

Bogler, Lisa et al.

Working Paper

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

Discussion Papers, No. 287

Provided in Cooperation with:

Courant Research Centre 'Poverty, Equity and Growth in Developing and Transition Countries', University of Göttingen

Suggested Citation: Bogler, Lisa et al. (2021) : 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, Discussion Papers, No. 287, Georg-August-Universität Göttingen, Courant Research Centre - Poverty, Equity and Growth (CRC-PEG), Göttingen

This Version is available at:

<https://hdl.handle.net/10419/248421>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

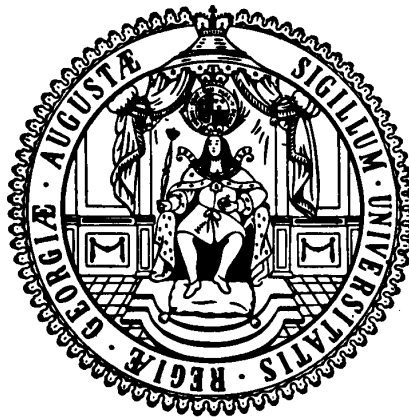
You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

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

December 2021

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 Bogler¹, Christian Bommer¹, Cara Ebert², Abhijeet Kumar¹, SV Subramanian³,
Malavika A. Subramanyam^{4*}, Sebastian Vollmer^{1*}

¹University of Göttingen, ²RWI – Leibniz Institute for Economic Research,
³Harvard T.H. Chan School of Public Health, ⁴IIT Gandhinagar

Abstract

Evidence from 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, whether PLA-based interventions remain effective after scale-up is only poorly understood. 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 at large scale by a government-supported agency. The intervention consisted of trained female facilitators conducting a series of 20 structured participatory meetings about key HNWASH topics in state-supported women's groups. In contrast to the strong results of small-scale trials, we do not observe systematic improvements in HNWASH knowledge, attitudes, behaviour or health outcomes but document irregularities in the implementation of the intervention. These 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 22401, lisa.bogler@cemis.uni-goettingen.de,

* joint principal investigators

I. Introduction

Poor maternal and child health and nutrition continues to pose a major public health challenge in many low- and middle-income countries. Despite decades of health and nutrition programmes, progress has been limited. 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 situation in 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). Rates of undernutrition are especially alarming: More than one third of children in India are stunted, an indication of chronic undernutrition (World Bank, 2019).

Evidence abounds that community participation is of 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 in women's groups has been found to be effective in improving maternal and newborn health in South Asia and Africa. The participatory approach intends to create a sense of empowerment and agency among participants and enhances their critical consciousness (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 the PLA approach 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 recent commentary (Paul, 2016) on a PLA trial targeting newborn health (Tripathy et al., 2016) advocated the use of women's groups as an approach ready for scale-up.

Yet, studies rigorously evaluating the effectiveness of community-based participatory interventions implemented at large scale are scarce. To our knowledge, the largest randomized PLA trial was implemented 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 as large-scale implementation may come with a change in implementing actors and an increased need for coordination and oversight. An important policy question, therefore, is whether PLA benefits sustain state-level implementation.

In fact, 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 lack of resources, qualified personnel for intervention implementation, and political will, as well as weak management. However, these challenges have not been sufficiently studied in the PLA context.

This study rigorously evaluates the effect of a PLA programme implemented in women's groups, called *Gram Varta*, on individual's health, nutrition, water, sanitation, and hygiene (HNWASH) in the north-eastern Indian state Bihar. *Gram Varta* was implemented in more than half of all districts of Bihar by state-supported agencies. To enable this uniquely wide geographic scope, *Gram Varta's* implementation was facilitated through existing women's self-help groups, setting 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). Since 2006, Jeevika has mobilised a large number of rural women to set up self-help groups, which focus on microfinance activities and offer a curriculum on women's empowerment and basic literacy and numeracy (Datta, 2015; Hoffmann et al., 2020).

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

Thanks to *Gram Varta's* wide geographic scope within Bihar and high-level governmental support, *Gram Varta* provides an ideal example to study the feasibility and effectiveness of scale-ups in the context of PLA interventions implemented by government-supported agencies. This study presents novel causal evidence for HNWASH- and health-related outcomes, highlights important differences in implementation design that may come with the scale-up of PLA programmes, and discusses potential avenues for future research.

The remainder of the paper proceeds as follows: Section II describes the setting, intervention, study design, and data collection. Section III presents the theoretical framework of the PLA approach while section IV presents the empirical framework of the study. Section V summarizes the results of the analysis. Section VI provides a discussion of the results and the final section VII concludes.

II. Study design and data collection

Setting

Gram Varta was first piloted in 2011 in Patna district of Bihar and has since been extended to more than half of the state's territory, resulting in the largest PLA programme implemented today. *Gram Varta* was implemented through village-based women's self-help groups affiliated with Women's Development Corporation, Jeevika, and Mahila Samkhya. To evaluate the programme, we accompanied the implementation of *Gram Varta* in six of the thirteen blocks (i.e., sub-districts) of Madhepura district by means of a randomized controlled trial between 2015 and 2016. In Madhepura, the implementing agency was Jeevika, a state-sponsored agency that started operating in Bihar in 2006. The agency mobilises rural women to set up self-help groups with a focus on microfinance activities and has been successful in reducing debts of the beneficiaries and empowering women in various dimensions (Datta, 2015).

Madhepura is largely rural and one of the most underdeveloped districts in Bihar, which itself ranks among the worst performing states on indicators of development, health, and nutrition. Data for Madhepura from 2015 show that 51.8 percent of children under-five were stunted and only 2.6 percent of children aged 6 to 23 months received an adequate diet (International Institute for Population Sciences (IIPS) & Macro International, 2016). Only 53.3 percent of households had electricity and 15.0 percent used improved sanitation facilities. Whereas the literacy rate for men was 76.9 percent, it was only 32.6 percent for women.

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 pre-structured meetings implemented through existing village-based women's self-help groups supported by the state government. One cycle took nine months to one year and consisted of four distinct phases, namely 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.

In Madhepura, Jeevika selected some of their staff members and trained them in facilitating the meetings. These trained facilitators first conducted a social mapping to identify the target population and potential meeting sites. Meeting sites were chosen to be comfortable, well-

known, and visible. They then invited community members to participate in the meetings and guided the self-help group through the PLA cycle using participatory techniques. This included games, picture cards, stories, demonstrations, and other activities to encourage participants to think critically and engage in discussions. Meeting contents related to women's agency, attitude towards working together, service utilisation, as well as HNWASH knowledge and practices. In each meeting one major topic was discussed, such as undernutrition in children or the importance of cleanliness. Table S1 in the supplementary material contains the agenda of the full PLA cycle. While meetings were ideally to be held fortnightly, the order and exact timing of the meetings varied between groups as each group decided themselves when meetings should take place.

During the meeting cycle, participants identified key problems related to HNWASH in their households and communities, and possible strategies to improve HNWASH practices. After health priorities and strategies were identified, participants drew up action plans, implemented strategies, and monitored progress. Importantly, while self-help groups provided the framework for holding 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 key messages learned with other community members including family members, neighbours, and friends. In addition to regular meetings, two official community meetings were scheduled, allowing participants to communicate previously identified problems to authorities and health service providers and to request changes.

Experimental design

The evaluation of *Gram Varta* was a cluster-randomized controlled trial set in Madhepura district. Madhepura is divided into 13 blocks (administrative regions) comprising 443 *gram panchayats* (village clusters), of which 6 blocks and their 68 *gram panchayats* were purposively chosen by Jeevika, the implementing agency, for potential implementation. The self-help-group structure of Jeevika 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.

Based on Jeevika's selection, we then randomized the implementation of *Gram Varta* at the level of *gram panchayats*. 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 other 34 *gram panchayats* with 90 villages were introduced to *Gram Varta* post evaluation and formed the control group. Randomization was done using Stata. Figure 1 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 cultural context and patriarchal system prevalent in the area prevented women from frequently travelling to other villages. This reduced the likelihood of contamination and spill-overs across *gram panchayats*.

For survey sample identification, enumerators visited each sampled village and recorded the number of households in each hamlet, a subdivision of a village. Probability proportional to size of hamlet was used to determine the number of households to be interviewed in each hamlet to reach a sample of on average 22 households per village. 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.

Power calculations

The total number of households was chosen given budgetary constraints. However, a power calculation was done at baseline to determine whether the survey capacity would allow us to determine minimum detectable effect (MDE) sizes for various indicators (Bommer et al., 2015b). We defined MDEs below 0.2 as small, between 0.2 and 0.8 as medium and above 0.8 as large and calculated MDEs for intra-cluster correlations (ICC) of 0.01, 0.05, and the ICC found in the baseline. Our calculations suggested that we would be able to detect small effects for indicators referring to hygiene, women's self-reported health outcomes, and women's health knowledge. Among anthropometric indicators and haemoglobin levels, we would be able to detect small effects among children below age 5, but only medium to large effects among women and men.

We re-did the power calculation given endline data in order to verify whether attrition changed MDE sizes. Results are summarized in table S5. We followed the method proposed by Duflo et al. for designs with randomisation over groups (Duflo et al., 2007). Given endline data, we adjusted the calculation to the actual number of clusters (column 4) and the actual number of observations per cluster (column 3) included in the analysis. As the number of clusters stayed the same for most indicators, this only increased MDEs slightly (compare column 2 to 5). At baseline, we had assumed compliance with the treatment, the share of treated households, to be 75 percent (compliance of 75 percent is assumed in columns 2 and 5). At endline, only 43 percent of women respondents claimed to be members of a self-help group. We used this reported fraction as a lower benchmark for compliance, although non-members may have attended meetings. This lower fraction considerably increased, often doubled, MDEs (column 6). It is worth noting that imperfect compliance only limited our ability to precisely estimate programme effects on those who actively participated in *Gram Varta* meetings. By contrast, the community-wide effects of *Gram Varta* are best captured by an intention-to-treat-analysis (the main focus of this article), whose precision is generally unaffected by imperfect compliance. We therefore re-did the calculation without correcting for imperfect compliance (column 7). Indeed, our calculations showed that we could detect small effects for most outcomes.

Survey tools

Data were collected during multiple survey rounds: A baseline survey was conducted before the intervention, in March/April 2015, and about two months after completion of the PLA cycle, in November/December 2016, the endline survey took place. In addition, in March 2016, a cross-section of 1,000 households selected from the baseline sample was surveyed for a midline. Due to delays in the implementation of the PLA cycle, only about three PLA meetings had taken place by March 2016 and the midline was not used in this analysis.

In each household, three survey tools were administered. The household head responded to questions about demographic, socioeconomic, and nutrition information at the household level, as well as HNWASH practices and attitudes. With the help of the household head, the enumerators identified the woman respondent. At baseline, a woman aged 15 to 49 with the youngest child in the household was determined as the woman respondent. The purpose of this selection was to capture women with children below 5 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 women respondent at baseline. The selection methodology was therefore changed for the midline and endline surveys to increase the number of respondents actually matching the group of interest, i.e. women with children below 5 years. The wife of the oldest son of the household head was then selected because this person had been identified as most likely to be matching the group of interest, whereas younger sons and their wives often have formed separate households. The woman respondent was not randomly chosen, conditional on matching the group of interest, and it was purposefully decided that very clear selection rules were needed, which could later on be checked against the household roster, to avoid any subjective leeway in the selection of the woman respondent and, therefore, to avoid the reoccurrence of mistakes taken place at baseline. 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. Detailed information was collected about her last-born child, while fewer questions were asked about her other children aged below five years. One the same day, anthropometric measurements were taken of all consenting household members. If a respondent was not present, the household was revisited once on another day.

In addition to the quantitative surveys, a qualitative study was carried out in two villages of the treatment group and in one village of the control group. At baseline and endline, interviews and focus group discussions were conducted with attendees of self-help group meetings, facilitators, ASHAs/AWWs, teenage girls, and men. These interviews aimed at understanding 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.

III. Theoretical framework

PLA is a development strategy that understands participation of the target group in the development process both as a means and an end (Wetmore & Theron, 1998). The PLA approach puts community priorities first and uses local resources for problem solving. The argument is that local actors might prioritise problems differently than external actors. They are also often able to identify solutions more suitable to their environment than outside experts (Kar, 2003). In addition, participation is seen as empowering, because it instils a feeling of ownership and therefore has the potential to lead to sustainable change once project funds are depleted. The sense of agency and skills created by PLA could be put to use in other domains outside the initial programme.

The theoretical foundation of PLA builds on Paulo Freire's concept of critical consciousness. The Brazilian educator advocated for participatory education, in which people are actors instead

of objects and identify problems and solutions themselves (Wallerstein & Bernstein, 1994). Critical consciousness means “learning to perceive social, political, and economic contradictions and to take action against the oppressive elements of reality” (Freire, 1973). Although the concept was developed for education, it has been applied to the health context (Minkler & Cox, 1980; Wallerstein & Bernstein, 1988).

The success of PLA hinges on several conditions. These start with a facilitator capable of creating awareness for the programme and encouraging the target population to participate in meetings. Husbands and other family members, travel costs, weather conditions and work or child care obligations must not be barriers for women who are willing to attend meetings.

Once in the meetings, participants engage actively with the topics discussed, are willing to learn new information, and act upon it. This requires the development of a sense of empowerment and the strengthening of problem solving skills. Messages learned are then shared with non-attending community members, leading to a change in norms and behaviour in larger parts of the community. If health care providers are part of the suggested solutions, they have to be willing to participate in the discussion and react to demands from the community. Finally, behaviour change leads to improvements in measurable health outcomes.

Given the vulnerability of the causal chain, the quality of programme implementation is crucial for its success. A priori it is unclear whether building on existing structures as in *Gram Varta* would strengthen the causal chain. While community members may be used to come together in meetings and engage with each other, the novel aspect of the programme might be lost in the regularity of self-help group meetings.

IV. Empirical framework

Outcomes

The key element of *Gram Varta* was the uptake of health knowledge among self-help group members through the use of participatory learning methods. We expected households in treatment areas to exhibit a better understanding of basic issues in health, hygiene, and nutrition, and implement their knowledge through the use of 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 a knowledge increased and practices improved. A list of outcomes analysed in this study is presented in the supplementary material table S2.

Self-rated health of adults, parental reports of their children suffering from diarrhoea or acute respiratory infections (ARI), measured body weight (adults), weight-for-height (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, 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 and sewage water, type and cleanliness of toilet. All outcomes are described in more detail in table S2 in the supplementary material.

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., 2015a). This plan proposed 7 groups of hypotheses, containing a total of 59 hypotheses. This study focuses on 12 hypotheses around 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 only those outcomes from the analysis that showed no variation at baseline or that could not be expected to be impacted, such as height of adult men and women.

To simplify reading the results, we created indices for outcome groups with 5 or more individual outcomes: nutritional beliefs and practice (attitude toward family feeding and feeding of newborns), prevention of diseases, hygiene, health knowledge, and child health based on anthropometric measurements. Each index is calculated on household or woman level by adding the number of outcomes, which are true for this household or woman and dividing this sum by the number of included outcomes. It therefore ranges from 0 to 1 with 1 being a more positive outcome. Outcomes for which 1 had a negative meaning were recoded such that 1 carried a positive meaning. Outcomes were omitted from an index if they only referred to a subgroup. To give one example “Frequency of breastfeeding” was excluded from the nutrition index as it was only asked for those children who were breastfed. The outcomes included in each index are marked in table S2. Several outcomes are measured on child level (all children) and one household may have multiple children. Such outcomes were summarised on household level before being included in the respective index (prevention of diseases). 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 for inclusion in the index of child health. The CIAF and the index of child health are reversed as 1 carries a negative meaning.

Statistical analysis

As the overarching goal of *Gram Varta* was to improve HNWASH practices and health outcomes on the community level, this study primarily focuses on the intention-to-treat effect of the intervention, which identifies the average effect on members of the community regardless of their actual participation in the meeting cycle. Participation in *Gram Varta* meetings was voluntary, thus actual treatment, i.e. attending PLA meetings, deviated from assignment to the treatment. Our analysis captured the effect of living in a *gram panchayat* where *Gram Varta* was implemented. Because assignment was random, a comparison of post-intervention outcomes (i.e. at endline) 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 standard errors clustered at the village level. Almost all outcomes were coded as binary indicators (with few exceptions) and most anthropometric indicators were continuous values. Table S2 in the supplementary material indicates the format (binary/ integer/ continuous) and respondent of each outcome variable.

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

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

where Y_{iv} is the outcome for unit i (e.g. woman respondent or household) in village v 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 village level. X_{iv} is a range of household and women-specific characteristics which we included as covariates in the analysis. These were household size, 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 woman respondent were taken from the endline questionnaire due to the change in selection methodology.

We further exploited the panel structure of our study 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.

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

where Y_{ivt} is the outcome for unit i in village v 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, taking the outcomes at baseline as reference values.

We assessed statistical significance at the 5 percent level. All statistical analysis was done using Stata 16 (StataCorp LP).

V. Results

Baseline characteristics

To assess whether randomisation successfully balanced observable characteristics, we examined baseline data for the treatment and control groups in the baseline sample of households. There were some statistically significant differences in background characteristics between treatment and control groups (see table 2). Households were slightly larger in treatment villages versus control villages, they owned slightly more types of assets, the woman respondent was slightly younger, and the household had slightly better education. However, the difference in standard deviation units was 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 variables (see table S3).

10.5 percent of the sampled households reported to belong to a scheduled caste, 7.0 percent to a scheduled tribe, and 38.8 percent to other backward classes (the three disadvantaged caste groups as per government classification). The majority of households was Hindu (70 percent) and less than one third was Muslim. Two thirds of women respondents had no education at all.

Survey and attrition

At baseline (approximately 10 months before the first group meetings), the household head and one woman were interviewed 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 (approximately two months after the last group meetings). 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 because there was no female member matching the selection criteria.

We checked for selective attrition by comparing characteristics of households that continued in the study with those that dropped out, both for control and treatment group 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. We therefore controlled for these characteristics in a robustness check. We also checked for selective attrition specifically among woman 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 both at baseline and endline, not necessarily the

same woman respondent had been interviewed because the selection criteria of woman respondents had been modified to improve the targeting of the group of interest. We therefore also 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 more children below the age of six years, reflected also in the larger household size. Respondents dropping out were older and less educated compared to those that stayed the same across waves. These differences are direct results of the modified criteria and affected treatment and control group equally. 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 the respective block. In Madhepura, PLA meetings were held in 3,129 self-help groups across all six blocks (see table 1). The population of the six blocks totalled 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, a higher number compared to 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 planned to be implemented in all existing self-help groups. The number of meetings actually held and the number of meetings cancelled were not consistently reported across meetings and blocks. Over time, average participation varied only slightly. In the first meeting the average participation across blocks was approximately 15 people, increasing to around 16 for meetings 3 to 9, then dropping again slightly before increasing to more than 17 for meetings 16 to 20. This pattern, however, was not consistent across blocks. This may be partly due to the fact that the actual order of meetings did not follow the scheduled agenda for all self-help groups.

According to the monthly progress reports, 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 of 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.

Overall, it appears that monthly progress reports were only partially filled by facilitators and therefore do not fully capture the actual implementation progress. Caution is especially warranted regarding information on who was truly present. Anecdotal observations suggest that meeting participants who were not part of the original self-help group members were not always reported in meeting notes.

Results of endline comparison

Table 3 presents indicator means in control and treatment groups at endline as well as the estimated intention-to-treat effects (difference in means at endline) along with two-sided t-tests. *Gram Varta's* impact on intake of micronutrients among children under five years was negative, although insignificant. The index of nutrition, summarising outcomes related to nutritional

beliefs and practices, suggests no impact of *Gram Varta*. The attitude of the household head on the importance of a balanced nutrition was not affected by *Gram Varta*. However, almost all household heads in both treatment and control group reported that a balanced nutrition was important even before the intervention. Effects on feeding practices were mixed, with some coefficients in the expected and others in the opposite direction, but none reached conventional levels of significance. The finding was similarly mixed regarding behaviours preventing diseases among children under 5. While indicators of childhood vaccinations and deworming treatment changed in the expected direction, we found a significantly negative effect on the probability of insecticide-treated bednet use for last-born children (-5.74 percentage points) which is large considering that less than 20 percent of last-born children in our sample slept under an insecticide-treated bednet. The index summarising these outcomes related to the prevention of diseases suggests no significant impact.

Our results showed a positive effect on water storage and treatment. *Gram Varta* appeared to significantly increase the probability that the household's water treatment was adequate (1.56 percentage points) which corresponds to a large relative effect size (59.8%) given low base levels. *Gram Varta*'s impact on hygiene behaviour, including use of soap was negative or mixed at best, the summary index suggests no impact. Whereas 11 hygiene indicators were insignificantly affected, the probability that the household practiced open defecation despite a toilet being available was significantly reduced (5.66 percentage points). Out of the seven indicators of knowledge of diseases, none was significantly improved and only four changed in the expected direction. The index summarising these knowledge outcomes suggests no impact. The two self-reported indicators on women's health were not affected, although the coefficients were in the expected direction. Similarly, recent diseases of the last-born child and other children below 5 years were not affected by *Gram Varta*.

Our analysis did not show an impact of *Gram Varta* on anthropometric measurements among women or men aged 20-49 years, or children below 5 years. Out of 16 indicators, 10 showed changes in the expected direction, but none were significantly different between treatment and control. The two indices capturing child health mirror this result.

The three significant effects on sleeping under a treated bednet, adequate water treatment, and practice of open defecation found in the main analysis lost significance when we controlled for baseline characteristics as a robustness check (Model 2) and in the difference-in-differences model (Model 3).

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 to participate regularly in meetings. While other self-help groups existed, *Gram Varta* was only implemented in groups led by Jeevika. Of all women respondents, 43.1 percent were a member of any self-help group, 92.1 percent of these reported that at least one of their groups was led by Jeevika, and 86.2 percent indicated to participate regularly in Jeevika-led groups. We then repeated the main analysis for this subgroup of exposed households. In alignment with the main analysis, this exercise did not reveal consistent effects on outcomes (see table S4 in the supplementary material).

VI. Discussion

Overall, there was little to no evidence that *Gram Varta* had the expected impact on HNWASH knowledge, practices, attitudes, and health outcomes in Madhepura, Bihar. Two months after the completion of the meeting cycle (about one year after the start of the intervention) no

changes were found regarding HNWASH attitudes, knowledge, and behaviour. As changes in knowledge and behaviour are necessary conditions for a change in health outcomes, the absence of programme impacts on the latter it is not surprising.

These results are not in line with previous studies that found a positive impact of participatory interventions in women's groups on HNWASH related outcomes, such as knowledge of diseases, breast-feeding, and morbidity (Fottrell et al., 2013; Lewycka et al., 2013; Saggurti et al., 2018; Younes et al., 2014). The missing impact on observed health outcomes further contradicts earlier findings of reductions in child mortality (Prost et al., 2013; Tripathy et al., 2016).

Previous trials and reviews have identified factors that influence the success of PLA interventions. One barrier to successful scale-up identified by the literature is low coverage, meaning a large population per self-help group (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). Three of the six blocks included in the evaluation district had similar or lower populations per self-help group compared to the successful trial in Bangladesh (Fottrell et al., 2013). This coverage was achieved by working with the existing structure of state-sponsored self-help groups already in place, which also allowed programme implementation on a larger scale (i.e. geographic scope) than any previous trial, reaching a population of 1,071,348 in the evaluation district alone. Actual programme implementation took place on an even larger scale than the RCT implemented for programme evaluation, spanning more than half of Bihar's state territory.

Building on this structure of self-help groups was one major difference to previous successful PLA trials. One of Jeevika's goals is to empower women and to encourage members to engage in collective action (Hoffmann et al., 2020). As we show in the theory of change, which was developed for the purpose of programme evaluation, empowerment and collective action are key mechanisms for the PLA approach to lead to wider impact (Figure S1). However, Jeevikaled self-help groups were formed with a focus on microcredit activities and health was not a focal topic of these self-help groups originally, potentially weakening the link between participatory learning and the action required for change in health-related outcomes. 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). While facilitators were encouraged to invite target groups, i.e. women of reproductive age, to the self-help group meetings, the make-up of the original group, who joined the group for a different purpose, may have contributed to the lack of impact. One indication of this is the average age of 35 years among women who self-declared self-help group membership. For women of this age group, topics related to child care may not be of primary interest, reducing potential impact. In fact, results from an accompanying qualitative study suggest that younger women were more open to activities and ideas shared in the meetings, and in particular those that challenge traditional norms. However, according to the qualitative work, self-help group members dominated the meetings, leaving less space for engagement of other attendants. The impact on outcomes related to neonatal care would then require that knowledge gained in the meetings is shared within families and with younger women in the community. However, the qualitative results suggest that information was rarely shared with people outside the self-help groups.

An additional difference in programme design is that *Gram Varta* did not include a component of strengthening existing health services as other interventions did. Functioning and responsive health services, however, are a prerequisite for intervention success if participants demand changes as part of the identified solution strategies. 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). This suggests the scale-up of interventions requires performance strengthening of the health system (Hanson et al., 2003).

Low participation of community workers in self-help group meetings may pose an additional reason for the lack of impact of *Gram Varta* which potentially had interacting effects with the absence of improvements in the health system. Engagement of 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, such as ASHAs (Tripathy et al., 2016). A review of successful PLA interventions even suggests community health workers as group facilitators for scale up (Houweling et al., 2019). In the case of *Gram Varta*, PLA meetings were facilitated by Jeevika staff, who were recruited for this task hastily and only had to meet the requirements of being female and literate. On average, facilitators were 27 years old and two thirds of facilitators had completed ten or more years of schooling. One third of facilitators had not worked as community mobilizer before the start of *Gram Varta*, pointing towards a lack of experience in this position. A study of Jeevika's impact on borrowing patterns in Bihar argues that a hasty pace of self-help group formation led to a reduced focus on collective action by facilitators and therefore weakened programme impact (Hoffmann et al., 2020). This effect may be mirrored in the hasty recruitment and training of *Gram Varta* facilitators. Facilitator motivation and her relationship with the community have been shown to enable the success of 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*. Insights from qualitative interviews indeed reveal that in villages where the facilitator was young and unmarried, women were less willing to listen to her, ask her questions, and take her seriously in comparison to villages where the facilitator was married. In some cases, facilitators did not show the necessary motivation to fulfil their tasks. Reasons for the lack of motivation may have been delayed honorarium payments and irregular schedules of recruitment and training of facilitators.

The implementation of *Gram Varta* was deficient in further ways. Several meetings in the PLA cycle were delayed and took place very irregularly. This was due to funding delays, staffing delays, change of government officers in charge of this programme, local elections, and the co-opting of *Gram Varta* functionaries by other government programmes. As per monthly progress reports, a few meetings were held after long delays followed by several meetings in quick succession, which may have deteriorated learning and lowered the motivation among participants and facilitators. These and more factors indicate a rather poor quality of implementation of *Gram Varta* and directly speak to limitations and challenges that had been discussed in the context of intervention scale-ups more generally (Mangham & Hanson, 2010; Milat et al., 2015; Yamey, 2012).

Gram Varta's poor implementation is also reflected in participants' programme awareness. In the treatment area, only 10.7 percent of women 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 rather low in other trials (Prost et al., 2013), *Gram Varta* ranks at the lower end in this regard. However, these awareness rates stand in contrast with women's self-reported self-help group membership of 43 percent, and 60 percent of those claimed to attend their group's normal meetings regularly. Potentially, respondents just did not associate PLA meetings with *Gram Varta* and misreported attendance when asked about *Gram Varta* meetings, even if they had participated in any of them. In fact, in qualitative interviews, women who reported not to have heard of *Gram Varta*, described meetings that followed the PLA approach. Further, funding constraints prohibited the provision of materials such as posters or flyers to improve the public visibility of the programme's name.

Although our study contributes novel and important evidence for the debate on scaling-up public health interventions, the analysis suffers from multiple limitations. First, among the potential factors that could explain *Gram Varta's* lack of impact, i.e., no strengthening of the health system, the use of existing self-help groups, and faulty implementation, we cannot clearly identify which of these factors specifically cause the absence of treatment effects. In part the design of the programme causes *Gram Varta* to differ from previous trials but also allowed its implementation on the unprecedented large scale in the first place. Challenges during programme implementation, which seem to have been more serious than in other trials, arose precisely because of the scale of the programme. Larger, state-implemented programmes require more coordination and oversight due to a plethora of actors and greater funding volumes. The results of this study highlight the increased implementation challenges and the implications these challenges have for the quality of the services provided that scaled-up programmes have.

Second, logistic challenges during the baseline data collection required a shortening and streamlining of questionnaires towards the endline. As a consequence, the analysis omitted several outcomes specified in the pre-analysis plan in order to shorten interviews or because they had shown no variation in the baseline data. Moreover, while we collected additional data for the purpose of evaluating *Gram Varta*, these were not discussed in this manuscript to maintain a focus on HNWASH outcomes. Specifically, this includes additional questions on female empowerment as well as surveys of facilitators and AWWs. An additional survey of a separate sample of pregnant women covered some of the same outcomes analysed in this study. For reasons of space, the analysis of this sample is not described here, however results were very similar. We also chose not to adjust for multiple hypotheses testing, because, given the absence of systematic effects, false positives are not a concern and adjustment for multiple hypotheses testing cannot influence our conclusions.

A further limitation relates to the lack of detailed data on actual self-help group participation. Note that our analysis is foremost an intention-to-treat analysis. Power calculations show that we are able to detect small effects with this approach, while we would only be able to detect large effects on those who actively participated in *Gram Varta* meetings. We maintain that an intention-to-treat perspective is most informative to evaluate overall programme effects (including spill-over effects within communities). Nevertheless, the lack of data on actual participation implies that we cannot conclusively evaluate the effect of *Gram Varta* conditional on participation. However, the treatment intensity analysis of the subgroup 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.

Moreover, in contrast to evaluations of other PLA trials, we did not collect data on neonatal or maternal mortality. However, we do not expect a reduction in mortality as found in previous studies, because we did not observe a change in behaviour and observed health outcomes.

Lastly, most indicators were self-reported and social desirability bias may be a concern. Data from qualitative interviews carried out in parallel suggested that while levels of knowledge about expected behaviours might have increased, norms had not changed to the extent that respondents might feel judged when reporting changes in behavior. In addition, low *Gram Varta* awareness suggests similar levels of desirability bias across treatment groups. We are therefore confident about the reliability of self-reports.

VII. Conclusion

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. Despite the support of government-supported institutions and initial political commitment, practical challenges easily interrupted programme implementation. Large scale interventions require a secure flow of funds, forward-looking planning for hiring qualified staff on the ground who are responsible for carrying out the programme, and continuity among senior staff to withstand turn-over in political positions. While some challenges could be resolved through additional funds, constraints related to programme management and performance may require a strengthening of procedures and systems (Hanson et al., 2003). Leveraging options in the existing system, for instance, by providing greater flexibility in moving funds and hiring, may be a first step. Future research is required to investigate the importance of specific aspects of programme design, such as the relevance of member composition if existing self-help groups with a focus on microfinance are used for a health intervention.

References

- Atkinson, J.-A. M., Fitzgerald, L., Toaliu, H., Taleo, G., Tynan, A., Whittaker, M., Riley, I., & Valley, 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](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
- Bommer, C., Subramanyam, M., & Vollmer, S. (2015a). *Evaluation of Gram Varta in rural Bihar—A mixed methods study. Analysis plan* [Internal document].
- Bommer, C., Subramanyam, M., & Vollmer, S. (2015b). *Evaluation of Gram Varta in rural Bihar—A mixed methods study. Baseline report* [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](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
- Duflo, E., Glennerster, R., & Kremer, M. (2007). Using Randomization in Development Economics Research: A Toolkit. In T. Schultz & J. Strauss (Eds.), *Handbook of Development Economics* (Vol. 4, pp. 3895–3962). North Holland. <http://www.sciencedirect.com/science/article/pii/S1573447107040612>
- 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

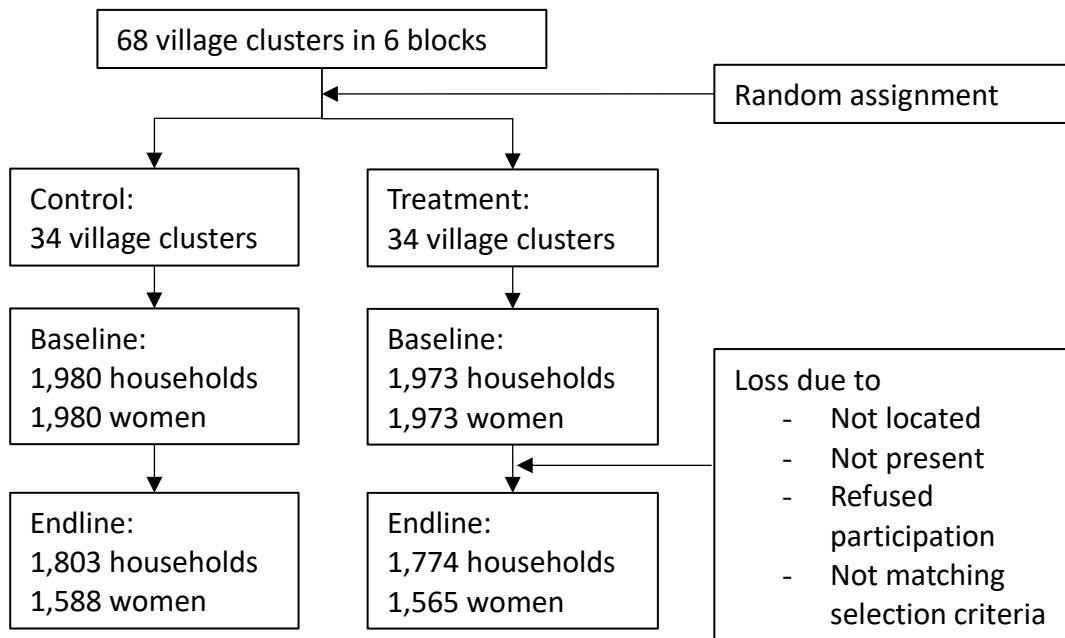
and Absorbable Zinc. *Food and Nutrition Bulletin*, 30(1_suppl1), S108–S143.
<https://doi.org/10.1177/15648265090301S107>

- 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. *Cochrane 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>
- Houweling, T. A. J., Looman, C. W. N., Azad, K., Das, S., King, C., Kuddus, A., Lewycka, S., Manandhar, D. S., Sah More, N., Morrison, J., Phiri, T., Rath, S., Rosato, M., Sen, A., Tripathy, P., Prost, A., Osrin, D., & Costello, A. (2019). The equity impact of community women's groups to reduce neonatal mortality: A meta-analysis of four cluster randomized trials. *International Journal of Epidemiology*, 48(1), 168–182.
<https://doi.org/10.1093/ije/dyx160>
- Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1), 5–86.
- International Institute for Population Sciences (IIPS) & Macro International. (2016). *National Family Health Survey (NFHS-4), 2015-16: State Fact Sheet, Bihar*. IIPS.
- 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>
- Lewycka, S., Mwansambo, C., Rosato, M., Kazembe, P., Phiri, T., Mganga, A., Chapota, H., Malamba, F., Kainja, E., Newell, M.-L., Greco, G., Pulkki-Brännström, A.-M., Skordis-Worrall, J., Vergnano, S., Osrin, D., & Costello, A. (2013). Effect of women's groups and volunteer peer counselling on rates of mortality, morbidity, and health behaviours in mothers and children in rural Malawi (MaiMwana): A factorial, cluster-randomised controlled trial. *The Lancet*, 381(9879), 1721–1735. [https://doi.org/10.1016/S0140-6736\(12\)61959-X](https://doi.org/10.1016/S0140-6736(12)61959-X)
- 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>
- 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>
- Minkler, M., & Cox, K. (1980). Creating Critical Consciousness in Health: Applications of Freire's Philosophy and Methods to the Health Care Setting. *International Journal of Health Services*, 10(2), 311–322. <https://doi.org/10.2190/023K-58E7-6TM8-3RRM>
- 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](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](https://doi.org/10.1016/S0140-6736(13)60685-6)
- Saggurti, N., Atmavilas, Y., Porwal, A., Schooley, J., Das, R., Kande, N., Irani, L., & Hay, K. (2018). Effect of health intervention integration within women's self-help groups on collectivization and healthy practices around reproductive, maternal, neonatal and child health in rural India. *PLoS ONE*, 13(8). <https://doi.org/10.1371/journal.pone.0202562>
- 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](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](https://doi.org/10.1016/S2214-109X(15)00287-9)
- Wallerstein, N., & Bernstein, E. (1988). Empowerment Education: Freire's Ideas Adapted to Health Education. *Health Education Quarterly*, 15(4), 379–394.
<https://doi.org/10.1177/109019818801500402>
- Wallerstein, N., & Bernstein, E. (1994). Introduction to community empowerment, participatory education, and health. *Health Education Quarterly*, 21(2), 141–148.
- Wetmore, S. B., & Theron, F. (1998). Community development and research: Participatory learning and action -a development strategy in itself. *Development Southern Africa*, 15(1), 29–54. <https://doi.org/10.1080/03768359808439994>
- 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>
- Younes, L., Houweling, T. A. J., Azad, K., Kuddus, A., Shaha, S., Haq, B., Nahar, T., Hossen, M., Beard, J., Copas, A., Prost, A., Costello, A., & Fottrell, E. (2014). The effect of participatory women's groups on infant feeding and child health knowledge, behaviour and outcomes in rural Bangladesh: A controlled before-and-after study. *J Epidemiol Community Health*, jech-2014-204271. <https://doi.org/10.1136/jech-2014-204271>

Fig. 1. 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.

Table 1. Process indicators

Indicator:	Block:						Total
	Biharianj	Gwalpara	Kumarkhand	Madhepura Sadar	Murliganj	Udakishunganj	
Population*	135534	126020	243629	191375	185631	189159	178558
Number of SHGs	415	310	843	541	670	350	522
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	366

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); 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 2. Sample characteristics at baseline

Characteristic	Number of observations	Mean (Total)	Mean (Control)	Mean (Treatment)	Difference in SD	P-Value
Household composition						
Household size	3953	5.42	5.22	5.62	-0.19	0.0000
Children below 6	3953	0.82	0.72	0.91	-0.19	0.0000
Household assets						
Land ownership (yes/no)	3953	47.13%	46.21%	48.05%	-0.04	0.2475
Types of livestock owned	3953	0.91	0.91	0.91	0.00	0.9367
Number of cattle owned	2765	1.60	1.58	1.61	-0.01	0.7545
Number of asset types owned	3953	4.43	4.32	4.54	-0.11	0.0003
Age of woman						
Age of woman	3875	37.37	37.99	36.76	0.10	0.0016
Share of women aged 18-29	3875	28.90%	26.50%	31.30%	-0.11	0.0010
Share of women aged 30-39	3875	30.79%	30.37%	31.20%	-0.02	0.5760
Share of women aged >40	3875	40.31%	43.13%	37.49%	0.11	0.0003
Education of woman						
No education	3953	66.68%	68.99%	64.37%	0.10	0.0021
Primary completed	3953	19.45%	17.37%	21.54%	-0.11	0.0009
Secondary or higher	3953	13.41%	13.13%	13.68%	-0.02	0.6097
Highest education in household						
No education	3953	11.48%	13.03%	9.93%	0.10	0.0023
Primary completed	3953	40.10%	38.69%	41.51%	-0.06	0.0702
Secondary or higher	3953	48.42%	48.28%	48.56%	-0.01	0.8639
Caste						
Scheduled castes	3953	10.45%	11.36%	9.53%	0.06	0.0593
Scheduled tribes	3953	6.98%	7.12%	6.84%	0.01	0.7310
Other backward classes	3953	38.78%	39.80%	37.76%	0.04	0.1886
General caste category	3953	7.29%	6.36%	8.21%	-0.07	0.0255
Religion						

Hindu	3953	63.07%	63.33%	62.80%	0.01	0.7273
Muslim	3953	26.16%	26.97%	25.34%	0.04	0.2445

Note: The table presents means of covariates at baseline for all households in the baseline sample and separately for control and treatment group, as well as the difference in standard deviations between control and treatment mean. The last column contains the p-value of a t-test of the difference between the mean of the control and the treatment group.

Table 3. Intention to treat effects for HNWASH outcomes of the *Gram Varta* trial

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)	0.90 (654)	0.79 (670)	-0.11 [-0.26,0.04]	-0.20 [-0.39,-0.02]	-0.34 [-1.36,0.68]
IFA tablet/syrup (last born)	32.05% (780)	27.16% (777)	-4.90 [-9.87,0.08]	-9.01 [-15.78,-2.24]	-2.44 [-29.22,24.34]
Received vitamin A dose (any child)	47.17% (1183)	45.94% (1232)	-1.23 [-6.58,4.13]	-2.42 [-9.04,4.20]	-1.83 [-28.42,24.75]
Multivitamin tablet/syrup (other child)	49.63% (540)	48.50% (567)	-1.13 [-8.33,6.07]	3.28 [-5.39,11.94]	-3.64 [-49.26,41.99]
Nutritional beliefs and practice					
<i>Index of nutrition</i>	58.17% (681)	58.12% (690)	-0.05 [-1.45,1.35]	0.31 [-1.10,1.72]	2.87 [-2.61,8.35]
Balanced food important	97.94% (1748)	97.86% (1733)	-0.08 [-1.14,0.98]	-0.39 [-1.52,0.73]	0.93 [-1.50,3.36]
Feeding thick breastmilk important	93.62% (815)	95.23% (818)	1.61 [-0.71,3.93]	2.01 [-0.64,4.67]	7.54 [-0.22,15.30]
Breastfed child within 24h (last born)	85.83% (847)	87.84% (839)	2.01 [-1.44,5.46]	4.20 [0.18,8.22]	9.30 [-1.86,20.46]
Complementary feeding after 6m important	95.70% (837)	95.45% (835)	-0.25 [-2.43,1.93]	-0.49 [-2.95,1.96]	0.34 [-10.31,10.98]
Gave pre-lacteal feeding at first day (last born)	30.63% (839)	27.23% (841)	-3.40 [-8.43,1.63]	-3.63 [-9.47,2.22]	-7.11 [-25.15,10.94]
Frequency of breastfeeding (last born)	63.64% (484)	61.87% (493)	-1.77 [-8.02,4.48]	-4.37 [-12.25,3.52]	7.26 [-51.46,65.98]
Days of exclusive breastfeeding (last born)	275.92 (223)	315.51 (253)	39.59 [-24.55,103.73]	9.05 [-55.38,73.47]	-13.30 [-596.54,569.95]
Number of meals per day of child (last born)	2.96 (508)	3.00 (516)	0.04 [-0.17,0.25]	0.04 [-0.21,0.28]	0.20 [-0.75,1.15]
Adding oil to meal of 6-59 months old important	79.49% (790)	74.45% (779)	-5.04 [-12.23,2.15]	-3.64 [-11.12,3.83]	-5.51 [-29.34,18.33]
Prevention of diseases					
<i>Index of disease prevention</i>	46.69% (846)	46.72% (842)	0.03 [-1.53,1.60]	-0.44 [-2.27,1.38]	1.51 [-1.98,5.01]
All vaccinations (any child)	59.25% (643)	61.16% (708)	1.90 [-4.64,8.45]	3.93 [-4.17,12.04]	3.19 [-52.21,58.59]
Vaccination card (any child)	66.36% (1418)	67.43% (1437)	1.07 [-4.26,6.40]	1.99 [-3.97,7.94]	-9.78 [-33.53,13.98]
Child sleeps under bednet (last born)	97.91% (860)	97.31% (855)	-0.60 [-2.13,0.94]	-0.36 [-2.07,1.35]	2.52 [-7.41,12.46]
Treated bednet (last born)	18.41% (766)	12.66% (766)	-5.74 [-10.74,-0.75]	-5.37 [-10.96,0.23]	-6.37 [-24.09,11.35]
Deworming past 6m (any child)	48.71% (1400)	50.71% (1406)	2.00 [-3.07,7.06]	-0.39 [-6.74,5.96]	2.83 [-19.62,25.28]

Water treatment					
Treats drinking water	3.39% (1800)	4.74% (1772)	1.35 [-0.19,2.89]	0.68 [-1.06,2.43]	-0.36 [-3.07,2.35]
Water treatment is adequate	2.61% (1800)	4.18% (1772)	1.56 [0.23,2.90]	1.25 [-0.30,2.80]	0.48 [-1.85,2.80]
Hygiene					
<i>Index of hygiene</i>	42.74% (1551)	43.12% (1549)	0.39 [-1.31,2.08]	-0.56 [-2.18,1.07]	-1.17 [-3.91,1.57]
Buys soap at least every 2 months	62.68% (1774)	59.24% (1737)	-3.44 [-8.09,1.21]	-5.29 [-10.05,-0.52]	-4.86 [-11.87,2.15]
Improved toilet at home	18.67% (1741)	19.74% (1707)	1.07 [-2.76,4.91]	-0.90 [-4.69,2.89]	-2.15 [-5.82,1.52]
Subjective cleanliness of toilet	69.88% (342)	69.32% (352)	-0.56 [-9.19,8.06]	-2.57 [-11.56,6.42]	-6.07 [-26.08,13.94]
Observed: stool piles	53.75% (1801)	51.38% (1773)	-2.37 [-8.93,4.20]	-2.05 [-9.20,5.11]	4.08 [-8.41,16.57]
Observed: Sewage water	24.94% (1800)	23.56% (1774)	-1.38 [-7.27,4.51]	-1.33 [-6.87,4.21]	2.79 [-7.40,12.98]
Disposes infant stool adequately	36.88% (320)	39.49% (314)	2.62 [-6.73,11.97]	3.64 [-6.55,13.83]	-1.94 [-31.06,27.18]
Household practices ODF but toilet available	17.23% (325)	11.57% (337)	-5.66 [-10.93,-0.39]	-5.08 [-11.12,0.95]	8.27 [-9.80,26.34]
Reprehends others for ODF	49.27% (1792)	49.01% (1763)	-0.27 [-5.80,5.27]	-2.46 [-8.28,3.35]	-2.05 [-12.90,8.79]
Believes ODF is health hazard	84.75% (1718)	86.00% (1700)	1.25 [-1.60,4.10]	-0.35 [-3.91,3.20]	3.71 [-1.63,9.05]
Toilet/covering excreta important	89.52% (1680)	90.66% (1670)	1.13 [-1.60,3.87]	0.63 [-1.63,2.89]	-2.31 [-8.83,4.20]
Uses soap after toilet	54.13% (1803)	54.23% (1774)	0.10 [-5.56,5.75]	0.00 [-6.25,6.26]	0.72 [-8.42,9.85]
Uses soap before meal	23.18% (1803)	21.98% (1774)	-1.20 [-5.41,3.01]	-1.85 [-6.76,3.06]	-1.30 [-10.37,7.76]
Health knowledge					
<i>Index of health knowledge</i>	42.43% (1334)	43.76% (1355)	1.34 [-1.85,4.52]	-2.44 [-6.19,1.30]	-5.50 [-15.37,4.37]
Heard of message: malaria/dengue	51.74% (1494)	55.15% (1494)	3.41 [-0.57,7.40]	-1.10 [-6.01,3.82]	-2.00 [-14.99,10.98]
Heard of message: diarrhea	56.85% (1497)	55.50% (1499)	-1.34 [-5.09,2.40]	-3.59 [-8.21,1.02]	-8.19 [-20.42,4.05]
Heard of message: ARI	47.83% (1478)	50.72% (1467)	2.88 [-1.18,6.94]	0.20 [-4.80,5.20]	-2.62 [-14.88,9.65]
Heard of message: STI	17.47% (1374)	18.37% (1383)	0.90 [-3.01,4.81]	-1.56 [-6.23,3.12]	-4.72 [-15.69,6.25]
Knows danger signs of malaria	49.94% (1588)	50.61% (1565)	0.67 [-3.50,4.84]	-4.24 [-9.29,0.81]	-3.51 [-15.89,8.87]
Knows danger signs of ARI	39.29% (1588)	38.66% (1565)	-0.64 [-5.07,3.80]	-3.78 [-8.92,1.37]	-7.53 [-18.21,3.15]
Knows treatment of diarrhea	36.15% (1588)	35.46% (1565)	-0.68 [-4.79,3.42]	-3.72 [-8.52,1.09]	-5.99 [-16.72,4.74]
Women's and child's health, self-reported					

Self-assessed health is good	52.69% (1579)	54.15% (1555)	1.46 [-3.39,6.30]	0.43 [-5.22,6.07]	0.43 [-12.11,12.97]
Feels chronically tired	77.59% (1575)	75.43% (1555)	-2.15 [-5.68,1.37]	-2.86 [-6.90,1.18]	5.59 [-5.67,16.84]
Child had no diarrhea in past 3m (any child)	80.56% (1404)	83.11% (1409)	2.55 [-0.99,6.10]	3.85 [-0.63,8.34]	0.49 [-18.24,19.23]
Child had no ARI in past 3m (any child)	88.16% (1393)	89.10% (1404)	0.95 [-2.83,4.73]	1.83 [-2.31,5.97]	-1.73 [-18.64,15.18]
Anthropometric measurement					
<i>Index of child health</i>	24.73% (372)	24.16% (365)	-0.58 [-3.70,2.54]	-0.27 [-3.90,3.36]	n/a
<i>CIAF</i>	65.61% (660)	62.08% (712)	-3.53 [-9.26,2.20]	-4.65 [-11.17,1.86]	n/a
Weight, female 20-49	45.03 (1321)	45.50 (1404)	0.47 [-0.26,1.20]	0.12 [-0.67,0.90]	n/a
Hemoglobin, female 20-49	12.03 (1245)	12.03 (1327)	-0.01 [-0.15,0.14]	0.01 [-0.14,0.17]	n/a
Weight, male 20-49	54.69 (518)	55.22 (540)	0.53 [-0.79,1.85]	-0.51 [-1.87,0.84]	n/a
Hemoglobin, male, 20-49	13.94 (494)	14.01 (511)	0.08 [-0.17,0.32]	0.08 [-0.21,0.38]	n/a
Weight, U5	10.58 (806)	10.54 (838)	-0.04 [-0.35,0.27]	-0.02 [-0.39,0.35]	n/a
Weight-for-age z-score, U5	-1.61 (794)	-1.65 (828)	-0.04 [-0.20,0.11]	0.05 [-0.14,0.24]	n/a
Underweight, U5	37.78% (794)	38.41% (828)	0.62 [-4.63,5.88]	-1.02 [-7.77,5.73]	n/a
Wasting, U5	17.91% (681)	15.52% (728)	-2.39 [-6.73,1.94]	-2.31 [-7.21,2.60]	n/a
Height, U5	82.72 (781)	82.08 (849)	-0.64 [-2.22,0.93]	-0.66 [-2.41,1.09]	n/a
Height-for-age z-score, U5	-1.72 (781)	-1.84 (849)	-0.11 [-0.37,0.15]	-0.03 [-0.32,0.25]	n/a
Stunting, U5	50.45% (781)	50.65% (849)	0.20 [-5.52,5.92]	-2.29 [-9.03,4.45]	n/a
Hemoglobin, U5	10.81 (554)	10.91 (559)	0.09 [-0.13,0.32]	0.09 [-0.16,0.34]	n/a
Oedema, U5	1.82% (877)	1.19% (926)	-0.64 [-1.97,0.70]	-0.55 [-2.16,1.06]	n/a
Arm circumference, U5	14.31 (656)	14.39 (668)	0.08 [-0.14,0.31]	0.11 [-0.17,0.38]	n/a
Arm circumference-for-age z score, U5	-1.01 (655)	-0.95 (666)	0.06 [-0.13,0.26]	0.08 [-0.15,0.31]	n/a
Arm circumference < 115 mm, U5	2.90% (655)	1.80% (666)	-1.10 [-2.72,0.52]	-1.58 [-3.39,0.23]	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. Model (3) could not be estimated for anthropometric indicators due to missing panel identification of other household members. Standard errors were clustered at the village 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

Supplementary Materials

Figure S1. Theory of change

Table S1. Meeting agenda of PLA cycle.

Table S2. Description of outcome measures.

Table S3. Outcome measures at baseline

Table S4. Results of the treatment intensity analysis for HNWASH outcomes of the *Gram Varta* trial.

Table S5. Minimum detectable effect sizes

PDF-File. Pre-Analysis plan.

Figure S1. Theory of change

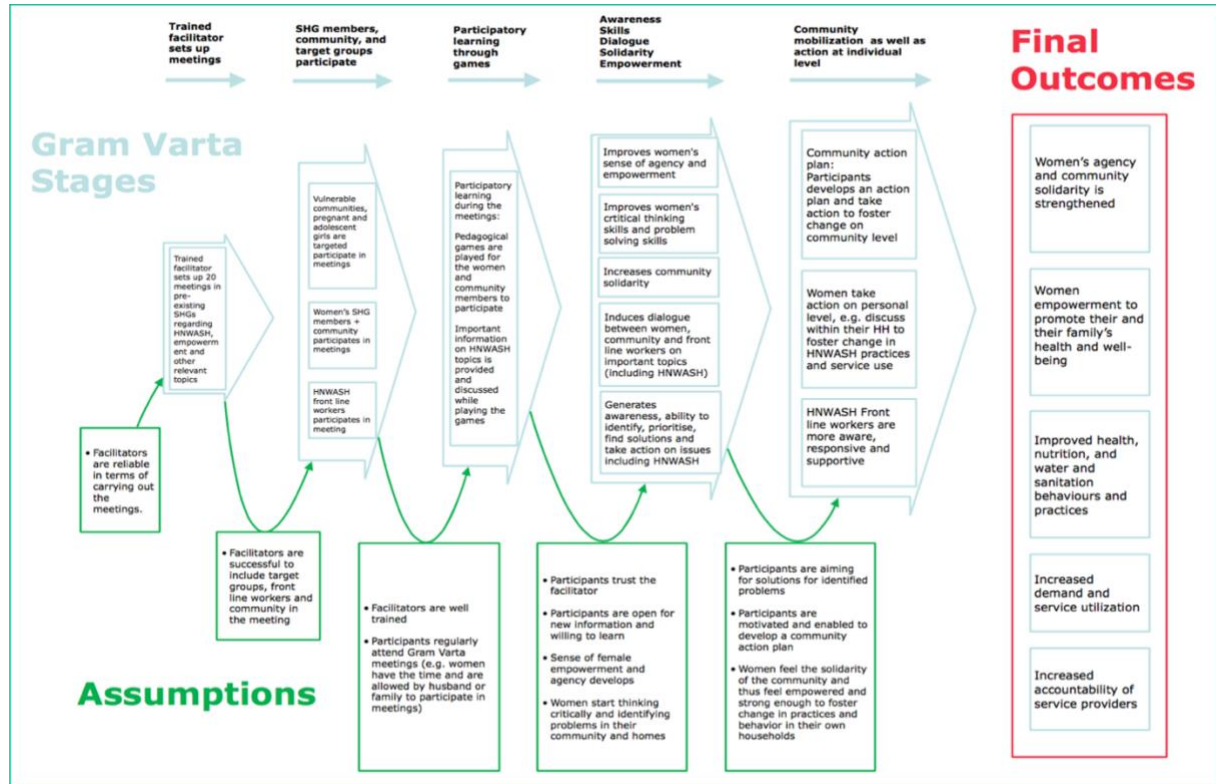


Table S1. Meeting agenda of PLA cycle

Meetings	Agenda
Meeting 1	Introducing the project
Meeting 2	Understanding the concept of equality in society
Meeting 3	Understanding the cycle of women and child malnutrition in society and understanding the current concepts about nutrition in society
Meeting 4	Using the MUAC tape to measure malnutrition and identify malnourished children
Meeting 5	Identifying the issues related to women and child malnutrition and their priority
Meeting 6	Introduction to primary problems, their reasons and their resolution
Meeting 7	Identifying problems and strategies
Meeting 8	Planning a community meeting and taking responsibilities
Meeting 9	Progress of strategy implementation and mapping available local food groups
Meeting 10	Strategy for caring for and providing adequate nutrition to infants (6-24 months)
Meeting 11	Discussion of possible strategies for preventing malnutrition in children
Meeting 12	Understanding the main reasons for malnutrition in mothers and breaking the vicious cycle of malnutrition through possible means
Meeting 13	Taking care of low birth weight children
Meeting 14	Taking care of a newborn, nutritional requirements in neonates
Meeting 15	Discussion on family planning measures - The benefits of various measures and a discussion on the availability of various means of contraception through the government health system
<i>Community directed total sanitation meeting, inviting local authorities and all community members</i>	
Meeting 16	Understanding uncleanliness
Meeting 17	The impact of uncleanliness
Meeting 18	Community directed total sanitation campaign - What to do (action plan)
Meeting 19	Community directed total sanitation campaign - How to do and who will manage
Meeting 20	Participatory evaluation and planning for the future

Table S2. Description of outcome measures

Measure/indicator	Description	Respondent	Format
Intake of micronutrients			
Number of vitamin A doses (last born)	Number of vitamin A doses received by respondent's last-born child	Woman	integer
IFA tablet/syrup (last born)	Whether respondent's last-born child received an IFA tablet or syrup	Woman	binary
Received vitamin A dose (any child)	Whether child (last-born or other) received any vitamin A dose	Woman	binary
Multivitamin tablet/syrup (other child)	Whether child (other than last-born) received a multivitamin tablet or syrup	Woman	binary
Nutritional beliefs and practice			
<i>Index of nutrition</i>	Index of nutritional beliefs and practice, scale 0 - 1		integer
Balanced food important*	Whether respondent thinks that a good balance of foods (pulses, vegetables, chapati, milk, curd) that helps growth, provide energy, and protect health is important	Household Head	binary
Feeding thick breastmilk important*	Whether respondent thinks that feeding the thick yellowish breastmilk of the first three days to babies is important	Woman	binary
Breastfed child within 24h (last born)*	Whether respondent breastfed her last-born child within 24 hours after birth	Woman	binary
Complementary feeding after 6m important*	Whether respondent thinks that starting complementary feeding of babies after they are six months old is important	Woman	binary
Gave pre-lacteal feeding at first day (last born)*	Whether respondent gave pre-lacteal feeding on the first day of last born's birth (e.g. honey, jaggery, ghee or ghutti)	Woman	binary
Frequency of breastfeeding (last born)	Whether respondent breastfeeds her last born 7 times or more per day	Woman	binary
Days of exclusive breastfeeding (last born)	How many days the last-born child was exclusively breastfed	Woman	integer
Number of meals per day of child (last born)	Number of times the last born ate solid, semi-solid, or soft foods the past day (during day or at night)	Woman	integer
Adding oil to meal of 6-59 months old important*	Whether respondent thinks that adding extra oil to every meal given to children who are 6 months to 59 months old is important	Woman	binary
Prevention of diseases			
<i>Index of disease prevention</i>	Index of disease prevention, scale 0 - 1		integer
All vaccinations (any child)	Whether child (last born or other) received all recommended vaccinations	Woman	binary
Vaccination card (any child)*	Whether child (last born or other) possesses a vaccination card	Woman	binary
Child sleeps under bednet (last born)*	Whether last-born child sleeps under a bednet	Woman	binary
Treated bednet (last born)	Whether last-born child sleeps under an Insecticide-treated Bednet (ITN)	Woman	binary
Deworming past 6m (any child)*	Whether child (last-born or other) was given any drug for intestinal worms in the past six months	Woman	binary

Water treatment			
Treats drinking water	Whether the household treats drinking water in any way	Household Head	binary
Water treatment is adequate	Whether the household treats drinking water adequately (boiling, adding chlorine/bleach tablets, using water filter, using electric purifier, keeping water covered, keeping it clean around the hand pump/tap)	Household Head	binary
Hygiene			
<i>Index of hygiene</i>	Index of hygiene, scale 0 - 1		integer
Buys soap at least every 2 months*	Whether household buys soap at least every 2 months	Household Head	binary
Improved toilet at home*	Whether household has an improved toilet at home (flush toilet, piped sewer system, septic tank, flush/pour flush to pit latrine, ventilated improved pit latrine, pit latrine with slab, composting toilet)	Household Head	binary
Subjective cleanliness of toilet	Whether enumerator observed toilet to be clean or very clean (as opposed to dirty or very dirty)	Observation	binary
Observed: stool piles*	Whether enumerator observed stool piles (faeces) nearby the household	Observation	binary
Observed: Sewage water*	Whether enumerator observed sewage water nearby the household	Observation	binary
Disposes infant stool adequately	Whether household disposes the stool of children adequately (rinsed into toilet/latrine, rinsed into drain/ditch, buried)	Household Head	binary
Household practices ODF but toilet available	Whether there any household members who defecate in the open even though you have a toilet facility is accessible	Household Head	binary
Reprehends others for ODF*	Whether respondent reprehended the individual, if he/she has seen someone openly defecating near his/her house	Household Head	binary
Believes ODF is health hazard*	Whether respondent believes open defecation represents a health hazard to him/her and his/her family?	Household Head	binary
Toilet/covering excreta important*	Whether respondent thinks that using a toilet for defecation, or, until a toilet is constructed, ensuring that excreta is covered with soil is important	Household Head	binary
Uses soap after toilet*	Whether respondent washes his/her hands with soap after using the toilet	Household Head	binary
Uses soap before meal*	Whether respondent washes his/her hands with soap before meals	Household Head	binary
Health knowledge			
<i>Index of health knowledge</i>	Index of health knowledge, scale 0 - 1		integer
Heard of message: malaria/dengue*	Whether respondent has seen/heard/read messages related to malaria/dengue programme	Woman	binary
Heard of message: diarrhea*	Whether respondent has seen/heard/read messages related to diarrhoea	Woman	binary
Heard of message: ARI*	Whether respondent has seen/heard/read messages related to acute respiratory infections/pneumonia	Woman	binary
Heard of message: STI*	Whether respondent has seen/heard/read messages related to sexually transmitted infections	Woman	binary

Knows danger signs of malaria*	Whether respondent could name any correct danger signs of malaria (fever, shaking chills, sweating, headache, vomiting, loose motions)	Woman	binary
Knows danger signs of ARI*	Whether respondent could name any correct danger signs of malaria (pain in chest or (productive) cough, difficulty in breathing, drowsy/difficulty to keep awake, fever, rapid breathing)	Woman	binary
Knows treatment of diarrhea*	Whether respondent could name any correct treatment for diarrhoea (give ORS solution, salt and sugar solution, plenty of fluids, continue breastfeeding, zinc, lime solution)	Woman	binary
Women's and child's health, self-reported			
Self-assessed health is good	Whether respondent rates her health as good or very good (as opposed to very bad, bad, or moderate)	Woman	binary
Feels chronically tired	Whether respondent feels chronically tired	Woman	binary
Child had no diarrhea in past 3m (any child)	Whether child (last-born or other) had no diarrhoea in past 3 months	Woman	binary
Child had no ARI in past 3m (any child)	Whether child (last-born or other) had no acute respiratory infection/pneumonia/fever in past 3 months	Woman	binary
Anthropometric measurement			
<i>Index of child health</i>	Index of child health, scale 0 – 1 (1 being negative)		integer
<i>CIAF</i>	Composite Index of Anthropometric Failure (1 being negative), based on wasting, underweight, stunting	Measurement	binary
Weight, female 20-49	Weight in kg, female household members aged 20 - 49 years, measured by enumerator	Measurement	continuous
Hemoglobin, female 20-49	Hemoglobin level in g/dL, female household members aged 20 - 49 years, measured by enumerator	Measurement	continuous
Weight, male 20-49	Weight in kg, male household members aged 20 - 49 years, measured by enumerator	Measurement	continuous
Hemoglobin, male, 20-49	Hemoglobin level in g/dL, male household members aged 20 - 49 years, measured by enumerator	Measurement	continuous
Weight, U5	Weight in kg, all children below 5 years, measured by enumerator	Measurement	continuous
Weight-for-age z-score, U5	Weight-for-age z-score, all children below 5 years	Measurement	continuous
Underweight, U5*	Whether weight-for-age z-score is below -2, all children below 5 years	Measurement	binary
Wasting, U5*	Whether weight-for-height z-score is below -2, all children below 5 years	Measurement	binary
Height, U5	Height in cm, all children below 5 years, measured by enumerator	Measurement	continuous
Height-for-age z-score, U5	Height-for-age z-score, all children below 5 years	Measurement	continuous
Stunting, U5*	Whether height-for-age z-score is below -2, all children below 5 years	Measurement	binary
Hemoglobin, U5*	Hemoglobin level in g/dL, all children below years, measured by enumerator	Measurement	continuous
Oedema, U5*	Whether oedema are detected by enumerator, all children below 5 years	Measurement	binary
Arm circumference, U5	Arm circumference in cm, all children below 5 years, measured by enumerator	Measurement	continuous

Arm circumference-for-age z score, U5	Arm circumference-for-age z-score, all children below 5 years	Measurement	continuous
Arm circumference < 115 mm, U5*	Whether arm circumference is less than 115 mm, all children below 5 years	Measurement	binary

Note: * Outcome was included in index was for this outcome group.

Table S3. Outcome measures at baseline

Measure/indicator	Number of observations	Mean (Total)	Mean (Control)	Mean (Treatment)	Difference in SD	P-Value
Intake of micronutrients						
Number of vitamin A doses (last born)	1386	0.77	0.71	0.82	-0.09	0.0795
IFA tablet/syrup (last born)	1791	28.48%	27.41%	29.38%	-0.04	0.3562
Received vitamin A dose (any child)	2373	40.88%	40.62%	41.10%	-0.01	0.8131
Multivitamin tablet/syrup (other child)	1035	38.36%	38.08%	38.57%	-0.01	0.8746
Nutritional beliefs and practice						
<i>Index of nutrition</i>	1551	53.93%	54.20%	53.66%	0.05	0.3012
Balanced food important	3823	97.31%	97.60%	97.00%	0.04	0.2549
Feeding thick breastmilk important	1860	93.76%	96.04%	91.81%	0.18	0.0002
Breastfed child within 24h (last born)	1927	88.79%	90.48%	87.39%	0.10	0.0325
Complementary feeding after 6m important	1835	86.98%	86.12%	87.72%	-0.05	0.3107
Gave pre-lacteal feeding at first day (last born)	1890	44.44%	44.48%	44.41%	0.00	0.9753
Frequency of breastfeeding (last born)	1298	47.15%	49.26%	45.30%	0.08	0.1539
Days of exclusive breastfeeding (last born)	549	611.13	625.53	599.08	0.08	0.3247
Number of meals per day of child (last born)	1544	2.02	2.12	1.93	0.12	0.0207
Adding oil to meal of 6-59 months old important	1657	65.48%	61.95%	68.76%	-0.14	0.0035
Prevention of diseases						
<i>Index of disease prevention</i>	1890	40.52%	41.23%	39.92%	0.09	0.0473
All vaccinations (any child)	1120	74.02%	75.25%	73.02%	0.05	0.3982
Vaccination card (any child)	1927	74.88%	73.62%	75.92%	-0.05	0.2467
Child sleeps under bednet (last born)	1927	91.13%	93.46%	89.19%	0.15	0.0010
Treated bednet (last born)	1619	1.48%	0.90%	2.03%	-0.09	0.0596
Deworming past 6m (any child)	2938	29.34%	30.94%	28.03%	0.06	0.0853
Water treatment						
Treats drinking water	3949	2.28%	1.41%	3.15%	-0.12	0.0003
Water treatment is adequate	3949	1.57%	1.06%	2.08%	-0.08	0.0099
Hygiene						
<i>Index of hygiene</i>	3645	55.47%	54.55%	56.48%	-0.20	0.0000
Buys soap at least every 2 months	3939	97.66%	96.91%	98.42%	-0.10	0.0017
Improved toilet at home	3953	13.33%	11.87%	14.80%	-0.09	0.0067

Subjective cleanliness of toilet	536	86.75%	85.65%	87.63%	-0.06	0.5046
Observed: stool piles	3928	17.36%	20.34%	14.38%	0.16	0.0000
Observed: Sewage water	3923	11.14%	13.22%	9.05%	0.13	0.0000
Disposes infant stool adequately	3238	6.86%	5.35%	8.24%	-0.11	0.0011
Household practices ODF but toilet available	536	20.90%	24.05%	18.39%	0.14	0.1101
Reprehends others for ODF	3922	74.20%	73.64%	74.76%	-0.03	0.4249
Believes ODF is health hazard	3860	92.38%	93.38%	91.36%	0.08	0.0178
Toilet/covering excreta important	3746	86.87%	85.10%	88.77%	-0.11	0.0009
Uses soap after toilet	3953	86.29%	86.72%	85.86%	0.02	0.4330
Uses soap before meal	3953	40.88%	40.51%	41.26%	-0.02	0.6307
Health knowledge						
<i>Index of health knowledge</i>	3380	31.03%	28.70%	33.42%	-0.13	0.0002
Heard of message: malaria/dengue	3671	48.35%	46.86%	49.86%	-0.06	0.0685
Heard of message: diarrhea	3661	43.62%	41.33%	45.94%	-0.09	0.0049
Heard of message: ARI	3656	38.18%	35.58%	40.81%	-0.11	0.0011
Heard of message: STI	3387	15.06%	12.66%	17.51%	-0.14	0.0001
Knows danger signs of malaria	3953	39.56%	38.38%	40.75%	-0.05	0.1283
Knows danger signs of ARI	3953	24.03%	21.46%	26.61%	-0.12	0.0002
Knows treatment of diarrhea	3953	25.37%	22.68%	28.08%	-0.12	0.0001
Women's and child's health, self-reported						
Self-assessed health is good	3953	46.55%	45.20%	47.90%	-0.05	0.0895
Feels chronically tired	3950	67.62%	70.20%	65.03%	0.11	0.0005
Child had no diarrhea in past 3m (any child)	3014	67.52%	66.00%	68.74%	-0.06	0.1101
Child had no ARI in past 3m (any child)	2967	78.80%	78.22%	79.28%	-0.03	0.4832
Anthropometric measurement						
<i>Index of child health</i>	73	21.23%	21.51%	21.03%	0.03	0.9045
<i>CIAF</i>	1329	65.54%	68.47%	62.59%	0.12	0.0243
Weight, female 20-49	3416	45.55	45.11	45.98	-0.10	0.0023
Hemoglobin, female 20-49	3337	11.39	11.47	11.31	0.11	0.0015
Weight, male 20-49	1847	54.34	53.41	55.17	-0.18	0.0002
Hemoglobin, male, 20-49	1770	12.97	13.04	12.92	0.07	0.1184
Weight, U5	1890	10.22	9.96	10.46	-0.16	0.0006
Weight-for-age z-score, U5	1745	-1.37	-1.52	-1.22	-0.16	0.0007
Underweight, U5	1745	37.82%	43.16%	32.85%	0.21	0.0000

Wasting, U5	1476	18.22%	20.03%	16.37%	0.09	0.0688
Height, U5	1740	81.52	80.98	82.05	-0.08	0.0807
Height-for-age z-score, U5	1740	-2.01	-2.03	-1.98	-0.02	0.6178
Stunting, U5	1740	54.20%	55.98%	52.45%	0.07	0.1391
Hemoglobin, U5	171	11.15	11.38	10.98	0.32	0.0411
Oedema, U5	2777	5.65%	5.12%	6.09%	-0.04	0.2749
Arm circumference, U5	2108	14.22	14.18	14.26	-0.05	0.2097
Arm circumference-for-age z score, U5	2050	-0.82	-0.84	-0.80	-0.03	0.4729
Arm circumference < 115 mm, U5	2050	2.73%	2.28%	3.16%	-0.05	0.2219

Note: The table presents means of outcome measures at baseline for the total sample and separately for control and treatment group, as well as the difference in standard deviations between control and treatment mean. The last column contains the p-value of a t-test of the difference between the mean of the control and the treatment group.

Table S4. Results of the treatment intensity analysis for HNWASH outcomes of the *Gram Varta* trial

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)	0.73 (179)	0.78 (232)	0.05 [-0.19,0.28]	-0.10 [-0.45,0.25]	-0.13 [-1.35,1.08]
IFA tablet/syrup (last born)	30.73% (218)	27.55% (265)	-3.19 [-12.34,5.97]	-5.07 [-18.03,7.88]	-2.85 [-42.42,36.71]
Received vitamin A dose (any child)	50.85% (352)	45.71% (420)	-5.14 [-14.73,4.45]	-11.20 [-22.52,0.13]	-9.84 [-51.56,31.89]
Multivitamin tablet/syrup (other child)	53.89% (180)	47.62% (189)	-6.27 [-18.12,5.58]	-6.95 [-21.23,7.34]	-7.30 [-67.20,52.60]
Nutritional beliefs and practice					
<i>Index of nutrition</i>	58.29% (179)	58.14% (232)	-0.15 [-2.50,2.21]	-0.58 [-2.99,1.83]	1.37 [-6.97,9.71]
Balanced food important	96.79% (499)	97.25% (510)	0.46 [-1.91,2.84]	0.57 [-2.34,3.48]	0.01 [-5.21,5.24]
Feeding thick breastmilk important	94.20% (224)	95.39% (282)	1.19 [-3.22,5.60]	0.68 [-4.21,5.58]	9.27 [-2.40,20.94]
Breastfed child within 24h (last born)	86.72% (241)	88.85% (287)	2.13 [-3.47,7.73]	3.64 [-3.94,11.21]	7.99 [-8.64,24.63]
Complementary feeding after 6m important	95.76% (236)	97.19% (285)	1.43 [-1.81,4.67]	-0.13 [-4.67,4.42]	-3.16 [-18.54,12.22]
Gave pre-lacteal feeding at first day (last born)	29.71% (239)	28.42% (285)	-1.29 [-9.96,7.39]	3.53 [-7.76,14.82]	-8.71 [-33.78,16.36]
Frequency of breastfeeding (last born)	64.39% (132)	59.42% (138)	-4.97 [-18.36,8.42]	-6.74 [-23.36,9.88]	21.05 [-78.28,120.39]
Days of exclusive breastfeeding (last born)	241.61 (69)	340.82 (97)	99.22 [12.58,185.86]	90.66 [-14.24,195.57]	211.54 [-537.12,960.19]
Number of meals per day of child (last born)	3.12 (161)	3.11 (215)	-0.01 [-0.36,0.34]	0.02 [-0.38,0.42]	0.25 [-1.20,1.69]
Adding oil to meal of 6-59 months old important	84.98% (213)	73.58% (265)	-11.39 [-20.36,-2.42]	-8.45 [-19.74,2.83]	-18.35 [-50.33,13.63]
Prevention of diseases					
<i>Index of disease prevention</i>	46.14% (241)	46.97% (290)	0.82 [-1.91,3.56]	0.19 [-3.14,3.53]	3.91 [-1.79,9.60]
All vaccinations (any child)	54.70% (181)	63.87% (238)	9.17 [-3.10,21.44]	10.55 [-4.69,25.79]	3.09 [-91.03,97.21]
Vaccination card (any child)	61.86% (430)	67.96% (490)	6.10 [-2.24,14.43]	7.64 [-2.06,17.33]	3.14 [-31.83,38.11]
Child sleeps under bednet (last born)	98.36% (244)	97.94% (291)	-0.42 [-2.66,1.81]	-1.64 [-4.87,1.60]	3.03 [-13.10,19.16]
Treated bednet (last born)	13.43% (216)	8.05% (261)	-5.38 [-12.42,1.66]	-2.30 [-10.72,6.13]	-6.09 [-31.24,19.06]
Deworming past 6m (any child)	49.65% (425)	53.73% (482)	4.09 [-4.00,12.18]	2.60 [-8.33,13.54]	1.29 [-33.24,35.82]

Water treatment					
Treats drinking water	2.17% (508)	4.36% (527)	2.20 [-0.31,4.70]	0.93 [-1.09,2.95]	-1.18 [-5.95,3.59]
Water treatment is adequate	1.97% (508)	3.23% (527)	1.26 [-0.74,3.25]	0.35 [-1.56,2.26]	-1.19 [-5.21,2.84]
Hygiene					
<i>Index of hygiene</i>	42.88% (440)	42.41% (470)	-0.47 [-2.76,1.82]	-0.66 [-2.95,1.63]	-2.12 [-6.03,1.79]
Buys soap at least every 2 months	64.73% (499)	58.22% (517)	-6.51 [-13.91,0.90]	-8.17 [-16.07,-0.27]	-8.07 [-19.08,2.94]
Improved toilet at home	17.62% (488)	16.41% (512)	-1.22 [-6.66,4.23]	-2.15 [-8.61,4.32]	-4.98 [-11.69,1.73]
Subjective cleanliness of toilet	65.91% (88)	69.32% (88)	3.41 [-11.84,18.66]	-2.34 [-19.75,15.06]	2.71 [-52.13,57.55]
Observed: stool piles	54.74% (506)	51.89% (528)	-2.85 [-11.84,6.14]	-3.68 [-14.27,6.92]	0.67 [-16.41,17.74]
Observed: Sewage water	24.21% (508)	24.24% (528)	0.03 [-7.68,7.74]	-4.67 [-12.58,3.24]	3.04 [-10.30,16.39]
Disposes infant stool adequately	35.53% (76)	42.67% (75)	7.14 [-8.36,22.64]	23.64 [3.09,44.19]	-1.39 [-51.34,48.57]
Household practices ODF but toilet available	17.44% (86)	9.52% (84)	-7.92 [-18.63,2.79]	-2.80 [-14.75,9.14]	18.27 [-27.42,63.96]
Reprehends others for ODF	48.32% (507)	49.05% (526)	0.73 [-7.13,8.58]	-1.47 [-10.25,7.31]	-3.43 [-17.50,10.64]
Believes ODF is health hazard	83.84% (489)	83.89% (509)	0.05 [-4.91,5.00]	-1.90 [-7.86,4.06]	2.57 [-5.73,10.88]
Toilet/covering excreta important	90.02% (471)	90.12% (506)	0.10 [-3.97,4.16]	0.05 [-4.18,4.29]	-3.66 [-12.53,5.21]
Uses soap after toilet	56.30% (508)	53.98% (528)	-2.32 [-10.37,5.72]	-1.52 [-10.38,7.35]	-1.72 [-14.69,11.24]
Uses soap before meal	22.64% (508)	21.97% (528)	-0.67 [-6.02,4.68]	-3.11 [-10.60,4.39]	-4.12 [-17.03,8.79]
Health knowledge					
<i>Index of health knowledge</i>	45.82% (443)	48.93% (480)	3.10 [-1.81,8.02]	-1.99 [-8.07,4.09]	-2.62 [-15.64,10.40]
Heard of message: malaria/dengue	55.86% (503)	60.26% (531)	4.40 [-2.18,10.97]	-0.56 [-8.71,7.59]	-4.41 [-21.04,12.22]
Heard of message: diarrhea	64.10% (507)	62.20% (537)	-1.91 [-7.91,4.10]	-4.33 [-12.38,3.73]	-9.68 [-26.49,7.14]
Heard of message: ARI	52.89% (501)	55.66% (521)	2.77 [-3.78,9.32]	-1.45 [-10.28,7.39]	-6.77 [-22.94,9.40]
Heard of message: STI	17.14% (455)	20.86% (489)	3.72 [-2.05,9.49]	1.11 [-6.05,8.28]	-1.59 [-16.45,13.26]
Knows danger signs of malaria	53.48% (531)	56.39% (548)	2.90 [-3.07,8.88]	-4.05 [-11.06,2.96]	-3.61 [-18.18,10.96]
Knows danger signs of ARI	44.26% (531)	43.25% (548)	-1.01 [-8.11,6.10]	-4.78 [-13.11,3.56]	-12.00 [-26.17,2.17]
Knows treatment of diarrhea	36.91% (531)	39.60% (548)	2.69 [-4.35,9.73]	-1.55 [-10.22,7.12]	-3.42 [-17.91,11.07]
Women's and child's health, self-reported					

Self-assessed health is good	47.64% (529)	50.83% (545)	3.19 [-4.29,10.66]	0.61 [-8.76,9.98]	0.69 [-14.32,15.69]
Feels chronically tired	80.34% (529)	76.47% (544)	-3.87 [-9.09,1.35]	-2.69 [-9.02,3.65]	1.55 [-12.96,16.07]
Child had no diarrhea in past 3m (any child)	79.67% (423)	81.99% (483)	2.32 [-3.79,8.42]	5.91 [-1.70,13.53]	-0.31 [-29.07,28.44]
Child had no ARI in past 3m (any child)	92.16% (421)	89.42% (482)	-2.74 [-7.95,2.47]	-3.21 [-9.35,2.93]	0.92 [-22.00,23.84]
Anthropometric measurement					
<i>Index of child health</i>	25.79% (106)	22.99% (116)	-2.80 [-8.11,2.51]	-3.47 [-10.43,3.49]	n/a
<i>CIAF</i>	65.07% (209)	61.40% (215)	-3.68 [-13.29,5.93]	-8.56 [-20.76,3.64]	n/a
Weight, female 20-49	44.39 (429)	45.44 (442)	1.04 [-0.09,2.17]	1.13 [-0.10,2.36]	n/a
Hemoglobin, female 20-49	12.01 (408)	12.08 (419)	0.07 [-0.15,0.28]	0.11 [-0.15,0.36]	n/a
Weight, male 20-49	54.74 (148)	56.17 (162)	1.44 [-0.68,3.55]	1.17 [-1.43,3.78]	n/a
Hemoglobin, male, 20-49	13.68 (145)	13.97 (158)	0.29 [-0.11,0.70]	0.46 [-0.03,0.96]	n/a
Weight, U5	10.60 (258)	10.94 (257)	0.34 [-0.23,0.91]	0.44 [-0.22,1.10]	n/a
Weight-for-age z-score, U5	-1.74 (252)	-1.65 (253)	0.09 [-0.20,0.38]	0.37 [0.04,0.70]	n/a
Underweight, U5	40.08% (252)	34.39% (253)	-5.69 [-15.29,3.90]	-14.55 [-26.46,-2.63]	n/a
Wasting, U5	18.06% (216)	18.10% (221)	0.04 [-7.85,7.94]	2.48 [-7.61,12.57]	n/a
Height, U5	83.68 (243)	84.71 (255)	1.03 [-1.59,3.64]	1.27 [-1.50,4.04]	n/a
Height-for-age z-score, U5	-1.66 (243)	-1.76 (255)	-0.11 [-0.57,0.35]	-0.10 [-0.60,0.39]	n/a
Stunting, U5	49.38% (243)	48.24% (255)	-1.15 [-10.56,8.27]	-4.76 [-16.81,7.28]	n/a
Hemoglobin, U5	10.62 (165)	10.93 (176)	0.31 [-0.05,0.68]	0.41 [-0.02,0.84]	n/a
Oedema, U5	2.60% (269)	1.43% (279)	-1.17 [-3.86,1.52]	-1.49 [-5.60,2.62]	n/a
Arm circumference, U5	14.25 (202)	14.32 (214)	0.07 [-0.29,0.42]	0.07 [-0.35,0.50]	n/a
Arm circumference-for-age z score, U5	-1.11 (202)	-1.08 (214)	0.03 [-0.28,0.33]	0.03 [-0.33,0.39]	n/a
Arm circumference < 115 mm, U5	2.48% (202)	1.40% (214)	-1.07 [-3.68,1.53]	-1.93 [-5.52,1.66]	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 who self-identified as active Jeevika-led SHG members. Model (1) is the simple endline comparison of means, Model (2) controls for covariates, Model (3) is the difference-in-differences model. Model (3) could not be estimated for anthropometric indicators due to missing panel identification of other household members. Standard errors were clustered at the village 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

Table S5. Minimum detectable effect sizes

Measure/indicator	(1) Baseline ICC	(2) MDE Baseline Report	(3) Observations per cluster at Endline	(4) Number clusters at Endline	(5) MDE at Endline	(6) MDE at Endline, low compliance	(7) MDE at endline, no compliance correction
Intake of micronutrients							
Number of vitamin A doses (last born)	0,0077	0,1613	7	180	0,2715	0,4735	0,2036
IFA tablet/syrup (last born)	0,1481	0,1086	9	180	0,1238	0,2159	0,0928
Received vitamin A dose (any child)	0,0559	0,0860	13	180	0,0981	0,1711	0,0736
Multivitamin tablet/syrup (other child)	0,2047	0,1329	6	180	0,1572	0,2742	0,1179
Nutritional beliefs and practice							
<i>Index of nutrition</i>	0,0664	0,0189	8	180	0,0245	0,0426	0,0183
Balanced food important	0,0151	0,0221	19	180	0,0233	0,0407	0,0175
Feeding thick breastmilk important	0,0378	0,0384	9	180	0,0512	0,0893	0,0384
Breastfed child within 24h (last born)	0,0722	0,0594	9	180	0,0735	0,1283	0,0552
Complementary feeding after 6m important	0,1304	0,0772	9	180	0,0893	0,1557	0,0669
Gave pre-lacteal feeding at first day (last born)	0,1073	0,1064	9	180	0,1257	0,2192	0,0943
Frequency of breastfeeding (last born)	0,0870	0,0996	5	180	0,1443	0,2517	0,1082
Days of exclusive breastfeeding (last born)	0,0566	54,9568	3	145	118,2451	206,2415	88,6838
Number of meals per day of child (last born)	0,0739	0,3069	6	179	0,4317	0,7529	0,3238
Adding oil to meal of 6-59 months old important	0,1381	0,1114	9	180	0,1280	0,2233	0,0960
Prevention of diseases							
<i>Index of disease prevention</i>	0,0451	0,0236	9	180	0,0309	0,0539	0,0232
All vaccinations (any child)	0,1925	0,1169	8	176	0,1338	0,2333	0,1003
Vaccination card (any child)	0,0411	0,0702	16	180	0,0767	0,1338	0,0575
Child sleeps under bednet (last born)	0,0621	0,0512	10	180	0,0625	0,1090	0,0469
Treated bednet (last born)	0,0381	0,0192	9	180	0,0256	0,0447	0,0192
Deworming past 6m (any child)	0,0579	0,0804	16	180	0,0866	0,1510	0,0650
Water treatment							
Treats drinking water	0,0344	0,0232	20	180	0,0239	0,0417	0,0179
Water treatment is adequate	0,0328	0,0192	20	180	0,0197	0,0344	0,0148
Hygiene							

<i>Index of hygiene</i>	0,0638	0,0175	17	180	0,0185	0,0323	0,0139
Buys soap at least every 2 months	0,0400	0,0243	20	180	0,0249	0,0435	0,0187
Improved toilet at home	0,0580	0,0601	19	180	0,0621	0,1083	0,0466
Subjective cleanliness of toilet	0,0073	0,0432	4	168	0,0988	0,1723	0,0741
Observed: stool piles	0,1771	0,0976	20	180	0,0985	0,1718	0,0739
Observed: Sewage water	0,1580	0,0776	20	180	0,0783	0,1366	0,0587
Disposes infant stool adequately	0,0360	0,0397	4	176	0,0749	0,1306	0,0561
Household practices ODF but toilet available	0,1031	0,0859	4	164	0,1358	0,2368	0,1018
Reprehends others for ODF	0,0837	0,0862	20	180	0,0877	0,1529	0,0657
Believes ODF is health hazard	0,0400	0,0427	19	180	0,0444	0,0775	0,0333
Toilet/covering excreta important	0,0973	0,0699	19	180	0,0716	0,1248	0,0537
Uses soap after toilet	0,0605	0,0615	20	180	0,0628	0,1095	0,0471
Uses soap before meal	0,1044	0,1042	20	180	0,1057	0,1843	0,0793
Health knowledge							
<i>Index of health knowledge</i>	0,0640	0,0671	15	180	0,0731	0,1275	0,0548
Heard of message: malaria/dengue	0,0746	0,0950	17	180	0,0999	0,1743	0,0749
Heard of message: diarrhea	0,0722	0,0934	17	180	0,0983	0,1715	0,0737
Heard of message: ARI	0,0664	0,0892	16	180	0,0955	0,1666	0,0716
Heard of message: STI	0,0817	0,0699	15	180	0,0753	0,1313	0,0564
Knows danger signs of malaria	0,0727	0,0922	18	180	0,0959	0,1673	0,0719
Knows danger signs of ARI	0,0639	0,0776	18	180	0,0810	0,1412	0,0607
Knows treatment of diarrhea	0,0638	0,0790	18	180	0,0824	0,1438	0,0618
Women's and child's health, self-reported							
Self-assessed health is good	0,0765	0,0955	17	180	0,1004	0,1751	0,0753
Feels chronically tired	0,0747	0,0890	17	180	0,0936	0,1633	0,0702
Child had no diarrhea in past 3m (any child)	0,0382	0,0746	16	180	0,0817	0,1425	0,0613
Child had no ARI in past 3m (any child)	0,0610	0,0732	16	180	0,0787	0,1373	0,0590
Anthropometric measurement							
<i>Index of child health</i>	0,2642	0,0501	5	152	0,0641	0,1118	0,0481
<i>CIAF</i>	0,0355	0,0745	8	166	0,1088	0,1898	0,0816
Weight, female 20-49	0,0181	1,1690	16	171	1,3499	2,3544	1,0124
Hemoglobin, female 20-49	0,0380	0,2290	15	170	0,2634	0,4594	0,1975
Weight, male 20-49	0,0353	1,5627	7	162	2,4357	4,2483	1,8268
Hemoglobin, male, 20-49	0,0599	0,2866	6	162	0,4389	0,7655	0,3292

Weight, U5	0,0220	0,4586	10	167	0,6393	1,1150	0,4794
Weight-for-age z-score, U5	0,0609	0,3254	10	167	0,4130	0,7204	0,3098
Underweight, U5	0,0410	0,0785	10	167	0,1037	0,1809	0,0778
Wasting, U5	0,0415	0,0627	8	167	0,0896	0,1563	0,0672
Height, U5	0,0176	1,7848	10	167	2,5272	4,4079	1,8954
Height-for-age z-score, U5	0,0187	0,3017	10	167	0,4256	0,7423	0,3192
Stunting, U5	0,0208	0,0709	10	167	0,0992	0,1730	0,0744
Hemoglobin, U5	0,0000	0,1505	7	160	0,2830	0,4936	0,2122
Oedema, U5	0,1569	0,0568	11	168	0,0643	0,1122	0,0482
Arm circumference, U5	0,0487	0,2531	8	165	0,3569	0,6226	0,2677
Arm circumference-for-age z score, U5	0,0833	0,2365	8	165	0,3108	0,5422	0,2331
Arm circumference < 115 mm, U5	0,0145	0,0221	8	165	0,0352	0,0613	0,0264

The table presents (1) intra-cluster correlations (ICC) calculated from baseline data, (2) minimum-detectable effect (MDE) sizes calculated from baseline data, (3) the number of observations per cluster at endline as used in the analysis, (4) the number of clusters at endline as used in the analysis, (5) MDE sizes given endline data (based on (3) and (4)), (6) MDE sizes given endline data and assuming compliance of 43 percent, and (7) MDE sizes given endline data without correcting for partial compliance, for each outcome.