The table presents intention-to-treat effect estimates with 95%-confidence intervals in square brackets as well as corresponding means of variables and sample sizes for control and treatment group members at endline. Model (1) is the simple endline comparison of means, Model (2) controls for covariates, Model (3) is the difference-in-differences model. Models (1) and (2) include block-level fixed effects. Model (3) could not be estimated for anthropometric indicators due to missing panel identification of other household members. Standard errors were clustered at the *gram panchayat* level. For binary indicators, means were expressed as percentages and intention-to-treat estimates as absolute marginal effects in terms of percentage points, while continuous and discrete measures were left untransformed. *Last born* refers to the last-born child of the woman respondent at the time of the respective survey, *other child* refers to all children under five years of the woman respondent except the last-born child, *any child* refers to all children under five years of the woman respondent. ITT = intention-to-treat; IFA = iron-folic acid. Asterisks (*) indicate statistical significance after Benjamini-Hochberg correction.

Table 5. Intention to treat effects for HNWASH outcomes of the *Gram Varta* trial – Health outcome indicators

Measure/indicator	**Control**	**Treated**	**ITT effect**	**ITT effect**	**ITT effect**
	Mean (N)	Mean (N)	Model (1)	Model (2)	Model (3)
Women's and child's health, self-reported					
Self-assessed health is good	52.69% (1579)	54.15% (1555)	1.29 [-3.03,5.61]	-0.10 [-5.56,5.37]	0.43 [-11.11,11.98]
Feels chronically tired	77.59% (1575)	75.43% (1555)	-2.73 [-6.08,0.62]	-3.44 [-7.25,0.38]	5.59 [-3.95,15.12]
Child had no diarrhoea in past 3m (any child)	80.56% (1404)	83.11% (1409)	2.90 [-0.28,6.08]	3.57 [-0.21,7.35]	0.49 [-20.96,21.95]
Child had no ARI in past 3m (any child)	88.16% (1393)	89.10% (1404)	1.29 [-1.84,4.43]	1.68 [-1.89,5.24]	-1.73 [-18.83,15.37]
Anthropometric measurement					
Index of child health	24.73% (372)	24.16% (365)	-0.69 [-3.60,2.22]	-0.29 [-4.17,3.60]	n/a
CIAF	65.61% (660)	62.08% (712)	-3.48 [-9.20,2.23]	-3.90 [-10.87,3.07]	n/a
Weight, female 20-49	45.03 (1321)	45.50 (1404)	0.49 [-0.22,1.20]	0.16 [-0.63,0.95]	n/a
Hemoglobin, female 20-49	12.03 (1245)	12.03 (1327)	0.03 [-0.08,0.14]	0.05 [-0.08,0.17]	n/a
Weight, male 20-49	54.69 (518)	55.22 (540)	0.70 [-0.66,2.05]	-0.28 [-1.61,1.05]	n/a
Hemoglobin, male, 20-49	13.94 (494)	14.01 (511)	0.15 [-0.12,0.42]	0.18 [-0.14,0.49]	n/a
Weight, U5	10.58 (806)	10.54 (838)	-0.06 [-0.37,0.25]	-0.07 [-0.47,0.32]	n/a
Weight-for-age z-score, U5	-1.61 (794)	-1.65 (828)	-0.04 [-0.18,0.10]	0.05 [-0.12,0.22]	n/a
Underweight, U5	37.78% (794)	38.41% (828)	0.50 [-4.46,5.45]	-0.46 [-6.78,5.85]	n/a
Wasting, U5	17.91% (681)	15.52% (728)	-3.47 [-7.47,0.53]	-2.83 [-7.06,1.39]	n/a
Height, U5	82.72 (781)	82.08 (849)	-1.03 [-2.47,0.41]	-1.12 [-2.73,0.50]	n/a
Height-for-age z-score, U5	-1.72 (781)	-1.84 (849)	-0.17 [-0.36,0.03]	-0.09 [-0.31,0.12]	n/a
Stunting, U5	50.45% (781)	50.65% (849)	1.08 [-3.85,6.00]	-1.15 [-6.43,4.13]	n/a
Hemoglobin, U5	10.81 (554)	10.91 (559)	0.14 [-0.05,0.33]	0.13 [-0.09,0.36]	n/a
Oedema, U5	1.82% (877)	1.19% (926)	-0.52 [-1.62,0.58]	-0.50 [-1.84,0.85]	n/a
Arm circumference, U5	14.31 (656)	14.39 (668)	0.08 [-0.13,0.29]	0.12 [-0.15,0.40]	n/a
Arm circumference-for-age z score, U5	-1.01 (655)	-0.95 (666)	0.07 [-0.11,0.25]	0.11 [-0.13,0.34]	n/a
Arm circumference < 115 mm, U5	2.90% (655)	1.80% (666)	-1.26 [-2.98,0.45]	-1.71 [-3.57,0.14]	n/a

The table presents intention-to-treat effect estimates with 95%-confidence intervals in square brackets as well as corresponding means of variables and sample sizes for control and treatment group members at endline. Model (1) is the simple endline comparison of means, Model (2) controls for covariates, Model (3) is the difference-in-differences model. Models (1) and (2) include block-level fixed effects. Model (3) could not be estimated for anthropometric indicators due to missing panel identification of other household members. Standard errors were clustered at the *gram panchayat* level. For binary indicators, means were expressed as percentages and intention-to-treat estimates as absolute marginal effects in terms of percentage points, while continuous and discrete measures were left untransformed. *Last born* refers to the last-born child of the woman respondent at the time of the respective survey, *other child* refers to all children under five years of the woman respondent except the last-born child, *any child* refers to all children under five years of the woman respondent. ITT = intention-to-treat; CIAF = composite index for anthropometric failure. Asterisks (*) indicate statistical significance after Benjamini-Hochberg correction.

Figure 1. Theory of change

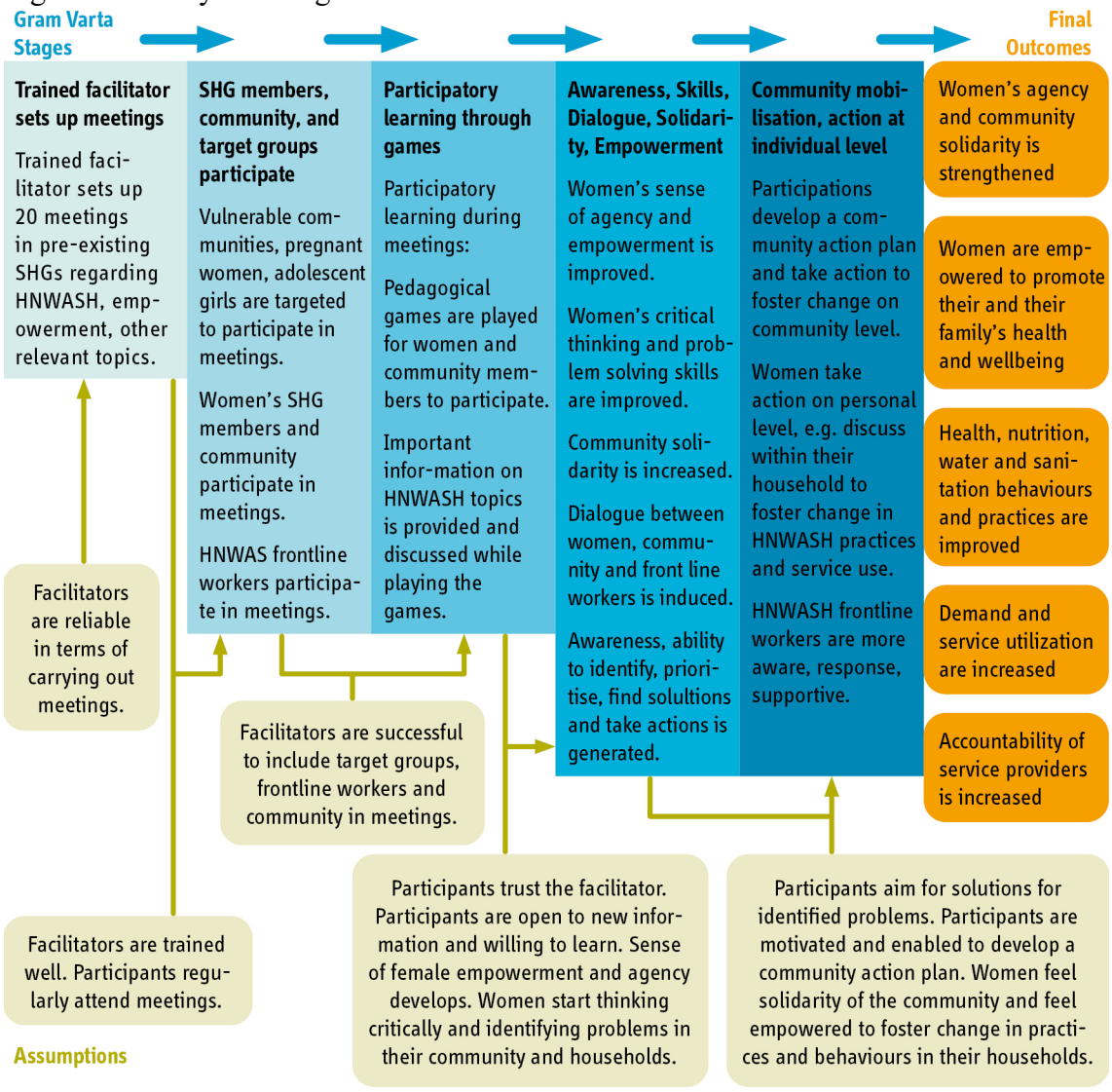
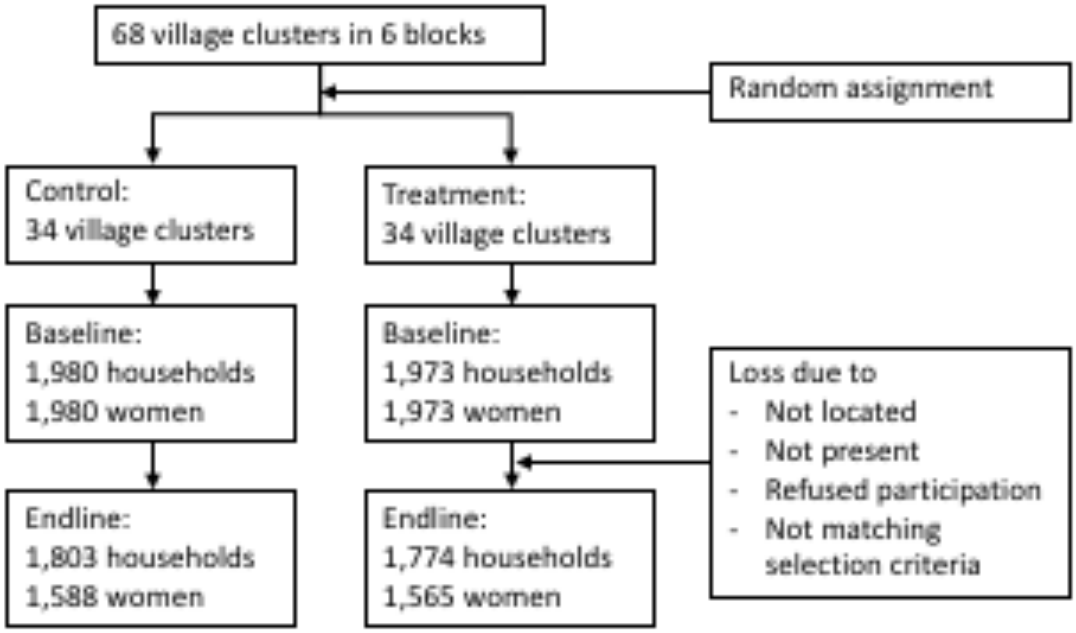


Figure 2. Study design and sampling



Note: Due to reasons explained in detail in subsection *survey tools*, the mechanism for selecting responds from the survey households changed at endline. Therefore, only a subset of women is present at both baseline and endline. Within the control group, 1,169 women were observed both at baseline and endline; within the treatment group, also 1,169 women were observed both at baseline and endline.