Impact of the Gram Varta programme on health, nutrition and women's empowerment in India

Malavika Subramanyam, Indian Institute of Technology (IIT) Gandhinagar

Cara Ebert, University of Göttingen

Christian Bommer, University of Göttingen

Lisa Bogler, University of Göttingen

Abhijeet Kumar, University of Göttingen

Sini Varghese, IIT

Sagar Atre, IIT

Sebastian Vollmer, University of Göttingen

Grantee Final Report

Accepted by 3ie: November 2017

Note to readers

This impact evaluation has been submitted in partial fulfilment of the requirements of grant CPW.01.GV.IE issued under Bihar Policy Window. This version is being published online as it was received. A copy-edited and formatted version will be available in the 3ie Impact Evaluation Report Series in the near future. All content is the sole responsibility of the authors and does not represent the opinions of 3ie, its donors or its board of commissioners. Any errors and omissions are the sole responsibility of the authors. All affiliations of the authors listed in the title page are those that were in effect at the time the report was accepted. Any comments or queries should be directed to the corresponding author, Sebastian Vollmer at svollmer@uni-goettingen.de

3ie has received funding for the Bihar Policy Window from UKaid through the Department for International Development.

Suggested citation: Subramanyam, M, L, Ebert, Bommer, C, Bogler, C, Kumar, A, Varghese, S, Atre, S and Vollmer, S, 2017. *Impact of the Gram Varta programme on health, nutrition and women's empowerment in India, 3ie Grantee Final Report.* New Delhi: International Initiative for Impact Evaluation (3ie)

Contents

Acknowledgments

We are grateful to the residents of the villages of Madhepura who participated in this research. The project benefited greatly from the remarkable contribution of the following students of the University of Göttingen: Miriam Albe, Mark Bohnes, Daniel Geissel, Esther Heesemann, Tatjana Kulp, Jennifer Pengel, Mareike Schön, Boryana Spasova, Michaela Theilmann, Kay Tuschen, Paula van Haaren, and Ann-Charline Weber. We gratefully acknowledge the timely efforts of the following students of IIT Gandhinagar and their colleagues: Prathu Bajpai, Susanna G., Jerene George, Vikash Patel, Vaidyanathan P.R., Debapriya Ray, Promita Roy, Ambarish Singh, and Ankita Shah.

We could not have completed this project without the able leadership of the Madhepura-based field team led by Abhijeet Kumar and supported by Pushkar Anand, Late Abhimanyu Das, Pratima Goswami, Budhsen Kumar, Rahul Kumar, Shivshankar Kumar, and Priyanka Kumari, many thanks to all of them!

We are grateful for funding from the International Initiative for Impact Evaluation (3ie) and the support from 3ie staff, particularly Francis Rathinam.

Executive Summary

The problem: HNWASH in Bihar

Women's and children's health (WCH) and water, sanitation and hygiene remain among the key health topics that attract and require attention in India. Reducing infant and neonatal mortality, maternal morbidities, and child mortality have remained major priorities of the Government of India and the governments of Indian states such as Bihar. Bihar continues to have conditions leading to high infant and neonatal mortality rates. In Madhepura district of Bihar, the focus of this report, the infant mortality rate was 64 deaths per 1,000 live births in 2013 (Government of India, 2013). Understanding of prevention and treatment of diseases and child care practices are known to be poor in this region.

The intervention: Gram Varta

Gram Varta is a programme that aims to improve critical indicators of health, nutrition, water, sanitation, and hygiene (HNWASH) by changing the relevant attitudes and behaviours in communities. It is a community-based, participatory programme, implemented in Bihar by state-supported agencies. It uses the participatory learning and action (PLA) approach and works by using women's self-help groups (SHGs) as agents of change. The implementing agencies train local women as facilitators to conduct structured meetings of the SHGs and guide participants through a cycle of participatory action. The meeting cycle addresses the identification and prioritization of problems related to HNWASH, the development and implementation of strategies to address these problems, and the evaluation of these strategies.

The underlying theory of change posits that the PLA approach fosters critical thinking skills in members, and by involving the entire community in the process, instills solidarity among community members. Gram Varta aims to change attitudes and behaviours, and enable communities to demand and utilize health services, while empowering communities to hold service providers accountable, and thus eventually improving health. The intended beneficiaries are women and children, with a special focus on pregnant women, lactating women, children under age five, and adolescent girls. However, the entire community, including men and service providers, are involved in the process. The idea is to empower communities to improve their health.

The evaluation: A randomized controlled trial

The core approach taken by Gram Varta has been tested in other parts of the world and in India. However, all previous trials, except for one in Bangladesh, were implemented on a small scale. Moreover, Gram Varta differed in several other aspects from the only other large scale trial. For this purpose, Gram Varta implementation in Madhepura district was embedded in a randomized controlled trial with subsequent phase-in.

The impact evaluation of Gram Varta tries to address the following gaps in current knowledge: Firstly, can Gram Varta be brought to scale in the Indian setting and what are the practical challenges arising from the large scale? Secondly, does the approach remain effective when its focus is on a broader range of HNWASH topics instead of the previously successful focus on maternal and child care only? Thirdly, is it feasible to use existing women's SHGs as a platfoprm for implementation instead of establishing new groups for the specific purpose of

delivering Gram Varta PLA meetings? Lastly, can this approach be effective without simultaneous strengthening of health service providers?

We randomly assigned Gram Varta implementation among 68 gram panchayats of Madhepura district. This randomization resulted in 90 villages assigned to the treatment group villages and 90 control villages. Pre- and post-implementation, we surveyed household heads, women of reproductive age, adolescent girls, and women pregnant at the time of pre-implementation. We also surveyed Gram Varta facilitators and Anganwadi workers in the treatment areas. A post-intervention comparison of the treatment and control group accounting for baseline characteristics allows us to measure the causal impact of Gram Varta. We have distinguished between the intention-to-treat effect (the impact on all households irrespective of their participation in Gram Varta) and the treatment effect on those who regularly participated in SHGs. Further, we have substantiated our quantitative findings with results from qualitative interviews with participants and facilitators.

The sample for the quantitative study included:

- 3953 household heads and women in reproductive age (household survey);
- 2000 pregnant women and their husbands (pregnant women survey);
- 316 community mobilizers (facilitator survey);
- and 265 anganwadi workers (Anganwadi survey).

This evaluation investigated seven groups of hypotheses regarding:

1. Participation in, acceptance and awareness of, **women's self-help groups** (SHG) as well as utilization of government health services through SHGs.
2. **Women's agency and empowerment** in terms of economic independence, bargaining power, recognition and confidence in the community, domestic violence, family planning and nutrition decisions of the women and adolescent girls.
3. **HNWASH knowledge and practices** in terms of own and child nutrition, awareness and prevention of diseases, risky consumption behaviour, domestic hygiene and sanitation as well as adolescent girls' and women's knowledge of sexuality and contraception.
4. Behaviour during **pregnancy** in terms of health, nutrition, antenatal care visits, and optimistic outlook.
5. **Anganwadi centers** in terms of utilization, undernutrition treatment and prevention, quality of health services, facilitation of routine check-ups, and Anganwadi worker related outcomes.
6. **Health outcomes** of women, their husbands and children.
7. **Social cohesion** in the community and neighbourhood.

Results

While we did not find consistent evidence for improvements in health indicators, we found some evidence that Gram Varta (i) increased women's involvement in the community; (ii) made them self-confident when it comes to refusing sexual intercourse with their husbands or demanding that they use a condom; (iii) reduced the practice of domestic violence and oppression; (iv) decreased women's preference for sons; (v) encouraged pregnant women to be mindful of their health, to take healthy and sufficient diet, to avoid stress and to avoid health risks; (vi) encouraged women to accept their pregnancy, making them more optimistic about their situation; and (vii) increased mutual trust within the community.

Discussion

The key learnings that emerge from our mixed methods study are that a PLA approach, implemented through existing SHGs supported by the state, is able to change women's social capital, self-confidence; and reduce the likelihood of them experiencing controlling behavior by husbands or even domestic violence. Such an approach could lead to an increase in participants' sense of trust in their communities. All of these effects are likely to be very beneficial in low-resource settings with historically high levels of patriarchy. However, whether such changes will lead to improved health knowledge, practices, and outcomes can only be known by investigating the long-term impact of Gram Varta participation.

We have confidence in these findings because data suggest minimal concern with respect to internal validity: there is a good balance of confounders post-randomization; spill-over and contamination though John Henry and Hawthorne effects are highly unlikely; the use of a few control questions and hard data to verify answers reduce social desirability bias; and there is no threat from attrition.

Moreover, our qualitative findings align well with the quantitative findings and give a context-rich insight in helping us understand a majority of the quantitative results.

Our findings are in contrast to those from previous trials examining the same PLA approach in women's groups. One explanation for the lack of consistent effects on HNWASH and other indicators could be that Gram Varta was not successful in bringing health providers and husbands along, who are the ones that make relevant decisions for the entire household.

List of figures and tables

Abbreviations and acronyms

AME	Average marginal effects
ANM	Auxiliary nurse midwife
ASHA	Accredited social health activist
AWC	Anganwadi center
AWW	Anganwadi worker
BPL	Below poverty line
BTAST	Bihar Technical Assistance and Support Team
CLF	Cluster level federation
CM	Community Mobilizer
DHS	Demographic Health Survey
DiD	Difference-in-differences
FGD	Focus group discussion
HNWASH	Health, nutrition, water, sanitation, hygiene
ICC	Intra class correlation
ICDS	Integrated child development services
ITT	Intention-to-treat effect
Jeevika	Bihar Rural Livelihoods Promotion Society
NFHS	National Family Health Survey
NMR	Neonatal mortality rate
PLA	Participatory learning and action
RCT	Randomised controlled trial
SHG	Self-help group
SMS	Short Message Service
STD	Sexually Transmitted Disease
ToC	Theory of change
VO	Village Organization
WCH	Women's and children's health
WDC	Women's development corporation

1. Introduction

Women's and children's health (WCH), water, sanitation and hygiene are among the key health topics that attract and require attention in India. Reducing infant and neonatal mortality, maternal morbidities, and child mortality remains a major priority of the Government of India and the governments of Indian states such as Bihar, Jharkhand, Assam, Rajasthan, Uttar Pradesh, Uttarakhand, Odisha, Chhattisgarh, and Madhya Pradesh. The Government of India conducts annual health surveys in these regions to monitor critical health indicators (Government of India, 2013). The state of Bihar has historically reported among the highest mortality and morbidity figures. It continues to report a high infant mortality rate of 48 deaths per 1,000 live births in 2013 (Government of India, 2013), and in 2015 (IIPS, 2016). In Madhepura district, the site of this evaluation, the infant mortality rate was even higher at 64 deaths at 1,000 live births according to census figures (Government of India, 2013) (2015 figures were not available for Madhepura). The focus on maternal and child health that multiple international agencies such as UNICEF and the Bill and Melinda Gates Foundation adopt in their activities in India demonstrates the importance of the topic in the eyes of policy-makers and academics alike. A discussion of health, nutrition, water, sanitation, and hygiene (HNWASH) can be approached from both the supply side and the demand side.

Supply side: Improving access to health care and sanitation

On the supply side, there have been many attempts to provide services in all areas of HNWASH. The Integrated Child Development Services (ICDS) scheme, Accredited Social Health Activists (ASHAs), and the Total Sanitation Campaign—these are just a few examples of the various government programmes that address maternal health, child health, and sanitation (Government of India, 2015). Village Health and Nutrition Days are an outreach activity at Anganwadi Centres—which implement the ICDS—that provides maternal and child care including nutrition. Approximately 890,000 ASHAs have been trained and are engaged across India to make health care services accessible for the poorest, most marginalized and vulnerable populations. Maternal and Child Health Wings are established at selected health facilities to improve the quality of care provided to mothers and children. Under the Total Sanitation Campaign, sanitation facilities are constructed in communities, in schools, and Anganwadi Centres, and the campaign includes financial incentives for latrine construction.

However, many top down schemes focusing on service supply have not proven impactful. Firstly, a lack of adequate and regular services and the resulting lack of trust reduce the effectiveness of such services. Secondly, in many cases, even existing services or facilities are not used for reasons of tradition, habit, culture, or a lack of understanding. For example, despite increased latrine coverage thanks to the Total Sanitation Campaign, open defecation remains widespread, simply because people prefer defecation in the open (Barnard et al. 2013). Women do not send their sick children to health facilities for care because they do not recognise symptoms of childhood diseases (Shah et al. 2011). Hands are not washed at critical times partly because it is not a norm to do so and the habit is not instilled at an early age (Curtis, Danqua, and Aunger, 2009). Whether a woman goes to a health facility for antenatal care or delivery is strongly influenced by beliefs and traditions (Tamang et al. 2001).

Therefore, increasing demand for services based on behaviour change is crucial to improve HNWASH outcomes. Moreover, programmes that allow for the participation of communities

show success in changing attitudes and behaviour related to HNWASH topics (Kar and Pasteur, 2005). Recognising these two points, the Gram Varta programme uses a participatory learning and action (PLA) approach to address the demand side of HNWASH services. This approach involves beneficiaries as agents of change in a cycle of problem identification and problem solving.

Demand side: Mobilizing the community with PLA

The Gram Varta programme is based on a model of PLA that gained broad recognition after its implementation in rural Bolivia. Prost and co-authors provide a meta-analysis of all randomised controlled trials (RCTs) following this approach (2013). In the Bolivian Warmi project, women's groups were organised and guided through a participatory cycle of action with the goal of improving maternal and child health (Howard-Grabman et al. 2002). The action cycle is directed by a facilitator and consists of skill development for the identification and prioritisation of problems, and the development and implementation of formal action plans. Auxiliary nurses from Save the Children (the facilitators) visited women's organisations at least monthly to conduct meetings. A before-after evaluation of the project found large reductions in perinatal mortality (by 65 percent), a doubling of participation in women's organizations, and improvements in pre- and postnatal practices (O'Rourke, Howard-Grabman, and Seoane, 1998).

The first randomised controlled trial of a programme modelled on the Warmi project was conducted in Makwanpur, rural Nepal, from 2001 to 2003 (Manandhar et al. 2004). Female facilitators, in this case literate women from the local area, were trained in participatory communication techniques and essential perinatal health issues. The facilitators organised women's group meetings every month and guided participants through the action-learning cycle in the course of ten meetings. The community-based participatory intervention reduced neonatal mortality by 30 percent. Moreover, maternal mortality decreased and home-care practices and health-care seeking for neonatal and maternal morbidity improved in the intervention area (Manandhar et al. 2004).

Successful interventions following the same PLA model with women's groups were also implemented in Malawi (Lewycka et al. 2013; Colbourn et al. 2013). In these cases, the PLA model was compared with a home visit strategy and a facility-based intervention, respectively. Although interactions were observed between the groups assigned to these strategies, reductions in mortality rates and improvements in neonatal care practices were observed in both trials.

The PLA approach was implemented and tested on a much larger scale in Bangladesh than in Nepal. Results from the randomised controlled trial revealed no reduction in neonatal or maternal mortality rates, contrasting previous findings (Azad et al. 2010). In an attempt to investigate whether the low population coverage of women's groups reduced its success, the intervention was repeated between 2009 and 2011 in the same three districts of Bangladesh with a higher population coverage of women's groups (Fottrell et al. 2013). Again, women's groups proceeded through the PLA cycle, identified problems related to maternal and neonatal health, and developed strategies to address these. Groups were newly formed in addition to those already set up in the intervention area as part of the earlier trial, resulting in a 4- to 5-fold increase in population coverage of women's groups. This intervention led to a reduction

in neonatal mortality of up to 39 percent, similar to the trial in Nepal. Improvements were also observed in essential newborn care and feeding practices (Fottrell et al. 2013).

In India, the approach was first introduced in Jharkhand and Odisha, rural states in eastern India, from 2005 to 2008 (Tripathy et al. 2010). As the intervention was implemented by the voluntary organisation Ekjut, it was called the Ekjut trial. Facilitators in this project were again local women who had received training on participatory communication and health issues. However, they each facilitated more groups than in the Makwanpur trial and conducted 20 meetings with each group. Results from the randomised controlled trial revealed a reduced neonatal mortality rate by 32 percent during the 3 years. Maternal depression, another main outcome analysed, was not significantly reduced overall, except for moderate depression. Use of health services did not increase, but hygiene and care practices around delivery improved (Tripathy et al. 2010).

After learning about the positive results of the Ekjut trial, the Bihar Government's Bihar Technical Assistance and Support Team (BTAST) implemented a trial of the programme, then called Gram Varta, in one block called Maner (near Patna). The endline evaluation found some evidence of impact, such as a rise in the proportion of women consuming iron tablets and the number of children attending Anganwadi centers (Vijayalakshmi et al. 2016). Based on these findings, the Gram Varta programme was scaled up to a larger number of districts and run simultaneously by the Women's Development Corporation (WDC) and the Bihar Rural Livelihoods Promotion Society (Jeevika). Implementation areas of Jeevika included the five districts Purnea, Madhubani, Madhepura, Supaul, and Gaya while the WDC implemented it in a number of other districts including Gaya, Jehanabad, and Rohtas.

Research questions: The remaining knowledge gap

As discussed above, earlier trials of the PLA approach in WCH in India were performed in smaller, more controlled environments; and in the case of Jharkhand and Odisha the programme was implemented by a voluntary organization in liaison with the government. However, to address the question of whether this programme could be taken to scale and delivered through existing government structures while remaining effective, a rigorous pre-post impact evaluation of the WDC and Jeevika implementation was necessary.

Gram Varta programme is similar in scale to the trial in Bangladesh, but different in other aspects; for example, Gram Varta uses existing SHGs instead of newly created ones. The impact evaluation of Gram Varta therefore tries to address the following gaps in current knowledge: Firstly, can Gram Varta be brought to scale in the Indian setting? What might be the practical challenges arising from the large scale? Secondly, does the approach remain effective when its focus is on a broader range of HNWASH topics instead of maternal and child care only? Thirdly, is it feasible to use existing women's SHGs as a platform for implementation instead of establishing new groups for the specific purpose of delivering Gram Varta PLA meetings? Lastly, is this approach effective without simultaneous strengthening of health service providers?

The Gram Varta evaluation

In order to evaluate the large scale Gram Varta implementation, the programme roll-out in Madhepura district of Bihar was embedded in a randomized controlled trial with a subsequent

phase-in. Jeevika was the implementing agency in Madhepura. We randomly assigned Gram Varta implementation to gram panchayats with a probability of 50 percent. Pre- and post-implementation, we surveyed household heads, women in reproductive age, adolescent girls, and women pregnant at the time of pre-implementation. In the treatment areas we also surveyed Gram Varta facilitators (so called community mobilizers (CM)) and Anganwadi workers (AWW). A post-intervention comparison of the treatment and control group accounting for baseline characteristics allows us to measure the causal impact of Gram Varta. We have distinguished between the intention-to-treat effect, which is the impact on all households irrespective of their participation in Gram Varta, and the treatment effect on those who regularly participate in SHGs. The latter was achieved in a subgroup analysis. The major part of the analysis is based on these quantitative surveys. This is substantiated with a qualitative study that helps us understand the quantitative results and places them in context of Madhepura.

The impact of Gram Varta was determined according to its impact on indicators related to maternal and child health; sanitation; hygiene; and proper nutrition of vulnerable sections of the population such as women, adolescent girls and children, especially those under five years of age. The Gram Varta process aims to improve women's knowledge and decision-making power in the community; enable community members to demand and utilize government services (such as the Anganwadi centers, the primary health centers and sub-centers, and the sanitation related services of the government); and inform and involve the community in the goal of improving HNWASH indicators.

The evaluation of Gram Varta builds on a previously set-up pre-analysis plan determining the hypotheses to be tested.[1] While we ensured adherence to the plan as closely as possible, during the course of the surveys it became necessary to make decisions in favor of survey quality over keeping every variable listed in the pre-analysis plan. Despite such decisions only two of the hypotheses announced in the pre-analysis plan were not investigated: hypothesis 15 consisting of a single indicator (comfortable with speaking out in public) and hypothesis 41 (awareness and understanding of abortions among pregnant women). Please see tables L.1 and L.2 in the appendix for a list of variables excluded from the report and the reasons for exclusion. In total we have analyzed seven groups of hypotheses regarding:

1. Participation in, acceptance, and awareness of **women's self-help groups** (SHG), as well as utilization of government health services through SHGs.
2. **Women's agency and empowerment** in terms of economic independence, bargaining power, recognition and confidence in the community, domestic violence, family planning, and nutrition decisions of the women and adolescent girls.
3. **HNWASH knowledge and practices** in terms of own and child nutrition, awareness and prevention of diseases, risky consumption behaviour, domestic hygiene and sanitation as well as adolescent girls' and women's knowledge on sexuality and contraception.
4. Behaviour during **pregnancy** in terms of health, nutrition, antenatal care visits, and optimistic outlook.
5. **Anganwadi centers** in terms of utilization, malnutrition treatment and prevention, quality of health services, facilitation of routine check-ups, and Anganwadi worker related outcomes.
6. **Health outcomes** of women, their husbands, and their children.

[1] The pre-analysis plan refers to an internal document cited under Bommer, Subramanyam, and Vollmer, 2015a.

7. **Social cohesion** in the community and neighbourhood.

The report is structured as follows. Section 2 describes the intervention and theory of change and lists the research hypotheses as stated in the pre-analysis plan. Section 3 introduces the setting of the implementation, followed by section 4, which presents the timeline of Gram Varta implementation and evaluation. In sections 5 and 6 we describe the design, methods, and implementation of the evaluation and the Gram Varta programme. Section 7 contains the impact analyses and results of our indicators, grouped by hypothesis. In section 8 we discuss results and address concerns related to them. Section 9 closes with remarks and recommendations for policy makers contemplating a PLA-based programme.

2. Intervention, theory of change and research hypotheses

2.1 Description of intervention

Gram Varta was piloted in 2011 with the goal of using participatory learning and action (PLA) to impact health, nutrition, water, sanitation, and hygiene (HNWASH) in Bihar. This programme takes a PLA approach to empower and mobilize communities. It is implemented through village-based women self-help groups (SHG) affiliated with Women's Development Corporation, Jeevika and Mahila Samakhya. It has already been implemented in several districts of Bihar. The scheduled implementation in Madhepura was accompanied by a randomized study to evaluate its impact.

The core of the intervention is a cycle of 20 pre-structured meetings. A few SHG members are selected and trained in facilitating these PLA meetings (in Madhepura the facilitators were existing staff of Jeevika called "Community Mobilizers"). Each facilitator invites the other SHG members and the entire village population to participate in these meetings. The facilitator uses games, stories, and activities related to women's agency, attitude towards working together, service utilization, as well as HNWASH knowledge and practices. Participants are encouraged to think critically, identify problems in their households and communities, and discuss how HNWASH practices could be improved. Pregnant women, lactating women, adolescent girls, and children under age five are of special interest. This process includes at least two meetings with the entire community including local authorities and service providers. The identified problems and solutions are discussed and a community action plan is formulated. Monitoring progress towards meeting pre-set goals is also a part of Gram Varta.

The ultimate aim of Gram Varta is to improve maternal and child nutrition status and health via changes in the community's ability to identify and address HNWASH issues that they themselves prioritize. The intended beneficiaries are therefore women and children, with a special focus on pregnant women, lactating women, children under age five, and adolescent girls.

One of the underlying assumptions is that Bihar's high levels of undernutrition and poor health among pregnant women, adolescent girls, and children under age five, are a result of risky HNWASH practices, including inadequate uptake of existing programmes targeting HNWASH.

The key outcomes of interest to the implementing agency were identified during our discussions with them, namely: mediators of change such as the sense of agency, empowerment, and critical-thinking skills; mediators that connect Gram Varta with demand for services; changes in attitudes of service providers; and changes in practices related to HNWASH.

2.2 Theory behind PLA

PLA is a development strategy that understands participation of the target group in the development process both as a means and an end (Wetmore and Theron, 1998). The PLA approach solves two issues. Firstly, it puts community priorities first and uses local resources for problem solving. The idea is that local actors might perceive different problems to be the most pressing compared to external actors. They are also often able to come up with solutions more suitable to their environment than outside experts (Kar, 2003). Secondly, participation is seen as empowering, because it creates a feeling of ownership of activities and processes, and therefore has the potential to lead to sustainable change once project funds are depleted. The sense of agency and skills instilled by PLA could be put to use in other domains outside the specific initial programme.

The concept of PLA is a cycle of sharing among local actors, analysing problems, finding solutions, implementing actions, evaluating these, and analysing again (Wetmore and Theron, 1998). It assumes that once people are aware of their situation and identify causes of problems, they have the motivation to act. The theoretical foundation of PLA builds on Paulo Freire's concept of critical consciousness. The Brazilian educator advocated participatory education, in which people are actors instead of objects and identify problems and solutions themselves (Wallerstein and 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, page 4). The concept includes critical reflection, the perceived ability to initiate change, and critical action (Watts, Diemer, and Voight, 2011). Although the concept of critical consciousness was developed for education, it has also been applied to the health context (Minkler and Cox, 1980; Wallerstein and Bernstein, 1988).

2.3 Theory of change (ToC)

Below we describe the causal pathway and the necessary assumptions linking each stage of the intervention to the anticipated final outcomes. The implementation stages have been separated out for the convenience of understanding the process. Many of these stages repeat over the cycle of 20 meetings including the two community meetings and the underlying theoretical concepts. By extension, the assumptions are applicable to each of the meetings. Stages 2 through 5 are presented separately only to highlight different assumptions. Theoretically they all go together and are the core of the PLA approach.

Stage 1 – Trained facilitator conducts social mapping

Process: As a first step, the implementing agencies select facilitators who are local to the area and who are from well-established women's self-help groups. The facilitators receive comprehensive training and necessary equipment in order to enable them to independently

implement the pre-structured Gram Varta meeting schedule in their communities. The trained facilitator conducts social mapping of the village where she lists all houses and identifies houses with SHG members, pregnant women, children under five and other target groups. She maps important sites such as the Anganwadi centre, water pumps, and potential sites for Gram Varta meetings.

Theory: The process of social mapping not only provides a tool for planning and activities during the later stages, but also allows the facilitator to interact with the community and spread awareness about Gram Varta, thus sparking curiosity in the community with the aim of seeing decent participation at the meetings.

Assumptions: For the intervention to be successful it is crucial to select a reliable facilitator who is highly motivated and accepted within the community. Further, reading and writing skills are a necessity for the facilitator to be successful in learning about and carrying out her tasks. Other assumptions are that the social mapping is inclusive of all subgroups and sub-areas of the village and that the facilitator possesses good skills to interact with community members and is able to carry out this task as planned.

Stage 2 - Trained facilitator sets up meetings

Process: Once the facilitator completes the preparations for Gram Varta she schedules 20 meetings, one at a time, where meeting sites and dates are set in consultation with SHG members. As the idea of Gram Varta is also to include the entire village population in these meetings, the meeting site is chosen by the SHG group such that it is comfortable, well-known and visible. Target groups such as pregnant women, adolescent girls and also men are informed about the PLA meetings. The Jeevika model is to focus only on SHG members when scheduling regular SHG meetings, but facilitators are expected to encourage other community members to participate in Gram Varta meetings. Moreover, front line workers such as Anganwadi workers (AWW) and Auxiliary Nurse Midwives (ANMs) are encouraged to participate in meetings.

Theory: The act of participation, from the very beginning, including being able to decide date and site of meeting, fosters a sense of empowerment. By holding meetings in the open and publicizing it through networks the community is informed and curious. This leads to increased participation and through participation to increased cohesion. The community-level meetings are inclusive, interactive, and strengthen the community's focus on priorities identified and action plans to tackle these priority areas. Front-line workers participate in PLA (regular Gram Varta meetings) as well as the community meetings; contribute their perspective; learn about the members' and community's views; and work with them to improve service delivery and uptake.

Assumptions: SHG members and community members learn about the meetings. SHG members and community members trust the facilitator. Members of SHGs and the community attend the meetings. Front-line workers learn about and attend the meetings. All target groups learn about and are willing and able to attend the meetings. This is assuming, for instance, that husbands and other family members are not barriers for women who are willing to attend meetings; that travel costs, weather conditions and work or child care obligations are not obstacles; that all target groups are interested to attend; and that they are open and willing to learn new information.

Stage 3 - Participation in meetings

Process: Once the SHG members, community members, and front-line workers attend the meetings, they participate actively. All of them engage deeply with and participate meaningfully in these meetings. Stories, games, and role-plays engage members in learning processes. Easily available materials like local household articles are used in activities. Members identify and prioritize the health problem in their village. They identify strategies to address problems and plan monitoring of action taken.

Theory: Over the course of the meeting cycle a sense of empowerment and agency develops in the women. They learn to speak up in front of a group of people and in general gain self-confidence as individuals and as a group. Their voices are heard and valued by the community. The community meetings and the dissemination of Gram Varta messages through person-to-person channels strengthen the feeling of cohesion and solidarity within the community. Front line workers such as AWWs and ANMs encourage participants to demand and utilize services.

Assumptions: Over the course of the meeting cycle a sense of empowerment and agency develops in the women. Members actively participate and are not passive attendees. Women start thinking critically and identifying problems in their community and homes. Sense of empowerment and agency strengthen over the meeting cycle. Women's critical thinking skills and problem solving skills strengthen over the meeting cycle. The community comes together to work on priorities identified. Front-line workers accept and welcome the changes seen in the SHG members and the community due to Gram Varta. Front-line workers value the contribution that Gram Varta is making.

Stage 4 - Participatory learning

Process: During the meetings the facilitator uses pedagogical games, stories, and other participatory learning and action techniques she learned about in her training. Simple material from their own homes such as vegetables and wheat flour are used in activities. Important information on HNWASH topics is discussed as part of PLA. Story cards, mapping, simple activities such as one on handwashing, and many more, are used to highlight different HNWASH topics. Attendees participate actively, share and discuss, frequently identify themselves with the characters in stories, gain knowledge, change attitudes and become armed with ways to engage in discussions. Two community meetings are initiated by the facilitator with the purpose of communicating identified problems to local authorities and suggest and demand solutions. Conversations on these issues occur on all levels and between all stakeholders (within the group, within the households, within the community, between community, front line workers and local authorities, etc.).

Theory: Learning occurs through the process of participation. No direct messages are given. However, the PLA sessions enable attendees to come to realizations about their own attitudes and those of their families and community; inspire them to pay attention; to learn about HNWASH; and to speak up about it at home and in the community. Knowledge is gained. The programme generates the ability to identify problems, prioritize problems, find solutions, and take action on issues including HNWASH. It induces dialogue between women, community and front line workers on important topics including HNWASH. The process increases community solidarity.

Assumptions: This is a crucial stage where many channels that are hypothesized to impact HNWASH practices are assumed to be affected at once. In particular, the assumptions are that:

1. Awareness is raised on the importance of relevant practices for health and nutrition of the whole family, e.g. open defecation vs. toilet use, hand washing, breast feeding, food intake during pregnancy, demand for health services.
2. Important knowledge is transferred, e.g. nutritional information on food (which food is important for energy, growth, a strong immune system, etc.), symptoms of important illnesses, information on anthropometric indicators of women's children, information on what increases safety of home delivery, information on family planning.
3. Personal attitudes towards, and beliefs about social norms regarding, the desirable practices change: (a.) The activities trigger women's critical thinking and problem solving skills. They individually are willing to commit to the desired behaviours. (b.) If a majority of participants speak up in the group and indicate their commitment to certain practices, all participants are assumed to have committed to it and to lead to a change of social norms within the community regarding the behaviour. This is assumed to help them once they are back in their household and have to discuss the change with their relatives, as they can refer to other village members who practice the desirable behaviours already and claim that norms are changing.
4. Participants are determined to find solutions to self-identified problems and are motivated and enabled to develop a community action plan. Women feel empowered and strong enough to foster change in practices and behaviour in their own households.
5. The community comes together in solidarity. The community supports all members in the health-promoting decisions they take. Participants speak up to demand improvements in service provision in the community. A crucial assumption here is that all these dialogues take place and that in the process the community is truly mobilized to jointly take action and solve the identified problems. Beliefs on obstacles and constraints preventing the desired behaviours are removed.

A prerequisite for all of the listed assumptions to hold are that Gram Varta activities are effective and powerful enough to trigger the listed changes, that the facilitator understands and remembers the information she is supposed to provide, and that the pre-designed activities for each meeting are carried out correctly.

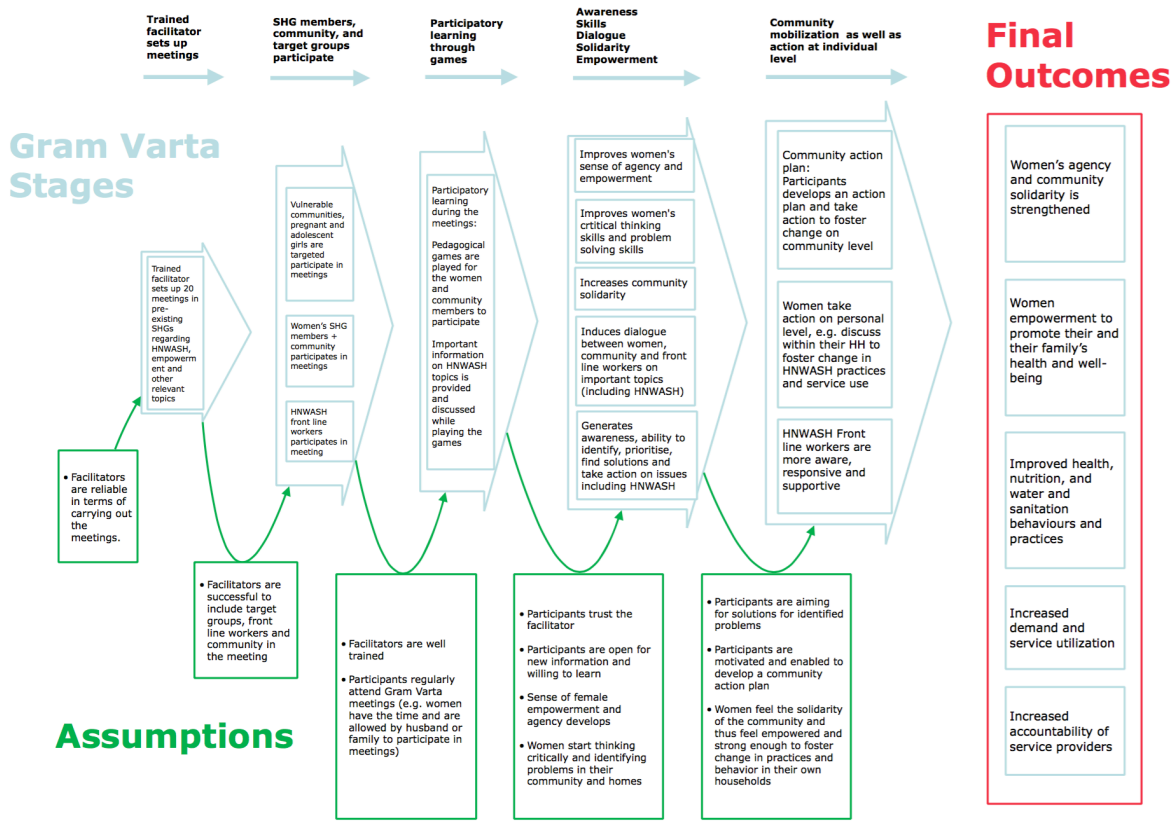
Stage 5 - Community mobilization, action at individual and community level and evaluation

Process: Health priorities and strategies to address them are identified and action plans are drawn up. Plans for monitoring action are created and followed up. A community action plan is developed and action is taken to foster change on the community level. Members monitor progress made towards achieving goals. Women take action on a personal level, e.g. discuss within their household about fostering change in HNWASH practices and service use. Front line workers are more aware, responsive, and supportive.

Theory: The process of PLA has led members and the community to learn more about their own attitudes towards HNWASH practices, introspect about them, be open to new information, gain knowledge, feel empowered to make a change, convince others to make a change, initiate and maintain dialogue in the community, and come together to take concrete action on priorities that they have themselves identified. The process also inspires them to also monitor progress towards meeting the stated goals of the community action plan.

Assumptions: Changes in norms are actually occurring and obstacles to the desired behaviours are actually removed. Women feel empowered to carry out the learned behaviours together with or even against the will of their husbands or other relatives. Identified problems on the community level are actively solved. HNWASH front line workers are more responsive and supportive. The evaluation mechanism is effective and is actually implemented.

Figure 2.1: Theory of Change



Stage 6 – Indicators of proximal determinants of HNWASH outcomes change and ultimate outcome indicators improve

Theory: Due to improved HNWASH practices; increased service demand and uptake; and solved problems at the household and village level; women and children are healthier and less malnourished.

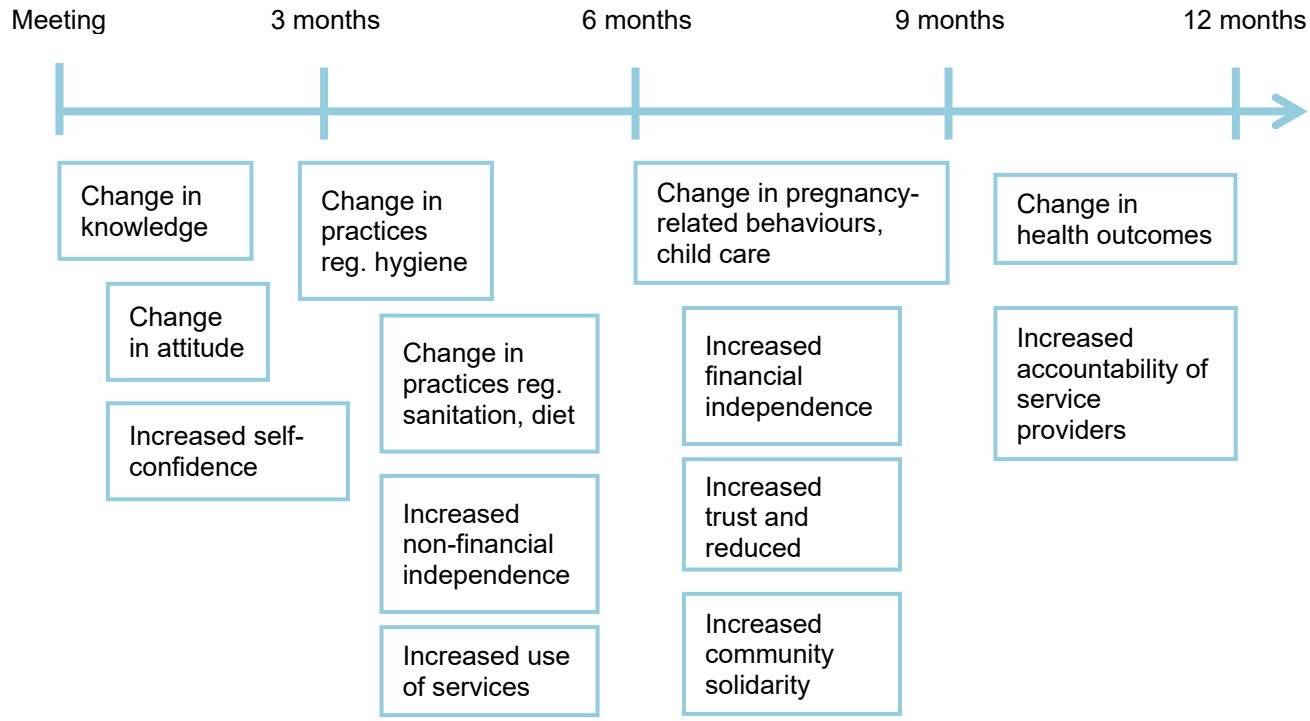
The above stages are summarized in Figure 2.1. Note that this figure of the ToC has not distinguished between shorter and longer term impact. However, we expect that the time taken for the impact of Gram Varta to manifest varies by the type of outcome. We discuss this in greater detail in the next section on impact timeline.

2.4 Potential impact timeline

Our evaluation examined the impact of Gram Varta on a variety of outcomes. We did not anticipate great changes in all outcome indicators given that we were expected to collect endline data almost immediately after the last PLA meeting was implemented. This tempering of our expectation was based on several reasons. Firstly, at the core of this theory of change is behavioural change, which is known to build on a change in social norms. These do not change quickly, in general. Secondly, the meetings were spread over a time period of nine months to one year, meaning that the last meeting where the last topic was discussed was held a few days or weeks before our endline survey data collection. This drastically reduced the time participants had to process, reflect, discuss, and adopt changes. Thirdly, it takes time before a change in health- and nutrition-related behaviour translates into better health, measurable for example by anthropometrics.

Memorable images and games used in meetings may have an immediate impact on participants' awareness and attitudes, for example, regarding the importance of handwashing and proper sanitation. However, changes in behaviour may not be seen until a few more weeks have passed. The translation of changed practices into improved objective health outcomes such as weight or height is likely to take even longer. Figure 2.2 presents a rough timeline of when changes could be expected. Please note that time has to be seen in relation to the time since the meeting where a specific topic was discussed.

Figure 2.2: Impact timeline



2.5 Research hypotheses

Our research hypotheses were grouped into seven blocks. Each block consisted of several related hypotheses, and each hypothesis was tested by examining one or more indicators of an outcome. Whenever possible, the data to test each hypothesis were taken from all the relevant groups in our sample such as women in households, pregnant women, community mobilizers and Anganwadi workers.

Group 1: Women's self-help groups

Women's self-help groups (SHG) were at the center of the Gram Varta programme. One part of our analysis therefore focused on the processes occurring within the SHG groups and looked at the interaction of SHG members with the community. To this end, we used data from the facilitator surveys at midline and endline. Specifically, we tested the following nine hypotheses:

H1: Gram Varta improves participation in SHG meetings.
H2: Gram Varta increases SHG acceptance in and cooperation with community.
H3: Gram Varta improves provision of information on health practices within SHG.
H4: Gram Varta improves provision of health finance information and practices within SHG.
H5: Gram Varta improves awareness and usage of government health services by SHG members.
H6: Gram Varta improves frequency of promoting enrolment to VO nutrition center.
H7: Gram Varta improves awareness and usage of VO nutrition centers by SHG members.
H8: Gram Varta improves facilitators' health knowledge.
H9: Gram Varta improves facilitators' opinion on their work.

Group 2: Women's agency and empowerment

Next, we looked at Gram Varta's effect on women's agency and empowerment. The ways in which the programme may potentially impact women's standing in the family and community are manifold. First, we anticipated that the participatory learning approach would encourage self-help group members to think more critically and to question societal standards as well as make them more comfortable at voicing their opinion in public. Second, we expected to see that adolescent girls indirectly benefit from their mothers' involvement and develop a stronger self-esteem and a more positive outlook on their future. Lastly, we expected that the involvement of men in some SHG sessions would also increase the willingness of husbands to support their wives and to rethink cultural norms about gender relations. The data used to test the following hypotheses were taken from the household and pregnant women surveys.

H10: Gram Varta encourages women to acquire paid work and to become economically more independent.
H11: Gram Varta increases women's bargaining power within the household.
H12: Gram Varta enables women to become more independent of their husbands.
H13: Gram Varta enables women to develop an identity of their own.
H14: Gram Varta increases women's involvement in the community.
H15: Gram Varta makes women become more comfortable at speaking out in public.
H16: Gram Varta makes women self-confident when it comes to refusing sexual intercourse with husband or demanding him to use a condom.
H17: Gram Varta reduces women's acceptance of domestic violence.
H18: Gram Varta reduces the practice of domestic violence and oppression.
H19: Gram Varta gives adolescent girls a more positive outlook on their future.
H20: Gram Varta reduces adolescent girls' preferred number of children.
H21: Gram Varta decreases women's preference for sons.
H22: Gram Varta makes it more likely that women desire a higher age at marriage for themselves or their daughters.
H23: Gram Varta reduces the likelihood of early pregnancies.
H24: Gram Varta improves attitudes towards and practices of care for daughters.
H25: Gram Varta makes adolescent girls more confident to cook for themselves in order to take care of their own nutrition.
H26: Gram Varta increases husbands' support for their wives' SHG membership.

Group 3: HNWASH knowledge and practices

The key element of Gram Varta was the uptake of health knowledge among SHG members through the use of participatory learning methods. In order to evaluate the programme's success, it was hence crucial to look at the changes in HNWASH knowledge and practices in SHGs and the community. We expected households and pregnant women in treatment areas to exhibit a better understanding of basic issues in health, hygiene and nutrition; as well as to be inspired by their knowledge and adopt healthier practices. The data we used to test the hypotheses in this section were taken from the household and pregnant women surveys.

H27: Gram Varta increases the intake of micronutrients.
H28: Gram Varta raises awareness of importance of balanced diet for family.
H29: Gram Varta improves knowledge and attitudes toward proper feeding of newborns.
H30: Gram Varta encourages parents to prevent diseases in children, e.g. through vaccinations and bednets.

H31: Gram Varta reduces risky consumption behavior (tobacco/alcohol).
H32: Gram Varta improves domestic storage and treatment of water.
H33: Gram Varta improves domestic hygiene (hand-washing/use of toilets).
H34: Gram Varta increases women's awareness about infectious diseases such as malaria.
H35: Gram Varta improves adolescent girls' and women's knowledge about sexuality and contraception.

Group 4: Pregnancy

An important group of beneficiaries of Gram Varta were pregnant women, given that their health knowledge and practices do not only affect themselves but also the health of their unborn children. This group of hypotheses therefore looked particularly at pregnant women's practices during pregnancy as well as their behaviours related to antenatal care.

H36: Gram Varta encourages pregnant women to be mindful of their health, to take healthy and sufficient diet, to avoid stress and to avoid health risks.
H37: Gram Varta increases the frequency of antenatal care visits as well as their quality.
H38: Gram Varta increases support to pregnant women for obtaining antenatal care.
H39: Gram Varta increases pregnant women's satisfaction with antenatal care.
H40: Gram Varta encourages women to accept their pregnancy, making them more optimistic about their situation.
H41: Gram Varta raises awareness and understanding of abortions among pregnant women.

Group 5: Anganwadi centers

The participation of local service providers in selected SHG meetings was a key element of Gram Varta. We expected that these meetings would improve the health knowledge of Anganwadi workers and increase the quality of services offered by them. To assess the following hypotheses, we used data collected directly from local Anganwadi workers.

H42: Gram Varta increases use of Anganwadi health centers.
H43: Gram Varta improves malnutrition treatment and prevention.
H44: Gram Varta improves the quality of child weighing practices.
H45: Gram Varta improves the cleanliness of and hygiene practices at the Anganwadi center.
H46: Gram Varta improves quality of work and activities related to preschool children.
H47: Gram Varta improves counseling for pregnant and lactating women.
H48: Gram Varta improves postnatal care.
H49: Gram Varta improves Anganwadi workers' health knowledge.
H50: Gram Varta improves immunization practices.
H51: Gram Varta improves facilitation of routine check-ups.
H52: Gram Varta increases job satisfaction of Anganwadi workers.
H53: Gram Varta improves participation of Anganwadi workers in community health events.
H54: Gram Varta improves perception of and respect for Anganwadi workers by community.

Group 6: Health outcomes

If Gram Varta was successful in improving health knowledge, service uptake as well as in altering health-related behavior, we expected to see modest improvements in the health status of women, men, and children at the endline. To assess whether this was true, we used a mix

of verbal autopsy data and anthropometric measures from the household and pregnant women surveys.

H55: Gram Varta improves women's health.
H56: Gram Varta improves husbands' health.
H57: Gram Varta improves child health.

Group 7: Social cohesion

We expected Gram Varta to strengthen the general social relationships within the community. To test our related hypotheses, we used data on indicators collected in the household surveys.

H58: Gram Varta increases mutual trust within the community.
H59: Gram Varta reduces tensions in the neighbourhood.

3. Context

3.1 Study site

The state of Bihar

Bihar is a predominantly rural state in northeastern India. According to the 2011 census, Bihar has a population of 104,099,452 (Population Census, 2011) and it is one of the poorest states in India (Datta, 2015). Bihar has been ranked last on the Indian government's human development index for several decades (Daniel, Masilamani, and Rahman, 2008). It is one of the states with the worst housing conditions, lowest latrine coverage, and lowest asset ownership (Population Census, 2011). According to the latest National Family Health Survey, the infant mortality rate is 49 per 1,000 live births in rural areas (IIPS, 2016).

The National Family Health Survey presents the population profile of the state as follows (IIPS, 2016). In rural areas of Bihar, 54.1 percent of the households have electricity and 20.7 percent use improved sanitation facilities. Only 10.8 percent of rural households use clean fuel for cooking, compared to 63.8 percent of urban households. Literacy is much lower among women than men, with a stark contrast especially in rural Bihar, with 46.3 percent of women literate and 75.2 percent of men. Women marry young, as 40.9 percent of rural women aged 20 to 24 years were married before they reached the age of 18 years. Among currently married women aged 15 to 49 years, 22.6 percent use a family planning method, but this method is mostly female sterilization. Child feeding practices and the nutritional status of children are suboptimal. Only one third of children under age three years were breastfed within one hour of birth. Almost half of children under five years are stunted and only 7.4 percent of all rural children aged 6 to 23 months receive an adequate diet. Gender-based violence persists as 43.7 of ever-married women have experienced spousal violence. About 75 percent of currently married women usually participate in household decisions and 60.9 percent of women own a house or land alone or jointly with others.

The district of Madhepura

The district of Madhepura, the site of our evaluation, is located in the Mithilanchal region in the northeastern corner of Bihar, adjoining Saharsa, Purnea, Supaul, Araria, and Katihar districts. Being next to the river Kosi, it is prone to devastating floods which have had an extremely detrimental impact on the district's economy, social capital and the livelihoods of its people (UNDP, 2009).

The region is one of the most underdeveloped regions of the state of Bihar, with an infant mortality rate of 64 per 1,000 live births (Government of India, 2013). A Jeevika survey classifies it as having one of the highest poverty rates in Bihar (Bihar Rural Livelihoods Promotion Society, 2002). Nearly 95 percent of the district population is based in rural areas of which 37 percent live below the poverty line (Directorate of Census Operations, 2012). The primary employment in the district is agriculture and related businesses. The main crops in the region are rice, corn, wheat, mustard, soybean, and bamboo. There are usually two cropping seasons (Bihar Rural Livelihoods Promotion Society, 2002). The district has an extremely low literacy rate at around 52 percent of the total population (Directorate of Census Operations, 2012).

Recently, the central government has announced and begun work on a Rs. 20,000 crore locomotive plant near the town of Madhepura, the district headquarters. The plant is expected to directly or indirectly employ nearly five thousand people and is set to be functional by 2019. This will be the first major investment in Madhepura in terms of an industrial production facility.

In terms of the local population demographics, the proportion of vulnerable population groups such as Muslims and Scheduled Tribes is not as high as in some other regions in Bihar. However, against a state-wide average of 15 percent Scheduled Caste population, the district population comprises of 17 percent of Scheduled Caste populations. Discrimination along caste lines persists (Daniel, Masilamani, and Rahman, 2008).

It is thus clear that Gram Varta is being implemented in low-income, resource-poor populations. Furthermore, it is being implemented with women as agents in a highly patriarchal setting (Government of India, 2016). However, there is strong political support for the programme and for the SHGs in general. In addition, state-supported societies implement the programme, suggesting a strong support structure. Jeevika, which is the implementing agency in Madhepura, is state-sponsored and started operating in Bihar in 2006, mobilising rural women to set up SHGs with a focus on microfinance activities (Datta, 2015). Jeevika has been successful in reducing debts of the beneficiaries and empowering women in various dimensions (Datta, 2015).

Overall, Madhepura and its characteristics are representative of the major districts in northeastern Bihar. Madhepura was chosen for the impact evaluation for two reasons. Firstly, the district's economic underdevelopment makes results about the programme's effectiveness indicative for other potential programme sites where extreme poverty is still prevalent. Secondly, the timing of the Gram Varta implementation in Madhepura allowed us to conduct a pre-post evaluation using a randomized trial. This created the opportunity for a more rigorous evaluation than one based on information collected only after implementation.

3.2 Sample description and external validity

The extent to which results from the Gram Varta evaluation can be transferred to Bihar or India as a whole depends on the representativeness of the sample population. A priori, there are reasons for concern: First, Bihar is one of the poorest states in India and findings from the district of Madhepura may therefore not generally be transferable to the national level. Second, the results of our study may not be generalizable to Bihar because our sample is limited to selected gram panchayats in one district that was chosen based on practical considerations, including the existence of local women SHGs, rather than a random draw.

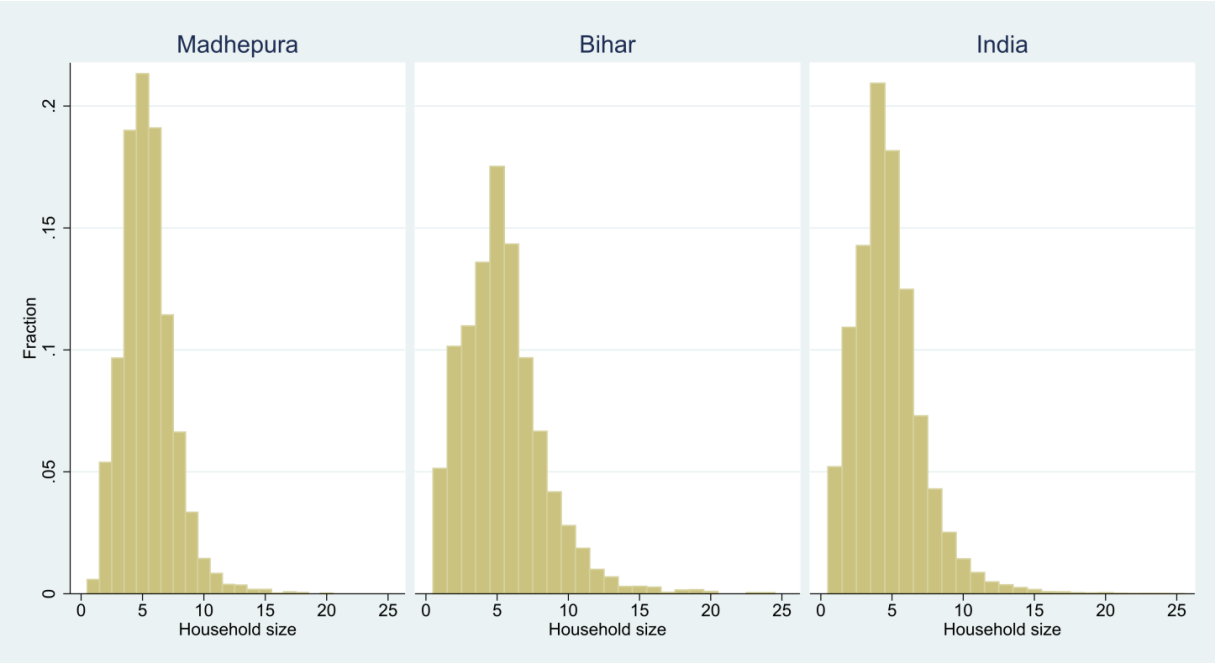
To alleviate these concerns, this section provides a detailed description of the sample characteristics and compares the quantitative data collected in our baseline household survey to results from the Demographic and Health Surveys (DHS). DHS are nationally representative cross-sectional surveys of women of reproductive age, their children and households and are administered by ICF International in cooperation with local governments of several countries. So far, India has participated in three rounds of DHS, with a fourth one currently ongoing. In India, the DHS are also called as National Family Health Surveys (NFHS). The latest publically available DHS in India was finalized in 2006 and the data from this survey (for Bihar as well as India) will serve as a comparison to our current study.[2] In this DHS, 109,401 households were randomly sampled overall, with 3,016 households surveyed in Bihar. These data from latest publically available DHS (NFHS-3) are slightly outdated, but provide more detailed information than the more recent National Family Health Survey-4. We therefore describe some general recent findings in the first part of this chapter, but are certain that the third DHS provides precise, albeit slightly outdated, estimates for the Indian and Bihari reference populations. In order to assess whether our study means are statistically different from those of the reference populations, we calculate 95% confidence intervals.

Demographics and gender

At baseline, we sampled 3,953 households from 68 Gram Panchayats. Households had an average size of 5.4 (95%-CI: 5.4; 5.5) members. This is in line with households in Bihar which also had an average of 5.4 members in 2006. At the national-level, the average size of households for India was 4.7 members, suggesting that our sample's average household size was more representative of Bihar than of India as a whole. While the national-level estimate lies outside the 95%-confidence interval, the difference is quantitatively small, suggesting that our sample is representative as far as household sizes are concerned.

Figure 3.1: Relative frequencies of household sizes

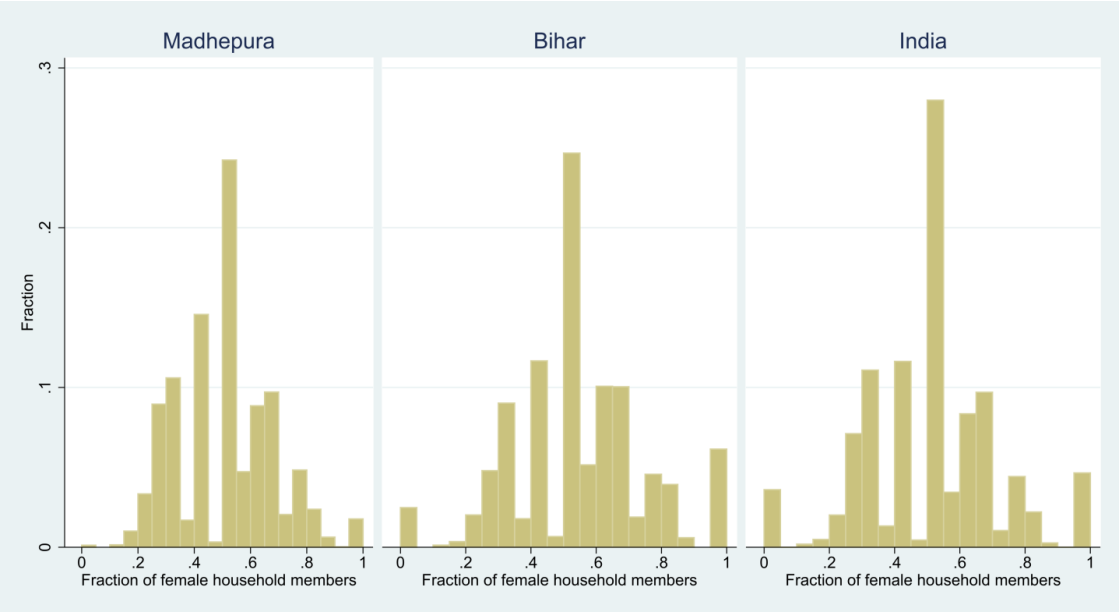
[2] DHS datasets are available for download from the DHS Program's homepage: http://www.dhsprogram.com/



This interpretation is supported by examining the relative frequencies of each household size, depicted in Figure 3.1. In our Madhepura sample, which we used to evaluate Gram Varta, and the two DHS populations, most households had between three to six members. Yet, at the national level four-person households were most common, while in Bihar and in the Gram Varta sample, households of five members were more frequent. What is striking, however, is that the shapes of the histograms depicted in Figure 3.1 are very similar for all three regions, with a strong concentration of medium-scale households.

In addition to the similarity in this crude measure of living arrangements, our sample households were largely comparable to the DHS population estimates in terms of gender ratios. On average, households in our sample had a female share of 49.5 percent (95%-CI: 48.9; 50.0), indicating no substantial bias to either gender. According to the 2006 DHS data, this is close to the state-level average, with a female fraction of 53.3 percent, and nearly identical to the national average of 50.0 percent. While the differences are statistically significant again, we do not find quantitatively relevant deviations. Nevertheless, it is interesting to analyze the entire distribution of values, as means could potentially hide important differences. This is done in Figure 3.2. Once again the distributions look very similar across regions. In our Madhepura sample, Bihar and India, the largest fraction of households had a female share of 50 percent to 55 percent.

Figure 3.2: Relative frequencies of shares of female household members



Next, we compared the age-structure of our household sample to the DHS population estimates for Bihar and India. The average of our study household member's mean age was 25.7 years (95%-CI: 25.3; 26.0). This is again very similar to the state level (26.7 years), albeit statically significant. Compared to India as a whole (29.6 years), however, our sample households were substantially younger. To investigate this issue in greater detail, we further considered the relative frequencies of the household-level age means. Taking household-level means before averaging, in contrast to simply plotting the relative frequencies of each age, allowed us to obtain valuable information about living arrangements. For example, if the mean age in a household is above 60, it is very likely that the household does not include any children or young adults. Histograms for Madhepura, Bihar and India are depicted in figure 3.3.

Despite the rather substantial deviations in the overall average between our study sample and the national level, all three histograms are very similar in shape. Notably, in all three datasets, households with a mean age of 20 to 25 years were most common and household with a mean age above 50 were rather seldom found. These similarities suggest that our sample is indeed largely representative of the age distribution and living arrangements in Bihar and India.

Figure 3.3: Relative frequencies of mean age of household members

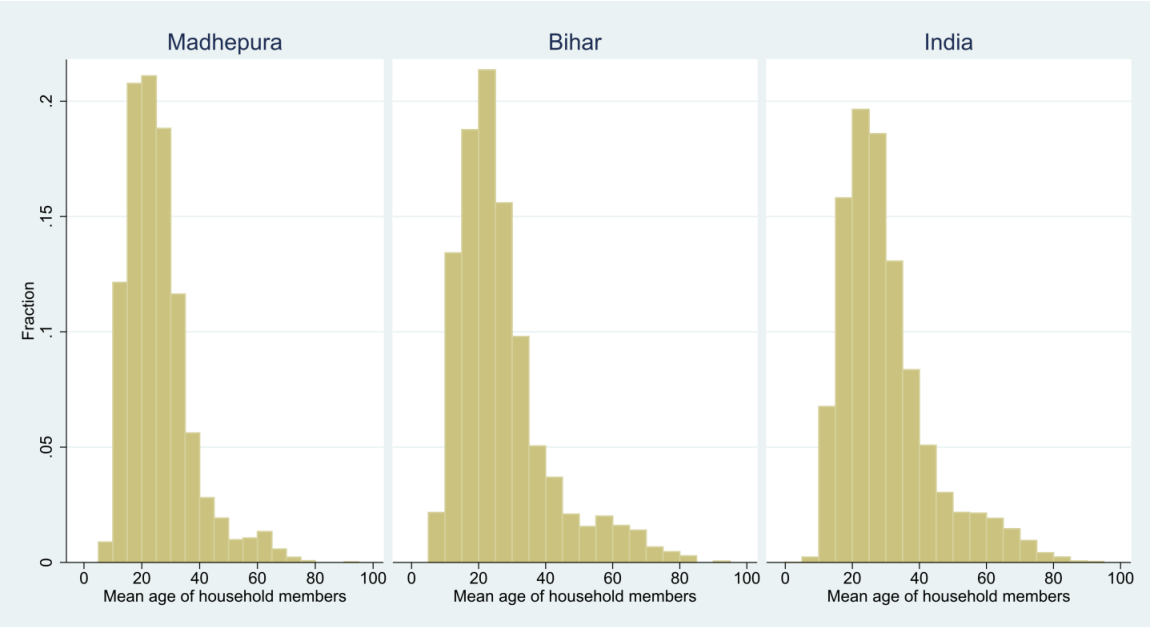
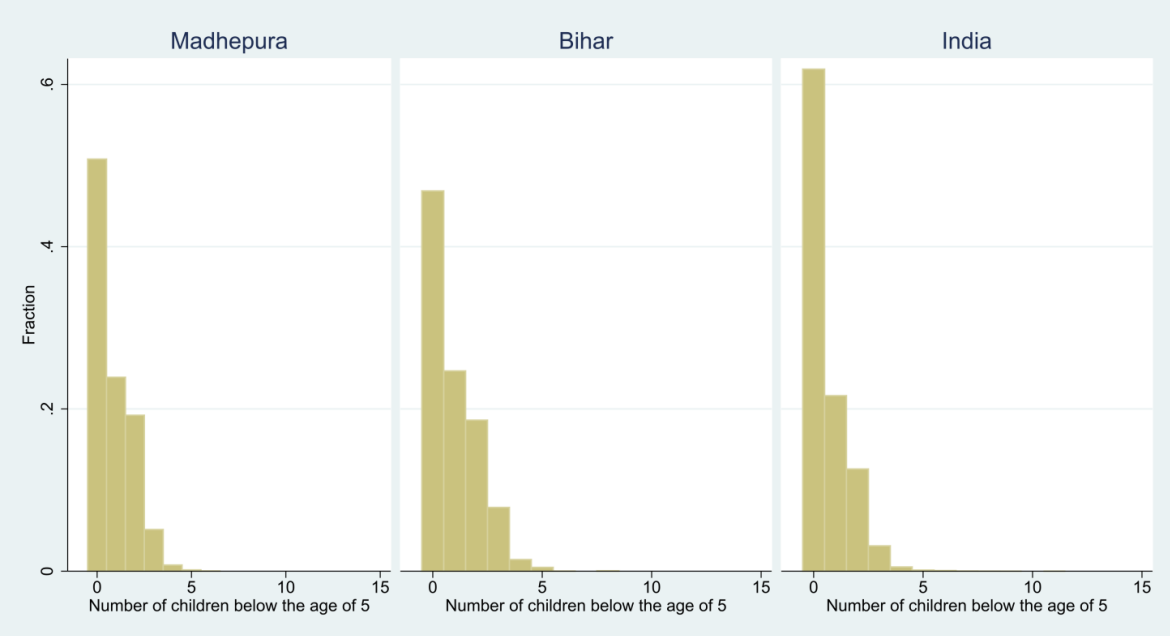


Figure 3.4: Relative frequencies of number of children below the age of five



Finally, we considered the number of young children in households. In our sample, the average household had 0.82 (0.79; 0.85) children below the age of five. This is slightly lower than the state-level average (0.94) and higher than the national-level average (0.62). Figure 3.4 plots the respective histograms which show that the relative frequencies of number of children below the age of five were nearly identical in the Madhepura sample and in Bihar. At the national level, families with no children below the age of five were more common, but the difference is not large. Importantly, in all three cases, it is very unlikely that in a household there were more than four children in that age group. Taken together with the previous figures, the concern that our sample may not be representative on the state or national level seems to be rather unfounded.

Education

An important dimension to assess the potential of our evaluation in producing meaningful results for the state and national level is representativeness of our sample in terms of educational attainment. Again, we focused on household-level averages in order to incorporate the actual living arrangements of individuals.

Table 3.1 lists (estimated) fractions of level of completed education with 95 percent confidence intervals in parentheses. Accordingly, there are rather substantial differences in education levels in the different datasets. While in the Madhepura sample on average 47 percent of household members were either uneducated or had not completed primary school, this fraction was higher in both Bihar (73 percent) and India (56 percent). Moreover, the share of household members with primary or incomplete secondary education was 34 percent in our study sample, while it ranges from 21 percent in Bihar to 33 percent in India. Finally, completing secondary education or more was the most common category in our sample, with on average 19 percent of household members falling into that category, while at the state level (5 percent) and the national level (10 percent) a substantially higher fraction of household members remained on average below this level.

Table 3.1: Average fraction of household members with selected education levels

	Madhepura	Bihar	India
No education/incomplete primary	0.47 (0.46;0.48)	0.73	0.56
Primary/incomplete secondary	0.34 (0.34;0.35)	0.21	0.33
Secondary or higher	0.19 (0.18;0.20)	0.05	0.10

While these differences are important, they may partly be due to the fact that the DHS data were collected in 2006 and general advancements in education since then may explain part of the differences, making our sample on average more educated than the state- and national-level reference populations.

Household assets

Lastly, we considered household wealth. Table 3.2 lists indicators of livestock ownership, agricultural land ownership and further assets. Again, (estimated) fractions are provided with 95 percent confidence intervals in parentheses below. Differences in ownership of livestock and agricultural land were rather small overall. While 44 percent of households in our sample owned agricultural land the fraction was slightly higher in Bihar (51 percent) and India (46 percent). Similarly, the fraction of households owning sheep/goats and the fraction of households owning donkeys did not vary substantially between the datasets. Cattle ownership, in turn, was slightly more common in the Madhepura sample than in Bihar and India, while fewer households in our sample owned poultry compared to the state and national level.

Regarding further assets, it is notable that ownership of mobile phones was substantially more common in our sample (76 percent) than in Bihar (9 percent) and India (17 percent). In

contrast, ownership of radios, watches and televisions was below 10 percent in the evaluation sample, while the numbers were substantially larger on the state and national level, especially concerning watches. Note, however that the nine year distance between the Madhepura sample and the DHS reference data may explain a large part of these deviations since mobile phones are increasingly common in all parts of the world and may for many people make watches or radios obsolete. In terms of transportation-related assets, which may be a better proxy for wealth due to their higher price, differences were less pronounced, especially when the sample was compared to the population estimates for Bihar. Owning at least one bicycle was very common in all three cases (>50 percent of households), while the ownership of motorbikes was rare in our sample (3 percent) and Bihar (8 percent) it was more common in India (17 percent). Moreover, in all three regions, households were roughly equally unlikely to own cars, animal-drawn carts, tractors or threshers, while sewing machines were unlikely to be encountered in Madhepura.

Table 3.2: Household assets (in percent) in study sample, Bihar and India

	Madhepura	Bihar	India
Livestock ownership			
Poultry	0.01 (0.01;0.02)	0.12	0.15
Cattle	0.61 (0.59;0.62)	0.49	0.37
Donkeys	0.01 (0.00;0.01)	0.01	0.00
Sheep/goats	0.24 (0.23;0.25)	0.31	0.17
Owns agricultural land	0.44 (0.42;0.45)	0.51	0.46
Further assets			
Mobile phone	0.76 (0.74;0.77)	0.09	0.17
Watch	0.07 (0.07;0.08)	0.63	0.78
Radio	0.03 (0.02;0.03)	0.31	0.31
Television	0.03 (0.03;0.04)	0.18	0.44
Electric fan	0.10 (0.09;0.11)	0.22	0.54
Bike	0.53 (0.51;0.54)	0.53	0.51
Motorbike	0.03 (0.02;0.03)	0.08	0.17
Animal drawn cart	0.05 (0.04;0.06)	0.02	0.05
Car	0.00 (0.00;0.01)	0.01	0.03
Tractor	0.01	0.02	0.02

	(0.01;0.02)		
Sewing machine	0.03 (0.03;0.04)	0.12	0.19
Thresher	0.01 (0.00;0.01)	0.03	0.02

All in all, while differences clearly exist in terms of asset ownership, the observed patterns do not indicate that our sample households were substantially poorer or richer than households at the state and national level. Instead, households appeared to make different consumption choices as far as assets of similar value were concerned, which can at least partly be explained by the time gap between the collection of these data.

Summary

The discussion in this section has shown that while our sampling strategy does not by itself guarantee external validity, there are reasons to believe that our main sample is still relevant and representative for Bihar and India as a whole. Most importantly, very little differences in terms of household demographics and gender compositions could be identified. Moreover, despite different asset ownership patterns, the data do not indicate that substantial wealth differences between our sample households and the DHS populations exist. One concern that needs to be raised, however, is the difference in the distribution of educational attainment. Although, these could at least partly be explained by advancement in general education levels since 2006.

4. Timeline

In terms of implementation, the Gram Varta SHG meetings were severely delayed, leading to major delays and changes in the evaluation. By January 2016, only four PLA meetings had been conducted in the treatment areas. The implementation timings had to be revised to conduct PLA sessions 5 to 8 with a gap of one week between meetings instead of the planned gap of a fortnight. This affected the content of the midline questionnaire which was scheduled to be administered in March 2016.

Secondly, the CEO of Jeevika changed twice in the first quarter of 2016. While this disrupted communication slightly, the evaluation team received necessary support from Jeevika and its officials throughout the evaluation due to the efforts of BTAST.

A Steering Committee was constituted to monitor the implementation of Gram Varta. It included members from the health ministry, the public health engineering department, the ICDS department, the rural development and social welfare ministry, and the implementing agencies of Jeevika, the Women's Development Corporation, and the Bihar Mahila Samakhya Society. Unfortunately, there was no regular communication between the relevant departments; and the communication between the implementation agencies and the government officials was mostly need-based. The Committee could have been a useful platform to ensure that the implementation was occurring on schedule, and that the service providers such as the health department and the ICDS scheme were kept in the loop about the implementation. Any

shortcomings in service could have been communicated to the relevant departments at these meetings.

Table 4.1: Planned and actual timing of baseline (B), midline (M) and endline (E) surveys

Questionnaire	2015		2016	
	Planned	**Actual**	**Planned**	**Actual**
Household	(B): March-April (M): October-November	(B): March-April	(E): June-July	(M): March 2016 (E): November-December
Pregnant woman	(B): March-April	(B): March-April	(E): June-July	(E): November-December
Community Mobilizer	(B): June-July	(B): December 2015-January 2016	(E): June-July	(E): January 2017
Anganwadi worker	(B): June-July	(B): October-November 2015	(E): June-July	(E): January 2017
Qualitative data	(B): June-July	(B): June-July	(E): June	(M): March (E): December 2016-January 2017

The initial evaluation timeline was planned as laid out Table 4.1. This timetable was strictly based on the assurance by the implementing agencies that the implementation would begin in March-April 2015 and be completed by March 2016 in the selected treatment areas. There were multiple delays during the implementation of Gram Varta beginning with delayed recruitment and training of the facilitators. The primary facilitator survey was planned in April 2015, while the actual survey had to be conducted nearly six months later in December 2015 and January 2016. The baseline qualitative study was also delayed and had to be conducted in July 2016 due to the delays in recruitment of the facilitators. The endline survey for Gram Varta was planned for June 2016 while it had to be actually held in November-December 2016. The endline CM and AWW survey were ultimately conducted in January 2017. This concluded the field activities for the Gram Varta evaluation on January 27, 2017.

5. Evaluation: Design, methods and implementation

The impact evaluation of Gram Varta combined a randomized design with a difference-in-differences approach. As actual participation in all SHG meetings could not be observed, our estimates generally represent intention-to-treat effects, i.e. they represent the effect of having the opportunity to participate in Gram Varta rather than the effect of the actual participation. While this potentially lowers effects, especially if treatment uptake is small, the intention-to-treat effect is a valuable estimate for policy makers, as it reflects the average effect of the treatment on the population that was offered the treatment. The quantitative analysis was substantiated by qualitative information, to facilitate the understanding of possible channels of change which can help to explain the presence or absence of certain effects. The following subsections provide detailed information on the estimation strategy, sampling, and data collection of our evaluation.

5.1 Estimation strategy

To quantify the impact of Gram Varta, we used a randomized controlled trial design, the gold standard methodology to identify causal effects. We randomly assigned Gram Varta implementation to gram panchayats with a probability of 50 percent. A post-intervention comparison of the treatment and control group, given baseline characteristics, allowed us to measure the causal impact of Gram Varta. Despite randomization, it is possible that imbalances between the treatment and control group in our experiment (in observable or unobservable characteristics) have occured by chance, thus leading to selection bias. We approached this problem in two ways: First, we reestimated all post-intervention comparisons controlling for a number of potential confounding variables. Second, we exploited the panel-structure of our study and employed a difference-in-differences (DiD) approach. The idea of this approach is to combine the advantages of before-after comparisons and between-group comparisons at endline by focusing on the difference in changes over time. This effectively eliminates all baseline differences in outcomes as well as shared trends between treatment and control groups. We further augmented our analysis by estimating separate intention-to-treat effects by subgroups. To this end, we stratified our sample by caste, block, age, and education and ran separate regressions of the types outlined above.

We performed additional analysis using the household and pregnant women sample by taking into account treatment intensity. The fact that participation in SHG meetings is voluntary made it likely that there would be differences in the treatment intensity across women. We exploited the information on whether the woman respondent was a member of a Jeevika SHG and whether she attended meetings “never,” “rarely,” “sometimes,” or “mostly.” A household was considered to be exposed to Gram Varta if the woman respondent was a Jeevika SHG member and went to meetings “sometimes” or “mostly.” Please note that this was in addition to the intention-to-treat analyses described above. For this additional analysis, we made the reasonable assumption that if the woman belonged to an SHG set up by Jeevika and it was in the treatment area, it would go through the PLA meeting cycle of Gram Varta. Women who went to these SHG meetings regularly (“mostly”) or sometimes were therefore considered as exposed to the intervention. We did not use questions that ask specifically about participation in Gram Varta meetings to define exposure. This is because the answers to this question did not seem reliable: Many respondents said they had not participated or even heard of Gram Varta (implementers did not publicize the name of the programme), but according to qualitative data, they described meetings in which they participated in a way that strongly resembles Gram Varta meetings. Many respondents therefore seemed not to know that specific meetings were Gram Varta meetings, although they participated in them. We then compared exposed women or other exposed household members in the treatment and control group using the estimation strategies outlined above. As the exposed households were those who actually experienced treatment, these estimated effects can be interpreted as the treatment-on-the-treated effect as opposed to the intention-to-treat effect.

5.2 Ethical clearance

As with every potentially advantageous randomized intervention, individuals in the control groups might be treated unfairly to the extent that they do not receive the treatment (in case the treatment is actually beneficial). However, without random assignment and an untreated

counterfactual group, rigorous impact evaluation of statewide implemented programmes or policies is hardly possible and it remains unknown if the treatment is actually beneficial.

To reduce the potential disadvantage to the counterfactual group to the minimum, a phase-in design was chosen and Jeevika staff in Patna shared that Gram Varta would be introduced to control villages starting early 2017. Ethical concerns related to informed consent and data confidentiality were actively addressed, especially given that sensitive information (such as opinions on service providers or SHG leaders) was shared by respondents. First, all field and research staff was trained on data sensitivity and confidentiality. Second, confidentiality of data was maintained by anonymizing it. All participants were assigned an ID number when they agreed to participate. Copies of forms linking ID numbers to names were stored separately.

Ethical clearance was provided by the Institutional Ethics Committee of the Indian Institute of Technology Gandhinagar upon submission of a detailed proposal.

5.3 Sample size determination

Given budgetary constraints, our survey capacity was 4000 households from 200 villages (including women and adolescent girls) plus an extra survey of around 2000 pregnant women. To draw conclusions about whether this was a feasible sample size we carried out power calculations for several important indicators prior to the evauation. We provide four examples here to give an idea on how much the sample size needed to detect statistically significant effects could differ. The calculations shown below are based on the following set of assumptions: (i) 200 villages in randomly selected gram panchayats [3], (ii) standard assumptions for power (0.8) and significance level (0.05), and (iii) an intra class correlation (ICC) parameter of 0.05. These are quite rough assumptions, however, reliable data on standard deviation within villages versus across villages in our study area were not available to us prior to our study.

[3] Eventually, 180 villages were randomly selected, a small difference from these power calculations.

Example 1: Assuming a number of 200 villages (100 in each group), we would need to collect information on 1800 children (9 per village) to detect a reduction of 5 percentage points in the child diarrhoea rate in the treatment group compared to the control group. This calculation is based on the children's diarrhoea rate for Madhepura for 2013 (13 out of 100 children had diarrhoea) (Government of India, 2013).

Example 2: Assuming a number of 200 villages (100 in each group), we would need to collect information on 8600 children (43 per village) to detect a reduction of 5 percentage points in the child fever rate in the treatment group compared to the control group. This calculation is based on the children's fever rate for Madhepura for 2013 from the Indian census (34 out of 100 children had fever) (Government of India, 2013).

Example 3: Assuming a number of 200 villages (100 in each group), we would need to collect information on 600 infants (3 per village) to detect a reduction of 5 percentage points in the infant mortality rate in the treatment group compared to the control group. This calculation is based on the infant mortality rate for Madhepura for 2013 from the Indian census (6 out of 100 infants died) (Government of India, 2013).

Example 4: Assuming a number of 200 villages (100 in each group), we would need to collect information on 3800 pregnant women (19 per village) to detect an increase of 5 percentage points in the neo-natal check-up rate in the treatment group compared to the control group. This calculation is based on the neo-natal check-up rate for Madhepura for 2013 from the Indian census (78 out of 100 mothers had any neo-natal check-up) (Government of India, 2013).

To get a better idea of how the required sample size varied with different ICC assumptions, we ran power calculations assuming an ICC of 0.03 and 0.07 as a robustness check. It turned out that sample size for some of the indicators varied substantially with this small change in the ICC parameter. The following textbox documents how the necessary sample sizes varied in our examples with small adjustments to the ICC parameter:

ICC parameter:	**0.03**	**0.05**	**0.07**
Example 1:	1600	1800	2000
Example 2:	4600	8600	72400
Example 3:	600	600	600
Example 4:	2800	3800	6000

In appendix P we calculate and discuss minimal detectable effect sizes based on baseline information. Appendix A includes a table with ICCs for all outcome variables for which causal impacts are estimated.

5.4 Quantitative sampling design

For the quantitative part of our evaluation, we collected a rich data set comprising different survey components:

- 3953 household heads and women in reproductive age (household survey),
- 2000 pregnant women and their husbands (pregnant women survey),
- 316 community mobilizers (facilitator survey),
- and 265 Anganwadi workers (Anganwadi survey).

Sample selection

As mentioned above, this evaluation of Gram Varta assesses the effects of this programme in Madhepura. The district is divided into thirteen blocks comprising a total of 443 gram panchayats. However, Gram Varta was implemented in only six of these thirteen blocks and due to budgetary constraints, we restricted the sample to 68 gram panchayats from these six blocks, each of which contained about 2.6 villages on average. This yielded a total of 180 villages.

Selection of household survey respondents

For the household survey, households were sampled and different respondents in these households were selected to be interviewed. We interviewed the household head, a woman of reproductive age with young children (preferably under the age of five), and adolescent girls. In addition, we measured health indicators of all household members. The exact process of sampling of households and respondents in the villages is described below.

When the enumerators reached the villages, they talked to local residents as well as the gram panchayat head (Mukhiya) and roughly mapped out the different hamlets (tolas) in the village which were based on caste in most cases. They also inquired about the number of households in each hamlet. We planned to sample 22 households per village. Based on the total households in the entire village and the number of households in the different hamlets, enumerators arrived at the number of households in each hamlet they needed to interview given our sample size calculations. The number of houses selected per hamlet was proportional to the total number of houses in each hamlet as compared to the total number of households in the entire village (probability proportional to size of hamlet).

Once the enumerators calculated how many houses they needed to sample in each hamlet, and the total number of houses in that hamlet, they randomly picked the first house (ninth household from the public facility in the hamlet) and then chose every ninth house. Once they reached the house, the different survey respondents were selected. Table 5.1 below summarizes the different respondents and the criteria for selecting them. First, the enumerators selected the head of the household for the household survey. If the head was not available, they talked to the adults available at home and picked one adult who seemed to have the necessary information about the household. When getting information from the head of the household, the enumerators also identified the woman respondent, who was interviewed the next day by a female enumerator. At baseline, a woman who was in the 15-49 age group and had the youngest child in the household, compared to all other women of similar age group in the household, was supposed to be selected. If no such woman was available, they interviewed the woman with the next older child and so on. However, in practice often the female household head was mistakenly selected, who was not always the woman with the youngest child and in some cases older than 49. To increase the number of respondents that fall into our group of interest, we changed the selection methodology at midline and endline. In these survey waves, the wife of the oldest son of the household head was selected. If there was no wife of the son of the head in the household, any other woman in reproductive age was selected. While these selection criteria resulted in a considerably better targeting of female respondents, it reduced the number of panel observations: 66 percent of baseline woman respondents were reinterviewed at endline. Adolescent girls were interviewed on the same day as the woman respondents and all girls in the appropriate age range were selected for the adolescent survey.

All household members were asked to participate in the measurement of health indicators, while arm-circumference and oedema were only measured in children under the age of five (six at midline and seven at endline) and the stool sample was only collected from the last born child of the woman respondent.

Of the 3953 households we interviewed at baseline, we were able to follow up 3340 with all questionnaires completed at endline. Households were counted as incomplete if either the interview with the household head or the interview with the woman respondent or both could not be completed. If no questionnaire for the woman respondent was completed because no suitable women respondent existed in the household, the case was considered complete. We have information on the reasons for incomplete woman's questionnaire in the household survey. In 59 percent of these incomplete cases this was due to the woman not being at home even when we returned later on the same day or a few weeks later to follow up incomplete cases. Two factors were responsible for not meeting people at home: first, it was paddy harvest season, and second, due to the then ongoing demonetization policy, many people spent hours at banks and ATMs in order to get rid of the money that was of no value anymore and to get new bills instead. In 18 percent of the cases, the household could not be located, meaning the enumerators were not able to find the house where the woman was supposed to be living. They had to rely on other residents in the village pointing them in the right direction, which was especially difficult in villages that were spread out across a larger area and where people did not know each other well. In 14 percent of the cases, the woman refused to participate in the interview. In cases where the woman respondent was not at home, but an adolescent girl was available, we conducted the interview with the adolescent girl. This was possible since consent for conducting the survey had been taken from the household head by the male enumerators the previous day.

Table 5.1: Survey respondents

Household sample: 3953 households					
Questionnaire	**Household characteristics**	**HNWASH indicators, WCH indicators, female empowerment, social capital**		**Family planning, female empowerment**	**Health indicators**
Respondent	Household head	Woman respondent		Adolescent girls	All members
Selection criteria	As pointed out by the household members	Baseline: Woman between 15-49 years and has youngest child	Midline and endline: Wife of the oldest son of the head	All girls between 13 and 19 years of age	

Pregnant woman sample: 2000 households			
Questionnaire	**Household characteristics**	**HNWASH indicators, WCH indicators, female empowerment, social capital**	**Health indicators**
Respondent	Husband	Pregnant woman	Pregnant woman, husband, children under 5

Selection criteria	Husband of pregnant woman	Based on Anganwadi centre list	

Selection of pregnant women survey respondents

At the village level, sample selection was different for the sample of pregnant women. Pregnant women were oversampled, given that the share of this group would have been too low in the household sample to achieve sufficient statistical power. At baseline, a list of currently pregnant women was requested from the Anganwadi centers of that village. Pregnant women were randomly selected from those lists such that the number of women selected from a village was proportional to the size of that village.

When the enumerators reached the household, the pregnant woman's husband was selected to respond to the first part of the questionnaire, while the pregnant woman answered the remaining questions immediately after. At baseline, health indicators of the pregnant woman and husband were measured, while at endline they were also measured for all children under the age of five. Table 5.1 summarizes the different respondents and the criteria for selecting them.

Of the 2000 pregnant women and their husbands interviewed at baseline we were able to follow up only 1,612 at endline (80.6 percent). For 47 percent of the lost cases this was due to people not being at home even after several attempts were made to reach them. The reasons for this were the same as described above. Also, young pregnant women often live in their parents' home until they give birth, and later move back with their husbands who often lived in a different village. While the data indicate that only 9.4 percent of the pregnant women permanently migrated, we suspect that this was frequently miscoded as temporarily migrated in cases where enumerators were told that the pregnant woman was spending the next few months at her parents' home. Lastly, a number of women had died, most likely due to pregnancy complications.

Selection of community mobilizers and Anganwadi workers

For the survey of the Community Mobilizers (CMs) and Anganwadi workers (AWWs) the strategy was to interview all CMs and AWWs in the treatment areas. However, some of them were not available at the time of the endline survey or refused to be interviewed. In total we interviewed 316 CMs and 265 AWWs at baseline and 282 CMs and 233 AWWs at endline.

5.5 Qualitative sampling design

The evaluation study involved a qualitative appraisal of the experiences and social conditions influencing empowerment, health, hygiene, nutrition and sanitation in Madhepura district as well as outside Madhepura district. Qualitative interviews and focus group discussions (FGDs) were used to elicit personal experiences, stories and subjective opinions about the respondents' family and social life. The two sections below outline the sampling for qualitative investigations inside and outside Madhepura district.

Sampling for qualitative study in Madhepura

Three villages in Madhepura district - Satverkupari, Murho and Gangora - were selected for the study. Out of the three villages, Satverkupari and Gangora were chosen from the villages in the treatment area of the RCT and Murho was chosen from the control area. The villages were similar as per several relevant factors such as the proportion of marginalized classes and the degree of flood proneness. The goal was to select three villages such that their socio-economic conditions would be relatively similar. The detailed process of selection of respondents is given in a figure in Appendix M.

In order to identify the SHG members, the qualitative interviewer (Sini Varghese) first approached the CM, the trained Gram Varta facilitator. After explaining the reason for the visit, the interviewer requested information on members of the SHG and sought permission to attend one of the meetings. Members of the SHG were selected randomly from the roster that the CMs shared with her.

The respondents from the baseline qualitative study were followed up during the midline qualitative study. Unfortunately, two of the respondents who were pregnant at baseline had died due to pregnancy complications and lack of adequate emergency care. Moreover, most of the teenage respondents had been married off to different villages and were not available. Therefore, new respondents were recruited to the study. Care was taken to recruit respondents from the target group (women, pregnant women, or teenaged girls). Various new stakeholders such as Gram Varta meeting facilitators and pregnant women were added to the study during this round.

The midline qualitative study includes data collected from ten SHG members (including pregnant women), one facilitator, one Anganwadi/ASHA worker, three teenage girls (aged 12-17), three men, and three older adults from each village. Overall, 25 interviews were conducted in each village. Interviews had a larger proportion of women because women were the primary beneficiaries of Gram Varta and were directly involved in the programme as members of SHGs. One FGD was also held in each village in order to understand and compare collective opinion and group dynamics.

The respondents in the selected villages were further followed up during the endline qualitative study. We also conducted more FGDs during the endline as it was found that we could gather perspectives of different social groups though FGDs rather than through interviews. The major aim of this exercise was to find out whether there were any changes in atittudes and perceptions regarding HNWASH in these villages and the initially selected villages and to find out the reasons for such changes, if any. During the endline, five more women were included in the study. These were respondents who had indicated in the quantitative endline survey that they had been to Gram Varta meetings and their responses to questions suggested that they had greater knowledge about malnutrition, hygiene, and health compared to their responses at baseline. These women were selected as ‘potential success stories.’ FGDs were also conducted among different members in the villages with ‘potential success stories.’

Sampling for qualitative study outside Madhepura

In addition to the qualitative study performed in the district of Madhepura, we sampled respondents for a qualitative survey of implementers and beneficiaries of Gram Varta outside

Madhepura. The primary goal of this second qualitative study was to gain insight into the implementation and impact of Gram Varta in districts other than Madhepura. Multiple factors played a role in determining the sample of districts for these interviews. The sampling was done in consultation with the BTAST team with some inputs from both Jeevika and WDC. The districts where the implementation was being performed by the Bihar Mahila Samakhya Society were not considered since their implementation had been halted due to funding insufficiency and due to major staffing issues arising out of these funding lapses. The study tried to cover more ground in the WDC areas since the RCT was being conducted in a Jeevika district. The study covered a total of twelve blocks in six districts. Three of these districts (six blocks) were areas where only WDC was the implementer; Gaya district was chosen since it had both WDC and Jeevika Gram Varta implementation areas; and the two northern districts of Madhubani and Purnea were chosen from Jeevika implementation areas.
The study aimed to cover diverse geographical areas within the Gram Varta implementation areas of Bihar. Although all areas did not receive proportional representation, the study included districts from the south, south west, and northern regions of Bihar. The selection of blocks was done in consultation with the BTAST, WDC and Jeevika team members, attempting to select blocks which were in different stages of the implementation. The selection of villages, however, was left to the block level officials. We avoided villages or blocks which had received extensive visits from senior officials from Jeevika, WDC or BTAST to avoid selecting any 'model' blocks or villages for the interviews.

The profile of respondents was uniform across all blocks. Among the WDC blocks, the interviews were carried out with the block coordinator at the block level and the MIS coordinator if the block coordinator was not adequately experienced. At the village level, one supervisor, one facilitator, and two community members were interviewed. We tried our best to interview one woman respondent, and preferably one male respondent or an adolescent girl who had attended the meeting. However, it was difficult to locate male respondents for interviews and we also found it difficult to reach adolescent girls since our visits were made mostly in the morning, when most of the girls were in school or college.

In the Jeevika blocks, we interviewed the district health and nutrition manager at the district level, and the Gram Varta Executive at the block level. The rest of the respondents remained the same as in the WDC blocks.

5.6 Treatment assignment

To be able to estimate causal effects, we randomly assigned Gram Varta participation. The random assignment was done at the gram panchayat level and, in order to maximize precision, we stratified by block. To this end, we first listed all the gram panchayats in our study area with the block they belonged to. Next, we randomly assigned participation in Gram Varta to gram panchayat, separately in each block, with a probability of 50 percent. Our randomization approach yielded 34 gram panchayats, 90 villages and 1,973 households in the treatment group; while the control group comprised 34 gram panchayats, 90 villages and 1,980 households. Moreover, we interviewed 1,234 pregnant women in the treatment group and 766 pregnant women in the control group.

Programme implementation was strictly aligned with treatment assignment. Thus, contamination was only possible if control group participants travelled to treatment gram

panchayats to attend a Gram Varta meeting. Our sampling design ensured that treatment and counterfactual communities were separated by sufficient physical distance. Additionally, the cultural context and patriarchal system prevents women from frequent or any travelling to other villages and gram panchayats. This reduces the likelihood of contamination and spill-overs across gram panchayats.

Participants in neither group were aware that they were part of an experiment to evaluate Gram Varta (informed consent was obtained for all procedures though). Control group members were not informed by us that Gram Varta was being implemented in other gram panchayats. This was the only feasible way for keeping the evaluation design clean and reducing bias arising from Hawthorne or John Henry effects. As mentioned above, Jeevika had planned to implement Gram Varta in control regions in early 2017 to minimize the potential disadvantages of members of the control group.

5.7 Data collection

Quantitative data collection

For the quantitative part of the study, data collection comprised a household survey, a pregnant women survey, a facilitator survey, and an Anganwadi worker survey. For the household survey, one questionnaire targeted the household head and the other targeted a woman in reproductive age as well as adolescent girls living in the household. For the pregnant woman sample, one questionnaire was administered to both: the husband of the pregnant woman and the pregnant woman. This is summarized in Table 5.1.

Altogether, these surveys covered six different modules:

1. *Household questionnaire, 1st part:*
   Contained a household roster to record demographic information and health indicators of all household members. In addition, household level socioeconomic and nutrition information was collected.

2. *Household questionnaire, 2nd part:*
   This questionnaire was concerned with women, adolescent girls, child health and nutrition indicators. We also collected blood and stool samples from children as well as anthropometric data from all present household members.

3. *Household questionnaire, 3rd part:*
   This part contained sections on outcomes (e.g. general HNWASH practices in the household, related attitudes, beliefs and knowledge, perceived social norms, HNWASH related consumption behavior) and enabling context (e.g. economic situation of household). Also included were: Gram Varta participation and potential obstacles to participation (only treatment group) as well as indicators related to mediators such as speaking up in groups, a sense of empowerment, critical-thinking, feeling of community cohesion or tension, and trust in service providers.

4. *Pregnant women questionnaire:*

The questionnaire for pregnant women inquired about knowledge, beliefs, social norms and practices relevant to pregnancy, child birth, and the post-natal period. We also collected blood samples and weight/height data from these women and their husbands at baseline. In the follow-up we collected data on the women who were pregnant at baseline and their newborns.

5. *Facilitator questionnaire:*
   We collected data from Gram Varta facilitators on participation of all target groups in meetings (pregnant women, adolescent girls, marginalized women) and other process indicators (meetings were implemented as scheduled, any complications, whether frontline workers, men and elderly participated, whether they observed any changes in the community etc.). This survey was restricted to the treatment areas.

6. *Anganwadi worker questionnaire:*
   We collected data from the Anganwadi worker on indicators of service use. This survey was restricted to the treatment areas.

Data were collected at baseline, midline, and endline. The complete household and pregnant woman sample was surveyed at baseline and endline, while at midline, only a cross-section of 1000 households from the household sample was surveyed. The Anganwadi worker and facilitator survey were conducted in October-November 2015 for the baseline and then finally again in January 2017. There was no midline for these two surveys.

We used structured pre-coded questionnaires for data collection. At baseline and midline data were collected using paper questionnaires. At endline, data were collected electronically on tablets, using the software ODK Collect.

Data were collected by local enumerators, recruited from local colleges, and selected through interviews. Criteria for their selection were: attendance or completion of college; knowledge of local languages; literacy in local language and Hindi; understanding of local culture and values; and confidence in conducting interviews. All enumerators received a six- to ten-day training related to the specific structured questionnaires and were instructed regarding their conduct when visiting households. The training ensured that interviews were conducted in a consistent and comparable way. In order to limit enumerator bias, we strongly emphasized avoiding leading questions and enumerators were instructed not to deviate from the wording described in the questionnaires. However, in cases where the respondent only understood and spoke the local dialect, questions needed to be retranslated.

For the endline household survey, one team of 29 male enumerators and one team of 28 female enumerators were separately trained. Male enumerators were trained to interview the household head, collecting information about the household composition, socioeconomic background, and family nutrition (household questionnaire, part 1). Female enumerators were trained to interview the main woman respondent and adolescent girls about WCH indicators, HNWASH indicators, reproductive health, female empowerment, and social capital (household questionnaire part 2 and 3). For the pregnant women questionnaire, 18 female enumerators were trained to interview the woman’s husband, collecting information about the household composition, socioeconomic background, and family nutrition, as well as the (formerly) pregnant woman about pregnancy, antenatal care, child health, and female empowerment. For the facilitator survey and the Anganwadi worker survey 10 male and 13 female

enumerators were trained for each survey along with 1 supervisor for each team of either gender. All female enumerators were specifically instructed about conducting interviews in privacy and avoiding direct influence of any other household member on the respondent. The training included exercises where experienced enumerators acted as a respondent, creating different scenarios for the practicing interviewer to manage.

For the collection of anthropometric measures (household questionnaire part 1 and pregnant women questionnaire), 24 female enumerators (18 for the household survey and 6 for the pregnant woman survey) were trained by a German medical student. The training started with an awareness session on hygiene and transmission of diseases, specifically the risk of disease spread through blood when measuring hemoglobin. The main part of the training focused on practical aspects of using the measurement equipment. The enumerators also received instructions on how to conduct themselves, especially how to approach children for taking the measurements.

Enumerators who did not perform well during the training were not allowed to continue with data collection. The groups of enumerators were split into smaller teams of 10 and each team was allocated a supervisor. This facilitated logistics such as allocation of cars for travel to the village and tracking of households in the village. All enumerators and supervisors were managed by the same local field manager throughout all waves and comparison groups as well as a group of student assistants from IIT Gandhinagar and the University of Goettingen.

The training did not differ between treatment and comparison groups as the questionnaires did not differ between the groups. In fact, all enumerators conducted interviews in both treatment and control regions and were not aware of which village belonged to the treatment area and which did not. These measures were taken to avoid enumerator bias.

In the household sample, the team of male enumerators established the first contact with the households. They identified the households in the villages using location information collected during baseline, noting down additional useful location information, and took consent from the household head for the interviews and anthropometric measurements. The team of female enumerators visited the same households one day after the male enumerators, using the more detailed location information provided by the male team. The female team was accompanied by the enumerators taking anthropometric measurements. Each household was therefore visited first by a male enumerator, followed on the next day by two female enumerators. In the pregnant woman sample, female enumerators established the first contact with the pregnant woman's husband, taking consent from the husband and the woman. The anthropometrics team visited the household the same day. Attention was paid to ensure that household members only learned about the outcomes of the anthropometric measures, especially blood test results, after the interview was completed. This ensured that information on health measurements did not affect the self-reporting of health outcomes.

Households were revisited on the same day if a respondent was momentarily unavailable or not at home at the time of visit. After the main survey round, a trailing round was conducted where households with respondents that were not present during the main round were followed up again. We did not pay any compensation for participating in the survey. When taking anthropometric measures of children, we gave them cookies. However, this was not meant as a compensation but to cheer up the children and encourage them to participate.

Qualitative data collection

A team member from IIT Gandhinagar with experience in the use of qualitative inquiry in the Indian context (Sini Varghese) collected the qualitative data of the study. She was guided in this exercise by Dr. Rosa Perez (qualitative methods) and Dr. Manisha Joshi (women's empowerment). To ensure responsiveness, interviews were carried out at a time convenient to the respondents, i.e. after 9 am (when they were free from their domestic work). The duration of interviews ranged from 45 min to two hours. The main qualitative approaches employed were in-depth interviews, FGDs, and village observations. These approaches were iteratively adapted to suit the local context under study.

At baseline, we interviewed 24 SHG members, three pregnant women, two men, three AWWs and three CMs from two treatment villages and one control village. Data were collected in Hindi as well as the local language. Data collection tools such as focus group topic guides and pre-designed questionnaires were developed in Hindi and the local language. The questionnaire and focus group topics were guided by previous studies on empowerment and agency of women (eg.: Nair et al. 2010; Rath et al. 2010; Kim et al. 2007; Bali Swain and Wallentin, 2009; Varkey et al. 2010; Kabeer, 2012; Sangeetha et al. 2013) and were also adapted from impact evaluation studies in similar settings (Rath et al. 2010; Morrison et al. 2010). The interviews and FGDs were conducted with the help of a translator. The interviews were audiotaped and later transcribed. Each interview lasted about 15 to 45 minutes. Only those respondents who provided informed consent were interviewed. Adequate care was taken to protect the privacy and confidentiality of the respondents. Drawing from previous experience of conducting FGDs in this area, FGDs were not audiotaped as we found that awareness of recording limited the interaction within the groups. Notes were taken during the FGDs and special care was taken to note down non-verbal cues such as gestures and body-language. Moreover, the investigator's observations of the meetings, health behaviours, and sanitation facilities in the villages were noted down.

The transcribed notes and observations were first translated to English. This was done with the help of an assistant who was fluent in the local language. These translated notes were again shared with another individual (who was fluent in Hindi and Tethi) who confirmed the cogency of the translated notes. The analysis of qualitative data was performed using a thematic framework approach for answering questions about the salient issues for particular groups of respondents (Green and Thorogood, 2014). The analysis involved the following steps:

(1) Listening to the recordings and re-reading field-notes and transcripts until the researcher was closely familiar with them.
(2) Thematic coding of data based on the objectives of the study and developing a systematic coding scheme.
(3) Indexing or applying the coding scheme to the data in textual form by noting the transcripts and observation notes.
(4) Charting, i.e., rearranging the data according to this thematic content, by themes identified.
(5) Mapping and interpretation: Exploring the relationships between defined themes and finding associations between them.

5.8 Data quality control measures during generation process

Several data quality control measures were in place during the data generating process. All enumerators participating in the data collection were split into small groups of up to ten persons. These groups were each headed by one supervisor, selected from the total group of enumerators. The supervisors' role was to organize the enumerators in the villages, ensure that interviews were completed in all sampled households, and accurately record the status of completed and not completed interviews. Additionally, a team of persons who were experts in administering specific questionnaires rotated among enumerators, accompanying them into the households. They noted mistakes or irregularities in the interviews and conduct with the respondents, which were then addressed either personally with individual enumerators or with the whole team of enumerators in case of common mistakes. At the early stage of the survey, the experts and the field manager held discussions every night on how to improve the quality of the survey.

Data checks were conducted on a regular basis on the newly generated data. This allowed detection of common inconsistencies, which were again addressed before the next field day. Particular focus was put on the correct tracking of individuals and households by randomly validating the collected information. Interviewer and supervisor codes allowed us to identify enumerators who performed badly. These mistakes were specifically addressed with the enumerators and they were intensively monitored during the interviews until their performance was at the same level as the other enumerators.

During the endline survey, data were recorded electronically using tablets. Enumerators were trained from the beginning in the use of the tablets. The programming ensured that questions were skipped only as required by the structure of the questionnaire, reducing data inconsistencies. Predetermined answer options for most questions removed potential typing errors during interviews. The automatic recording of interview start and end time as well as random time stamps allowed us to detect cheating. However, due to close monitoring by supervisors and the social stigma associated with cheating, we did not face this problem. The use of tablets also eliminated the need for data entry from paper questionnaires and thus removed that as a source of potential mistakes.

At midline and baseline, paper questionnaires were used. For the midline, a team of locals with technical expertise was hired for data entry. They were thoroughly instructed about data entry templates and the coding of questions. Frequent cross verifications of entered data were performed to avoid inconsistencies and errors. Templates were designed by researchers from IIT Gandhinagar and the University of Goettingen, and crosschecked by various team members. At baseline, the survey company Morsel conducted the data entry.

Data cleaning was carried out by researchers at the University of Goettingen. During the survey, duplicates of observations and the reasons for duplication were carefully noted and these were deleted during data cleaning. All variables were tabulated to check inconsistencies in the coding and values not corresponding to answer options were coded as missing. Note that this was not necessary for the electronic data collection at endline. Additional checks were done by crosstabulating variables that were related due to the sequence of questions in the questionnaire or the content. When merging data waves, questions and answer options in baseline, midline, and endline questionnaires were carefully compared and recoded in case of discrepancies.

6. Programme or policy: Design, methods and implementation

6.1 Programme content

The core of the Gram Varta intervention is a cycle of 20 pre-structured meetings, which were scheduled over a year. The meeting contents and outlines were mainly developed by Ekjut, a non-profit organization in Odisha, India. The meeting contents were not greatly modified for the Bihar intervention, according to many interviews conducted with officials implementing Gram Varta in Bihar. The meetings usually take about two to three hours for PLAs 1-15, and can stretch up to four hours per PLA for PLAs 16-20. For some meetings such as PLA 3 where the mid-arm circumference of children below five years are measured, the meetings can stretch longer due to non-members of SHGs likely bringing their children to the meetings. A table providing a brief description of each meeting is provided in the appendix. The meetings were expected to be held with a desired gap of a fortnight, and the PLAs from 16-20 were planned to be held with a gap of one week each.

A meeting usually begins with a welcome by the facilitator and with thanking those present for attending the meeting. The meetings are held in a discussion format and once the facilitator has explained the details of the meeting, the women or community members are encouraged to engage in discussions or ask questions regarding anything they have not understood. In some meetings, the attendees are invited to play games, others include stories and activities related to women's agency, attitude towards working together and service utilization, as well as HNWASH knowledge and practices. Participants are encouraged to think critically, identify problems in their households and communities, and discuss how HNWASH practices could be improved. Problems in their area regarding pregnant women, adolescent girls, and children under age six, are of special interest and participants belonging to these groups are encouraged to participate actively.

For some meetings such as the last five (PLA 16-20), the entire village community is requested to attend, including local authorities and service providers. In these meetings, the facilitator conducts a mapping of the village and also discusses, through examples and demonstrations, how habits such as open defecation are harmful to the health of the community. The identified problems and solutions are discussed and a community action plan is formulated. Community members attending the meetings are expected to set health and sanitation goals for the next few days or months. In later meetings, the community members are expected to report and discuss these goals and the progress made. Monitoring progress towards meeting pre-set goals is also a part of Gram Varta.

The meeting aids include flip charts, banners, posters, drawings, and simple stationery which can be used to elucidate any concept to the women and the community members. The venue of the meeting could be someone's courtyard, a public square, a panchayat office, a farm or any public place easily accessible (and visible) to the community.

6.2 Implementation

Gram Varta's PLA approach was implemented through village-based women SHGs affiliated with Jeevika in the case of Madhepura. This SHG based approach had already been implemented by WDC, Jeevika and Mahila Samakhya in several districts of Bihar before implementation started in Madhepura near the end of 2015.

The SHG structure of Jeevika is as follows: when eight Jeevika SHGs with at least 12 members each are formed among a couple of Gram Panchayats or a group of two to three villages, the next level of organization, the village organization (VO) is formed. A cluster level federation (CLF) is formed with around 15 VOs functioning in nearby villages and Gram Panchayats. The VOs of an area are allowed to form a CLF only after two years of regular meetings and no major gaps in their activities. The CLFs are at the sub-block level, thus each administrative block may have more than one CLF. Finally, the CLFs converge at the block level and form a block level federation, which then provides training for various livelihood activities, undertakes activities for SHGs to produce some goods which can then be sold or distributed for economic profit, maintains financial records, ensures quality of goods and disburses loans among the SHGs, among other activities.

SHGs which had matured and were considered ready for the intervention were chosen for Gram Varta implementation in certain areas. Notably, Jeevika chose to saturate entire clusters with SHGs and then implement Gram Varta in selected clusters. However, the objective criteria for beginning Gram Varta were not clearly defined.

Facilitators

In our earlier discussions with the implementing agency, the Jeevika team had shared that the intervention would be implemented by a cadre of Jeevika Sahelis, responsible only for Gram Varta implementation, and who were already expected to be present in the field by the time the intervention was rolled out. However, it was later decided that the Gram Varta meetings would be facilitated by a different cadre of staff called community mobilizers (CMs) who would need to be recruited. This presented a major change to the facilitator profile we anticipated as well as the time schedule. The CM was responsible for conducting the regular SHG meetings in addition to Gram Varta meetings. The recruitment of the CMs was either done through nomination among the SHGs or by inviting fresh applications which was published widely by Jeevika in their VOs and CLFs. The minimum qualifications were that the CM should be a woman, who should have been schooled at least until the 7th standard (grade) of school, should be able to perform basic calculations and writing, should be mobile, sympathetic, patient to women's queries, and should not hold any public office.

In Madhepura, the facilitators were usually trained at the block level by designated officials, such as the Gram Varta executive and the management information systems coordinator, and for some trainings special trainers were invited to convey the content. At some districts, including Madhepura, the training was held at the Jeevika training center in Purnea. Common training manuals and materials were used across districts and queries of the facilitators were addressed to enable them to conduct the meetings smoothly. Improvisation and innovations were not usually necessary for conducting the SHG meetings, since the content discussed and the materials used to discuss it were streamlined for all districts by the state level officials.

As per standard ethical guidelines, we informed the facilitators during our survey that the information they share would be used in a research project, although it is not known if they were told of the research aspect when the Gram Varta implementation began.

Programme beneficiaries and attendees

The programme target groups were low-income, resource-poor populations of rural Bihar, where the patriarchal system is still pronounced. The intended beneficiaries were women and children, with a special focus on pregnant women, lactating women, children under the age of five, and adolescent girls.

The Gram Varta intervention was announced to the community by the field staff of the implementation agencies after the facilitator recruitment for the programme was completed. The facilitator began publicizing the programme during the regular SHG meetings and encouraged the community members and women to participate in this programme. She emphasized that they would gain very useful information about health, nutrition, water, sanitation and hygiene. No compensation or tangible incentives were promised or given to any participants of the intervention.

The facilitator invited SHG members and the entire village population to participate in the Gram Varta meetings. Attendance for the meetings was voluntary, although the facilitator encouraged women to attend the meetings as much as possible. Attendance was not limited to SHG members only, and nor were any community members discouraged or turned away if they were listening to the meeting from nearby spots, or participating in the meeting directly. The date and time of the next meeting was decided based on convenience of the members during the earlier meeting itself; and as the date approached, the facilitator sent reminders to the women through other women or if possible, traveled to their houses herself. On the day of the meeting, she was encouraged to go door to door to persuade the women to join the meeting.

For most meetings, the intended participants were women since the setting of the intervention was the women's SHG. However, men and other community members were also encouraged to listen and participate in sessions. However, when meetings focused on topics such as contraception, family planning, sexual health of women; it was anticipated that women would be shy or unable to open up in the presence of men and community members. Such meetings were encouraged to be private and have only women in attendance. The entire community was invited to meetings related to community hygiene, toilets and community sanitation.

The ideal situation was that service providers such as the AWW, the Auxiliary nurse midwife and the ASHAs would attend the meetings and provide inputs regarding the government schemes, so that attendees were up-to-date regarding the information given. The community members were expected to seek services from these service providers—improving service uptake and delivery was one of the major intentions of Gram Varta.

Process indicators

On average, around 15-20 community members attended the meetings of Gram Varta in Madhepura according to the monthly progress report data. Monthly progress reports were

compiled by the implementing agency Jeevika for monitoring the number of meetings conducted per SHG, attendance at meetings, and details such as gender of attendees. These data also show that the meetings in Madhepura were primarily attended by women (men formed 10 percent or less of total attendees) and that there was decent participation by adolescent girls, pregnant and lactating women (about 10 percent to 35 percent of total attendees).

Table 6.1: Meeting participation and coverage

Indicator	Bihariganj	Gwalpara	Kumarkhand	Madhepura Sadar	Murliganj	Udakishunganj
Average total participation	6829.05	4537.55	17737.25	8991.05	7876.70	5671.65
Average participation of ANMs	0.00	1.25	1.35	0.00	1.90	9.00
Average participation of ASHAs	17.65	3.05	35.00	1.05	6.85	74.25
Average proportion of women	0.9234	0.9173	0.9165	0.9286	0.8746	0.9266
Average proportion of men	0.0766	0.0827	0.0835	0.0714	0.1254	0.0734
Average proportion of adolescent girls	0.0994	0.1921	0.1482	0.1090	0.1949	0.0987
Average proportion of lactating women	0.1434	0.2476	0.1122	0.1268	0.2303	0.1618
Average proportion of pregnant women	0.0489	0.0582	0.0430	0.0346	0.0661	0.0485
Average proportion of SHG members	0.6794	0.8218	0.5571	0.6762	0.8368	0.6709
Average proportion of ASHAs	0.0026	0.0007	0.0020	0.0001	0.0009	0.0131
Average proportion of ANMs	0.0000	0.0003	0.0001	0.0000	0.0002	0.0016
Average proportion of scheduled tribes	0.0130	0.0064	0.0150	0.0000	0.1706	0.0000
Average proportion of scheduled castes	0.4121	0.3897	0.4372	0.3544	0.2803	0.2724
Average proportion of other backward classes	0.5398	0.5728	0.5215	0.5903	1.0218	0.6263
Average proportion of general caste	0.0317	0.0296	0.0261	0.0271	0.0196	0.1009
Coverage	326.59	406.52	289.00	353.74	277.06	540.45
Number of SHGs	415	310	843	541	670	350

Note: These participation figures are based on monthly progress reports aggregated across all SHGs in one block. To calculate coverage, we used population figures from the Madhepura population, Census 2011 (Directorate of Census Operations, 2011).

Many monthly progress reports of the Madhepura Gram Varta meetings report no attendance of service providers at all, an extremely important indicator of the inclusiveness of the meetings, and a point which ties into the necessity of having departmental convergence between the stakeholders of Gram Varta. This was repeatedly shown to be lacking in our qualitative interviews in Madhepura as well as in six other districts where Gram Varta was implemented by Jeevika and WDC. In our qualitative surveys across Bihar, Gram Varta implementers reported a similar number of attendees, around 15-25 depending on the number of SHGs the facilitator was conducting the meeting with.

Table 6.1. presents process indicators, specifically participation of different population groups and coverage of SHGs, aggregated across all meetings and SHGs for each block. To calculate coverage, we used population figures from the population census in Madhepura district in 2011 (Directorate of Census Operations, 2011). Coverage varied across blocks, but the rates are similar to or even slightly better than the trials in Jharkhand and Odisha which reported one group per 468 population (Tripathy et al. 2010) and one group per 309 populaton in the second trial in Bangladesh (Fottrell et al. 2013).

Challenges in the implementation

During the course of the implementation, there were two major unforeseen changes. First, the change in the planned cadre of facilitators mentioned above. Second, the delay and changes in timing of PLAs (meetings).

The implementation of Gram Varta in Madhepura faced severe delays right from the start. The community mobilizers (CMs), the main facilitators for Gram Varta in Madhepura, were to be recruited before May 2015. However, that task was not completed in most blocks even until September 2015. An added complication that delayed hiring of CMs in several areas was that male CMs had begun facilitating Gram Varta meetings in those areas. They were later replaced by female CMs, although the recruitment process was ongoing even in December 2015. By December 2015, only three PLAs had been completed in most areas, which was a severe delay as per the original timeline. There was a long gap in implementation after November 2015 when there was a funding crunch at the local level in the Madhepura office of Jeevika. This funding crunch was solved in February 2016 and the gap between mettings for PLAs 4 to 8 was reduced, from fifteen days as originally intended to only seven days, to cover for the delay in implementation. In Gwalpara, as per the Gram Varta monthly progress reports, PLAs 9, 10, 11, 12, 13, and 14 were all held in the month of July. Similarly, six PLAs were reportedly held in July in Udakishanganj. The combination of long delays between meetings followed by several meetings held within a short period of time is highly likely to have caused poor retention of key messages and loss of enthusiasm among the Gram Varta participants. Notably, we have no quantitative data on how the meetings were conducted and are unable to gauge meeting quality. Members of the evaluation team had conversations with SHG members, CMs, and other community members during their field visits. They were given the impression that not all meetings were held in the manner expected and a few community members mentioned that no Gram Varta meetings at all were held in a few areas. Our team members also reported hearing that the refresher training for CMs (organized due to delays in implementation) was not attended by all CMs and the make-up refresher sessions were perceived to be inadequate by a few CMs.

Another critical aspect of the implementation, which was corroborated by our primary and secondary data, was the fact that men and elderly people, who are critical decision-makers in the community, did not participate in the meetings to the extent that was desired. Further, service providers such as the ANMs, ASHAs did not participate in the meetings in a regular manner. Hence they were not as alert about the Gram Varta process and the responsibility that it put on them to support community members, provide adequate, timely services, and work with the community to persuade higher officials to resolve any difficulties. According to the Department of Administrative Reforms and Public Grievances, there were some deficiencies in service provision during the rollout. The lack of inter-departmental cooperation

to ensure timely and adequate service provision is likely to have affected the Gram Varta process and the morale of the people in seeking such services in the future.

Additionally, in other districts such as Gaya, Madhubani, and Purnea, interviews with the district health and nutrition managers revealed that the PLAs 16-20 were conducted after meetings 1-3, since the rollout of Gram Varta in those areas coincided with the rollout of the Union Government's Swacch Bharat Abhiyan. This is likely to have had an impact since one of the major aims of the Swacch Bharat Abhiyan was to encourage people to build toilets, thus coinciding with a major outcome of Gram Varta. This gives us reason to believe that the comparability of Gram Varta implementation in Madhepura with other districts may be reduced. However, the Gram Varta manual, which covers the content and proceedings of the 20 meetings and was used in Madhepura, was the same as the one used in other districts of Bihar.

Tying into this discussion were the likely weak links which may have affected the effectiveness of the implementation and may have led to the lack of desired impact:

1. Two critical elements of Gram Varta, as mentioned above, were that important community members, including the powerful and decision-makers such as the men and the elderly, as well as relevant vulnerable members such as adolescent women, pregnant and lactating mothers, attend these meetings. However, as the Madhepura monthly progress reports show, there was low participation by men. Our qualitative data suggest that attending Gram Varta meetings was perceived to be an activity restricted to women SHG members.
2. It is unlikely that a diverse selection of community members participated in the meetings. Along with this, the frontline workers such as the ASHA and the ANM also did not participate in the meetings as much as desired. This participation was critical to ensure that community members increase their demand for service delivery and to spur improvements in service delivery which were critical to actually improve health indicators of the respondents.
3. Towards the end of the theory of change, the cycle mentions that frontline HNWASH workers are expected to be more aware, responsive, and supportive. This lacuna of a lack of convergence between all government departments and the frontline HNWASH workers being unaware or aloof from the Gram Varta process has likely affected these outcomes of the Gram Varta process.

7. Impact analysis and results of the key evaluation questions

7.1 Specification

General estimation approach

By virtue of random assignment of the intervention, a simple comparison of post-intervention outcomes between treated and non-treated households will on average provide us with an unbiased intention-to-treat effect. Therefore, our first specification took the following form:

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

where Y_{iv} is the outcome for unit i (e.g. pregnant woman/household head) in village v measured at endline. The coefficient β presents the intention-to-treat effect. Note that this is a simple comparison of post intervention means.

However, unbiasedness only ensures that random selection errors induced by chance average out when randomized controlled trials are conducted repeatedly. In contrast, for any single trial, it is possible that random imbalances between the treatment and control group regarding observable or unobservable characteristics distort the estimates. Therefore, our second specification controlled for household and individual background characteristics:

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

where X_{iv} is a vector of control variables. While the baseline report[4] overall did not suggest that large imbalances exists between the treatment and control group, small deviations between groups may lead to considerable distortions if the prognostic strength of the respective covariates is high. Therefore, in the household sample we controlled for household size, the highest education achieved by any member of the household; whether the household owns land, livestock or durable household goods; and the age of the woman respondent, as we suspected that demographic and household characteristics are strong predictors of the outcomes under study. Similarly, to capture potential confounders with a high relevance for female empowerment and HNWASH indicators, in the pregnant woman sample, we controlled for pregnant woman specific variables – age, educational attainment, whether she can read an SMS or short sentence, whether she works on the farm or in the household, and age when first child was born – as well as household level characteristics – religion, caste, having a below poverty line card, an asset index[5], ownership of land and ownership of livestock.

The majority of indicators were binary variables. While linear probability models may provide a good approximation to the average marginal effects of Gram Varta, models relying on non-linear link functions are typically more efficient for modelling probability. We therefore modeled the probability that a binary indicator equals one as a logistic function and calculated average marginal effects. In some cases, indicators were coded on an ordinal scale, requiring the use of ordered logit models which do not impose a cardinal interpretation of estimated effects. For binary outcome variables, the estimated average marginal effect can be interpreted as the percentage point change in the probability that the indicator of interest holds true. Similarly, for ordinal outcome variables, marginal effects reflect the percentage point change in the probability that the respective outcome category is true.

While the pre-analysis plan specified that we will interpret effects in terms of odds ratios, we decided to report average marginal effects (AME) for two reasons: First, AME have, similar to linear estimates, a straight-forward interpretation in terms of absolute effects on the probability that the outcome equals one. Second, odds ratios are scaled by the unobserved heterogeneity not accounted for by the respective regression model such that comparisons across different models or sub-samples are generally not adequate (Allison 1999). In contrast, AME are unaffected by this problem (Mood, 2010).

[4] The baseline report refers to an internal document cited under Bommer, Subramanyam, and Vollmer, 2015b.
[5] The asset index places equal weights on having electricity, a radio, a television set, a mobile phone, a landline phone, a refrigerator, a watch or clock, a motor cycle, an animal drawn cart, a car or a truck.

For further analysis we exploited the panel-structure of our study and employed a difference-in-differences approach in order to control for baseline differences and a common time trend:

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

where Y_{ivt} is the outcome for unit *i* (e.g. pregnant woman/household head) 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 is a dummy variable that equals 1 if the data comes from 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.

In the household sample we used individual fixed effects for the main woman respondent. For outcomes regarding the last born child we defined the panel through the mother instead of the child for three reasons. First, because the last born child at baseline may not be the same at endline and questions about the last born child differ from those about other children under the age of five, simply looking at the last born child of the same mother avoids these issues. Second, it allowed us to increase the number of observations included in the analysis as opposed to only looking at children who are under the age of five in all waves. Third, it is rather the mother than the child characteristics and behaviour that influence our outcomes. For outcomes regarding adolescent girls, we did not estimate the difference-in-differences specification. This is because adolescent girls were difficult to track from baseline to endline and we observed substantial errors in tracking during data collection. For the pregnant women sample, we used individual fixed effects for the woman and her husband.

We estimated the difference-in-differences specification with ordinary least squares even for binary and ordinal outcomes. This is because the standard difference-in-differences framework does not apply to non-linear estimators and alternative approaches, such as the changes-in-changes model, are not able to provide point estimates without additional assumptions (Athey and Imbens, 2006). A linear probability model estimated by OLS is hence likely to produce a better approximation of causal effects.

In order to reflect the spatial segregation of households due to our clustered sampling design, all standard errors are clustered on the village-level.[6] While the pre-analysis plan indicated that we would additionally cluster on the individual or household level in all difference-in-differences specification, estimations turned out to be too computational intensive to conduct them for all outcomes. However, we found standard errors to be nearly identical whenever these additional cluster levels were introduced, suggesting that this is unlikely to affect our results.[7]

In the main analysis we made use of the whole samples of households and pregnant women. For further analysis, we created a sub-sample of Gram Varta exposed households or individuals. As explained in Section 5, the fact that participation in SHG meetings is voluntary

[6] As alternative strategy in would have been possible to cluster on the Gram Panchayat level. We decided against this approach as the geographical distance between villages of the same Panchayat was large enough to justify the assumption that the correlation of unobserved heterogeneity between these villages would be small. Clustering on the Gram Panachayat level would therefore have induced an unnecessary loss in efficiency.

[7] While efficiency gains might be realized by using a random effects (RE) model instead, the RE approach assumes that the unobserved village-level heterogeneity is uncorrelated to all of our right-hand side variables which is unlikely to be true given our sampling design. Clustered standard errors, while being inefficient, do not require such assumptions on the correlation structure.

increases the likelihood of differences in the treatment intensity. A household or individual was considered directly exposed to Gram Varta if the woman respondent indicated to be a Jeevika SHG member and attends meetings mostly or sometimes (opposed to never or rarely). This resulted in a smaller sample of 1,079 panel observations. We then compared exposed women or household members in the treatment and control group using the estimation strategies outlined above. These estimations can be interpreted as treatment-on-the-treated effects. Because selection into SHG meeting intensity might be higher in the treatment group over the course of the study due to Gram Varta we tested for differential trends in membership between treatment and control group since midline, which is the first time we observed the frequency of SHG meeting attendance. We found no evidence for a differential trend in the probability of “sometimes” or “mostly” attending Jeevika SHG meetings. Please note that this estimation strategy makes two assumptions. First, that the trend in SHG intensity does not differ between treatment and control group between baseline and midline. We argue that this holds true because only a few Gram Varta meetings had been conducted at the time of the midline and, if anything, selection into Jeevika SHGs would have been larger for the treatment group between midline and endline, which we do not observe. Second, we had to assume that this finding of non-differential trend could be applied to the sample of pregnant women; which is credible considering that the household and pregnant women sample cover the same villages. In the results section we present comparisons of the findings from the main analysis with the findings from the exposed sub-sample. However we have included the regression tables of the exposed sub-sample in the appendix T for the sake of brevity. We refer to women that frequently attend Jeevika SHG meetings as ***active SHG members***.

For an analysis of heterogeneous treatment effects, we split the full samples into subgroups according to four variables and used the model specifications as described above for each subgroup. The first variable was caste, where we differentiated between households whose head’s caste was (1) a Scheduled Caste, (2) a Scheduled tribe, (3) an Other Backward Class, and (4) a general category. The second variable was age, where we formed two subgroups such that their size was as equal as possible. In the household sample, this resulted in a group of women aged 18 to 35, and a second group aged 36 and above. In the pregnant women sample, this resulted in a group of women aged 15 to 22, and a second group aged 23 to 45. The third variable for subgrouping was education of the woman, divided into three groups. In the household sample the groups were (1) no education completed, (2) primary or middle school (8^{th} grade) completed, and (3) secondary completed or higher. In the pregnant women sample, the groups were (1) no education completed, (2) primary school completed and (3) middle school completed or higher. Lastly, the fourth variable was the administrative block of residence: (1) Bihariganj, (2) Gwalpara, (3), Kumarkhand, (4) Murliganj, (5) Madhepura Sadar, and (6) Uda Kishanganj. We refer to results from the subgroup analysis in the text in the instances where we found robust and significant results across specifications and indicators supporting or opposing our analysis. Subgroup analysis regression tables are available upon request.

The following sections describe the results of our investigation of the impact of Gram Varta as per the hypotheses outlined in the pre-analysis plan and in Section 2. As indicated in Section 1, indicators that did not work well in the field or were directly related to the current pregnancy of the woman in the pregnant women sample are not part of the analysis. Please see tables A.1 and A.2 for a list of indicators and the reasons for dropping them. Results for hypotheses 6, 7, 15, 41, and 48 are not included in this report.

Note that the vast majority of indicators considered in this section are self-reported. This predominant reliance of self-reporting is both a strength and a weakness of this evaluation. On the one hand we have the opportunity to gain important insights into the lived experiences of respondents with Gram Varta. On the other hand, self-reported indicators, particularly those related to health, may be confounded by pre-conceived opinions and expectations of subjects. The objective measures for hygiene or health, including anthropometic measurement and observations of enumerators, therefore provide important additional insights allowing us to analyse both subjective and objective dimensions of the impact of Gram Varta.

Finally, please bear in mind that in contrast to the analysis described above, results from the Anganwadi worker survey and the facilitator survey are based on a baseline-endline comparison only, as these modules were not administereded in the control areas. Consequently, readers should treat results based on these surveys as complementary to the causal evidence generated by the main parts of the evaluation.

Sample size variations

Notably, sample sizes varied substantially across outcomes and specifications. Specification 2 in general included a smaller number of observations than specification 1 because it included a vector of covariates. These control variables were not available for all observations. They were not available if the respondent did not answer the relevant question or if that question was answered with ‘don’t know’ and coded as missing value. Specification 3, the difference-in-differences model, contained more observations than the other specifications because it used both baseline and endline data.

Moreover, the household sample was larger than the pregnant women sample. Within the household sample, more interviews were completed with household heads than with the women respondents. Reasons for this are discussed elsewhere in this report and include the change in selection criteria for the woman respondent. The sample of adolescent girls, last born children, and other children were also smaller than the full sample. This is because not every household had a female member in the appropriate age range for the adolescent girl interview, and not every woman respondent had a child under the age of five years at baseline or other children in this age range. The sample size varied across indicators depending on the dataset (household or pregnant woman) from which the indicator was taken.

Lastly, sample sizes vary across indicators within one specification and within one survey sample. This is because not all respondents answered all questions. Some questions were not asked in the first place, because they were not applicable to a household. For example, if a household did not own a toilet, the respondent was not asked whether any household members practiced open defecation despite the presence of a toilet. Other questions were not answered because the respondent refused to answer. For instance, a few adolescent girls refused to answer sensitive questions related to their knowledge about sexuality and contraception. Other questions were answered with ‘don’t know’ if the respondent did not have the information, for example the number of visits to the Anganwadi centre in the past months. Both non-response and “don’t know” were coded as missing values.

Multiple testing adjustment

One potential concern with the present analysis is that the large number of tested hypotheses may lead to an accumulation of type-1 errors such that a number of hypotheses could be rejected by chance rather than due to the existence of real effects. We therefore employed a correction for multiple hypotheses testing to guard against this problem. Please note, however, that traditional error correction approaches controlling the family wise error rate, such as the Bonferroni correction, are likely to be too conservative in most settings. Instead we made use of the Benjamini-Hochberg method which aims to control the false discovery rate (see Fink et al. (2014) for a discussion). The method of Benjamini and Hochberg is simple to implement: First, we ranked in ascending order all p-values of a set of hypotheses. Then, every hypothesis for which the following inequality held was rejected:

$$p(i) \leq \frac{i}{m}\alpha$$

where $p(i)$ is the p-value with rank *i*, *m* is the number of hypotheses in the multiple testing set and α is the target type-1 error probability (e.g. 5 percent).

We considered hypotheses to be part of one set if they met the following two conditions: (a) they referred to the same survey and (b) they referred to the same model. Condition (a) was imposed as separate surveys (i.e. household survey and pregnant women survey) were subject to different sampling processes. Condition (b) was chosen as the hypotheses tests for the same indicator across different models (no controls, controls and difference-in-differences) were highly correlated, making multiple testing corrections too conservative. Further please note that results from the Anganwadi and CM surveys as well as all sub sample analyses were exempt from the multiple testing procedure as the intention behind these analyses was to augment the main results and the inclusion of auxiliary hypotheses should not weaken the confidence in the main hypotheses. To highlight significant effects which were robust to the multiple testing correction in regression tables, we added significance stars (* = 10 percent and ** = 5 percent) to p-values which remained significant after applying the Benjamini-Hochberg approach.

7.2 Results related to women's self-help groups

The first group of hypotheses investigates the effect of Gram Varta on acceptance and awareness of women's self-help groups (SHG) as well as utilization of government health services through SHGs. As pointed out before, these results are based on the community mobilizer (CM) survey and are, while being helpful to understand the general context of Gram Varta, not necessarily indicative of causal effects due to the lack of a control group.

Hypothesis 1

We first investigated the hypothesis that Gram Varta improves participation in SHG meetings. This relates to one of the assumptions in the ToC that "All target groups learn about and are willing and able to attend the meetings" (Stage 2). For this, we computed the difference in the reported SHG attendance during endline (January 2017) and attendance reported by the same CMs at baseline (October 2015). On average, attendance decreased by 55 members (IQR: -93 to -10). We are therefore unable to find evidence to support Hypothesis 1.

Hypothesis 2

We futher investigated the hypothesis that Gram Varta increased SHG acceptance in and cooperation with community.

Table 7.1: Distribution of responses about SHG member cooperation

	Endline				
Baseline	Lots of cooperation	Some cooperation	Not much cooperation	No cooperation	**Total**
Lots of cooperation	71	11	1	1	84
Some cooperation	69	15	4	0	88
Not much cooperation	0	2	0	0	2
No cooperation	0	4	0	0	4
Total	140	32	5	1	178

Table 7.1 shows the distribution of responses to the question “How much cooperation was there between SHG members and the village community when it comes to solving problems (e.g. regarding health, infrastructure, poverty) together?” The rows represent the responses of 178 CMs during the baseline while the columns show the responses of the same CMs during the endline. As can be seen 69 of the 88 who reported “some cooperation” at baseline moved to the “lots of cooperation” column by the endline. This supports our hypothesis that Gram Varta increases the cooperation between the SHG and the community. This ties in with the ToC assumption that the community comes together in solidarity (Stage 4). Please note that these are self-reports by the CMs, who are facilitators of Gram Varta.

Hypothesis 3

Based on the CM survey, this section investigates the hypothesis that Gram Varta improves provision of information on health practices within SHGs.

Table 7.2: Frequency of discussions about water and sanitation related topics

	Endline			
Baseline	Every meeting	Often	Not very often	**Total**
Every meeting	72	22	4	98
Often	18	3	1	22
Not very often	48	10	1	59
Total	138	35	6	179

Table 7.2 summarizes responses to the question “How often are water and sanitation related topics discussed in Gram Varta meetings?”. Of the 98 CMs who said “every meeting” at baseline, 72 continued with the same answer at endline, 22 said “often,” and 4 said “not very often.” Among the 22 who said “often” at baseline, 18 moved up to “every meeting,” while 48 moved to “every meeting” out of the 59 “not very often.” When asked “Is training SHG members on proper water usage and sanitation measures in the household really necessary?” 178 out of 182 CMs answered “yes” at baseline while 100 percent of them said “yes” at endline. Tables 7.3 to 7.6 show the distribution of responses from the same group of CMs, at baseline versus endline, to questions related to provision of health information.

Table 7.3: Frequency of discussions about neonatal & infant care practices

	Endline			
Baseline	Every meeting	Often	Not very often	**Total**
Every meeting	76	22	4	102
Often	28	3	0	31
Not very often	28	14	3	45
Total	132	39	7	178

Table 7.4: Frequency of discussions about vaccination in pregnancy and early childhood

	Endline			
Baseline	Every meeting	Often	Not very often	**Total**
Every meeting	57	23	5	85
Often	21	5	1	27
Not very often	41	19	5	65
Total	119	47	11	177

Interestingly, questions on neonatal and infant care practices, vaccination in pregnancy and early childhood, nutrition and contraction followed a similar pattern where responses from CMs in the endline tended to move to “every meeting” or “often” from other categories.

Table 7.5: Frequency of discussions about nutrition related topics

	Endline			
Baseline	Every meeting	Often	Not very often	**Total**
Every meeting	70	17	5	92
Often	22	4	0	26
Not very often	40	15	4	59
Total	132	36	9	177

Table 7.6: Frequency of discussions about contraception

	Endline			
Baseline	Every meeting	Often	Not very often	**Total**
Every meeting	42	12	9	63
Often	10	6	1	17
Not very often	44	18	12	74
Total	96	36	22	154

Taken together, the presented evidence supports our hypothesis that Gram Varta improves provision of information on health practices within SHGs. Again, please bear in mind that these are self-reports from Gram Varta facilitators. This conclusion ties in with the ToC assumption that awareness is raised on the importance of relevant practices for health and nutrition of the whole family (Stage 2).

Qualitative evidence also supports the idea that provision of health-related information improved with Gram Varta. Interviews with the facilitators suggested that the phased manner of the Gram Varta meeting cycle and the structured content within these meetings ensured that there were focussed discussions on health issues within the community. Conversations with SHG members suggested that HNWASH related messages were given to the respondents through the use of picture cards, stories, demonstrations and games.

Interestingly, the respondents seemed to remember demonstrations, games and picture cards rather than stories. The following quotes from the respondents illustrate this:

"*I remember the games such as voting for diseases and being shown pictures to identify health problems in small children" (SHG member, married, 33 years, treatment village 3).*

"I think that showing the effects of open defecation through pictures worked better than telling people stories as they were able to see and understand things. Then, these messages will be engraved in their minds." (Facilitator, treatment village 1)

"These games helped us to mingle with each other and have fun. We laughed so much during the power walk and the donkey games. CM didi and other people also explained to us the messages behind these games." (Adolescent girl, treatment village)

Hypothesis 4

We also investigated whether Gram Varta improved provision of health finance information and practices within SHGs. Specifically, we asked whether CMs had discussed the Health Risk Fund with SHG members during training. A majority of CMs (95.6 percent at baseline versus 98.9 percent at endline (N=181)) reported that they had indeed discussed it (McNemar's chi square p value>0.05). However, the trend in practices suggests that the use of health risk fund did not pick up in Madhepura. We analyzed the response to the question "How many members of your SHG/SHGs keep aside money as health savings?" using paired t tests and found that the average number of members per CM, who kept aside money as health savings, decreased from about 94 (SD=3.65) at baseline to about 71 (SD=5.14) at endline ($p<0.001$).

This is corroborated by our findings related to the number of community members enrolled in health risk funds and the number of deposits made into health risk funds. In response to the question, "How many people in your community are currently enrolled in a health risk fund?" the average number of community members who were enrolled in health risk fund decreased from 91 at baseline to about 56 at endline (paired t test p value<0.001). The number of deposits made the previous month into health risk funds also decreased, on average, from the baseline value of about 76 to the endline value of about 64 (paired t test p value<0.05). Interestingly, CMs reported a decrease in the number of people who needed to access their health risk fund due to illness from about 6 at baseline to about 4 at endline.
Taken together, while we find some indicative evidence that the provision of heath finance information improved, we do not find an increase in utilization, thus we cannot confirm our hypothesis.

Hypothesis 5

Based on the CM survey, we further investigated the hypothesis that Gram Varta improved awareness and usage of government health services by SHG members. The average number of SHG women who had used government health facilities in the previous month decreased from 20.66 at baseline to 13.58 at endline (paired t test p value=0.005, N=184) based on CM reporting. While different seasons may explain some of this difference, the data do not support the hypothesis. At baseline, 180 of 229 CMs (78.6 percent) reported that members of their community most frequently visited government health facilities for treatment when they have a serious injury or illness. This proportion reduced at endline to 162 of 237 CMs (68.4 percent).

On the other hand, the proportion of CMs who reported that SHG member visited government health facilities for minor health issues increased from 54.2 percent at baseline 68.4 percent (N=237) at endline. When we asked about their opinion: “Do you think the community makes full use of government health services?” of the 166 who had answered “yes” during baseline 162 again said “yes” during the endline; notably, of the 16 “no” at baseline, 15 turned to “yes” at endline (McNemar’s chi square p value=0.02). We saw a similar pattern when the same question was asked about Anganwadi centers (of the 24 “no” at baseline, 21 turned into “yes” at endline, McNemar’s chi square p value=0.01).

This hypothesis is related to the ToC assumptions that awareness is raised on the importance of relevant practices for health and nutrition of the whole family, including demand for health services, and that participants are determined to find solutions to self-identified problems (Stage 4). Overall, we do not find sufficient evidence to support our hypothesis and reject that Gram Varta improves awareness and usage of government health services by SHG members.

Hypothesis 8

Based on the CM survey, this section investigates whether Gram Varta improved facilitators' health knowledge. When we asked the CMs, “How many times should a pregnant woman go to antenatal check-up?” the average number answered at baseline was 3.84 and at endline it was 3.45. Almost 100 pecent of the CMs answered in the affirmative at both baseline and endline when asked, “Do you think it necessary that pregnant women take iron and folic acid tablets during their pregnancy?” This was also true for a few other questions such as “Do you believe open defecation or open sewage water represents a health hazard to you, your family or people in the community?” and “Do you think a good balance of foods will protect families from diseases and malnutrition?”.

All CMs who had said “no” at baseline to the question whether condoms are an effective way of protecting against sexually transmitted diseases like HIV/AIDS, said yes by endline (36 out of 157, McNemar’s chi square p value=0.00). This pattern was also observed among those who had said “no” at baseline to the question, “Is holding the baby in contact with mother’s warm skin a good practice?” (15 out of 157, McNemar’s chi square p value=0.002).

While 110 of 232 CMs at baseline did not believe that choosing to breastfeed babies instead of giving them water to drink can protect them from some diseases, 96 of them indicated that they believed this to be true in the endline. Moreover, while 19 of 232 CMs who responded at baseline that feeding colostrum was “not” or “somewhat” important responded with “very important” at endline, 6 CMS had regressed from “very important” at baseline to lower categories at endline. Lastly, the majority of CMs who had responded with “not important” (87.5 percent of them) and 91.7 percent of those who had responded with “somewhat important” at baseline, felt that getting a tetanus injection before giving birth was “very important” at endline.

Taken together, based on the presented results we can confirm the hypothesis that Gram Varta improves facilitators' health knowledge.

Hypothesis 9

Lastly, based on the CM survey, we investigated whether Gram Varta improved facilitators' opinion on their work. When asked if she was satisfied with her job as a community mobilizer,

5 of 183 CMs who had answered no at baseline responded affirmatively at the endline. All 183 CMs, who responded to this question, reported feeling respected by the SHG members and the community, at both baseline and endline. A similar pattern was seen in response to the question, “Do you feel that your work is appreciated by SHG members?” Only 17 of 182 had said yes at baseline, when asked if they had experienced something unpleasant in the community while doing their jobs. All of them said no to this question at the endline (McNamara’s chi square p value=0.02). Corroborating our finding regarding lack of convergence with other services and departments, the proportion of CMs who said yes to the question, “Do the other decision makers in the village like ASHA, ANM, Sarpanch, Gram Sevak seek your opinions during decision-making processes?” changed from about 50 percent at baseline to about 65 percent at endline (p value=0.0013). However, when asked, “Do people listen to you and seek your opinion during Gram Sabha meetings?” the proportion of CMs saying yes changed from 86.9 percent at baseline to 94.3 percent at endline (p value=0.024).

Taken together, these results suggest that while the CM had a very high opinion of her job even before Gram Varta, the programme has likely increased this further, thus supporting our hypothesis.

Summary

The results presented above have shown that while Gram Varta was not associated with an increase in the participation in SHG groups, there is some evidence that cooperation between SHG groups and the community improved. Moreover, CMs reported more discussions on important health and nutrition-related topics and generally exhibited better health knowledge. Nevertheless, this did not result in increased uptake of government health services by SHG members. Similarly, despite increased discussion on health finance, the CM survey did not find any increase in the utilization of financial services.

7.3 Results related to women’s agency and empowerment

An increase in the level of women’s empowerment and their sense of agency was hypothesized to be an important channel through which Gram Varta makes an impact on HNWASH related attitudes and practices. For instance, important assumptions made in Stages 4 and 5 of the ToC are that women feel empowered and strong enough to foster change in practices and behaviour in their own households; and that women feel empowered to carry out the learned behaviours together with or even against the will of their husbands or other relatives.

This section of results tests the hypotheses that Gram Varta increases the level of women’s empowerment and sense of agency. It presents statistical models that takes advantage of the randomized design to compare women living treatment and control areas and thus allows a causal interpretation of the results.

We use data from both the household survey and the pregnant women’s survey in section 7.3. The next few sub-sections discuss the statistical results in detail. A summary of these results is given after presenting results related to Hypothesis 26.

Hypothesis 10

Based on data from the household sample, this section describes the result of testing the hypothesis that Gram Varta encouraged women to acquire paid work and to become economically more independent.

Table 7.7: Paid work and economic independence, household sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Woman earns agricultural income	0.0197	0.314	2996	0.0245	0.134	2992	0.0293	0.189	4954
	(0.0195)			(0.0164)			(0.0222)		
Woman earns non-agricultural income	0.0001	0.987	2996	-0.0027	0.697	2992	0.006	0.554	4954
	(0.0071)			(0.007)			(0.0102)		
Woman has any savings	-0.0203	0.444	1934	-0.0244	0.346	1932	0.0774	0.036	4153
	(0.0265)			(0.0259)			(0.0366)		
Woman's level of savings	-248.8949	0.625	397	-288.7596	0.564	397	-1268.250	0.198	806
	(507.4414)			(499.7627)			(981.3683)		
Woman has any loan (business)	0.0100	0.547	3117	0.0137	0.405	3113	0.0382	0.099	4960
	(0.0167)			(0.0165)			(0.023)		
Husband trust woman in money matters	-0.0133	0.513	2451	-0.0109	0.596	2449	-0.0422	0.253	4380
	(0.0203)			(0.0206)			(0.0368)		
Hours unpaid housework (adolescent girl)	1.2749	0.162	666	1.2209	0.179	666			
	(0.9072)			(0.9048)					

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

Table 7.7 shows that none of the effects on economic empowerment indicators were significant when considering endline means only. This did not change after adding control variables. According to the difference-in-differences specification, one effect was significant at the 5 percent level, although not robust to multiple testing correction. Treatment increased the probability of a woman having any savings by 7.8 percentage points. The sign of the coefficients in specifications 1 and 2, however, showed a decrease. The effect on the level on savings was negative in all three specifications, although not significant. In line with the hypothesis, the effect on having own agricultural income was positive in all specifications although insignificant. With a significance level of 10 percent in the difference-in-differences specification, the probability of having any loan increases, by about 4 percentage points, and this increase was consistent across specifications. Notably, the significance disappeared once the Benjamini-Hochberg correction is applied. In contrast to the hypothesis, we found a

positive, albeit non-significant, effect on the hours of unpaid housework done by the adolescent girl.

This picture changed when we included only self-declared active SHG members in the analysis. For this subsample we found a consistently positive and significant effect on whether the woman has any own non-agricultural income and a significant increase in having agricultural income in specifications 2 and 3. While the probability of having any savings significantly increased by almost 20 percentage points according to specification 3, the level of savings significantly decreased in all specifications as found in the full sample.

Among the women who were active SHG members, Gram Varta increased the chance of the woman having any own non-agricultural income and increased the level of her agricultural income as well. Some suggestion of a similar effect was seen when all women in the household sample were studied, but results were not statistically significant. No evidence was found to support the hypothesis that Gram Varta improved other indicators of paid work and economic independence.

Overall, we do not find convincing support (consistent effects across multiple indicators and models) for our hypothesis that Gram Varta encourages women to acquire paid work and to become economically more independent.

Hypothesis 11

This section reports on the hypothesis that Gram Varta increases women's bargaining power within the household using data from the household and pregnant women samples. In the household sample, the effects on women's participation in deciding about her own income were positive, as expected, and consistent across specifications (table 7.8). However, none of the effects were significant in Model 1. Including baseline controls did not change much in terms of coefficient signs or standard errors. In the difference-in-differences specification, two effects were significant at the 10 percent level. Treatment increased the probability that the woman had any say in the use of agricultural income earned by herself. It also increased the probability that the woman participates in the decision about visits to relatives by 5 percentage points. Note, however, that significance disappeared once multiple testing was accounted for.

Table 7.8: Bargaining power within the household, household sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Say in use of agricultural income...									
Not allowed	-0.0133	0.376	828	-0.0136	0.373	826	0.2917	0.073	1379
	(0.015)			(0.0152)			(0.1620)		
Decide on less than half	-0.0175	0.370	828	-0.0177	0.367	826			
	(0.0196)			(0.0197)					

Decide on more than half	-0.003	0.366	828	-0.003	0.362	826			
	(0.0033)			(0.0033)					
Decide on total amount	0.0338	0.370	828	0.0343	0.367	826			
	(0.0377)			(0.0381)					
Say in use of non-agricultural income									
Not allowed	-0.0356	0.153	108	-0.0289	0.287	107	-		
	(0.0249)			(0.0271)					
Decide on less than half	-0.0657	0.202	108	-0.0507	0.340	107			
	(0.0514)			(0.0531)					
Decide on more than half	-0.0149	0.229	108	-0.0108	0.360	107			
	(0.0124)			(0.0118)					
Decide on total amount	0.1161	0.167	108	0.0904	0.311	107			
	(0.0841)			(0.0893)					
Say in use of savings									
Not allowed	0.0032	0.881	990	-0.0024	0.913	989	0.1393	0.636	1347
	(0.0212)			(0.0215)			(0.2940)		
Decide on less than half	0.0019	0.881	990	-0.0014	0.913	989			
	(0.0125)			(0.0125)					
Decide on more than half	0.0004	0.881	990	-0.0003	0.913	989			
	(0.0025)			(0.0025)					
Decide on total amount	-0.0054	0.881	990	0.004	0.913	989			
	(0.0362)			(0.0365)					
Own decision on savings	-0.0131	0.718	1016	-0.0021	0.954	1015	-0.0063	0.952	1366
	(0.0363)			(0.0366)			(0.1060)		
Participate in decisions regarding									
Health care	-0.0025	0.895	3127	-0.0019	0.920	3123	0.0357	0.197	4977
	(0.0190)			(0.0191)			(0.0276)		
Household purchases	-0.0031	0.881	3066	-0.0007	0.973	3062	0.0348	0.255	4939
	(0.0211)			(0.0214)			(0.0304)		
Visits to relatives	0.0008	0.962	3135	0.0015	0.935	3131	0.0498	0.079	4990
	(0.0180)			(0.0180)			(0.0282)		
Farm matters	-0.0338	0.192	2591	-0.0323	0.209	2589	0.0040	0.920	4517
	(0.0259)			(0.0258)			(0.0399)		
Use of own earnings	0.002	0.906	2043	0.0043	0.801	2040	0.0382	0.212	4134
	(0.0169)			(0.0170)			(0.0305)		
Use of husband's earnings	-0.0105	0.588	3063	-0.0076	0.700	3059	0.0248	0.434	4918
	(0.0193)			(0.0198)			(0.0316)		
Land/house purchase and use	-0.0052	0.823	2632	-0.0022	0.926	2628	-0.0145	0.729	4488
	(0.0231)			(0.0234)			(0.0418)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

When we looked at the subsample of active SHG members, no effect was found to be significant at any conventional level, although coefficients pointed in the expected direction for the first three indicators. Participation in decision making seemed negatively affected for decisions about health care, farm matters, and the use of husband's earnings. Looking at the indicators for bargaining power overall, including participation in decision making, Gram Varta does not seem to strengthen the women's position visibly in the household sample.

Table 7.9: Bargaining power within the household, pregnant women sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Makes decisions alone or jointly with husband about…									
Health care	0.0056	0.879	1605	0.0186	0.626	1531	0.0366	0.453	3595
	(0.0366)			(0.0382)			(0.0486)		
Household purchases	-0.0152	0.705	1607	-0.0085	0.838	1533	-0.0503	0.302	3598
	(0.0402)			(0.0414)			(0.0486)		
Visits to relatives	0.0120	0.751	1605	0.0215	0.593	1531	0.0318	0.505	3588
	(0.0379)			(0.0401)			(0.0476)		
Farm matters	0.0096	0.827	1352	0.0118	0.787	1287	-0.0519	0.365	2059
	(0.0437)			(0.0436)			(0.0571)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

In the pregnant woman sample, we did not find a consistent effect of Gram Varta on the woman's bargaining power within the household across specifications and indicators (Table 7.9). Treatment had a positive yet insignificant effect on making decisions about health care and visiting relatives or friends alone or jointly with the husband. The effect on making decisions about household purchases alone or jointly with the husband was negative but insignificant in all three specifications. The sign of the estimate regarding making decisions about farm matters changed from positive to negative when estimated in the difference-in-differences framework. Based on this subsample analysis of active SHG members we conclude that there is no effect of Gram Varta on the woman's bargaining power within the household in the pregnant women sample.

In the subgroup analysis by caste, we found a robust decrease in the probability of being involved in decision making for people of scheduled tribes. This was significant for decisions about health care and household purchases in the difference-in-differences specification and for decisions about farm matters in specification 2. Moreover, in the subgroup analysis by

education, we found a robust positive impact of Gram Varta on involvement in decision making for women who have completed primary schooling. However, the effects were only significant for decisions about visits to family in specification 3 and decisisons about farm matters in specifications 1 and 2. When we looked at women who have completed junior secondary education or higher we found convincing evidence that Gram Varta had significantly reduced involvement in decisions. The probability of participating in decision-making dropped by 12 to 16.5 percentage points for women in this education group.

Based on these results, we were unable to confirm the hypothesis that Gram Varta increases women's bargaining power within the household. In the case of Scheduled Tribe members we even found that Gram Varta reduced bargaining power in the household.

Hypothesis 12

Results of testing the hypothesis that Gram Varta enabled women to become more independent of their husbands is described in this section. The data used were from the household and pregnant women samples. In the household sample, none of the effects were significant in specifications 1 and 3 and most effects were negative i.e., ran opposite to our hypothesis (table 7.10). Including baseline controls did not really change the coefficients or standard errors. This was confirmed when looking at the subsample of active SHG members where no significant effect was found for any indicator. However, among active SHG members treatment effects had the expected positive sign for all indicators.

Table 7.10: Independence from husband, household sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Allowed to go									
Alone to market	-0.0041	0.822	3145	-0.0065	0.717	3141	-0.0061	0.838	4996
	(0.0181)			(0.0179)			(0.0299)		
Alone to health facility	-0.0011	0.952	3144	-0.0031	0.858	3140	-0.0297	0.180	4997
	(0.0178)			(0.0174)			(0.0221)		
Alone to neighbour's	-0.0011	0.952	3141	-0.0024	0.900	3137	-0.0119	0.569	4995
	(0.0187)			(0.0187)			(0.0209)		
Alone outside village	0.0019	0.897	3146	0.0015	0.917	3142	-0.023	0.204	4996
	(0.0148)			(0.0145)			(0.0181)		
Alone to place of worship	0.0093	0.678	3145	0.0087	0.695	3141	-0.0126	0.628	4998
	(0.0224)			(0.0221)			(0.0260)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

In the pregnant woman sample, Gram Varta had a positive yet insignificant effect on being allowed to go to the market, health facility, neighbour's home or place of worship alone (Table 7.11). While this points towards an increased independence, the difference-in-differences specification of active SHG members showed that Gram Varta reduced the probability of being allowed to go to places alone (except for going to the neighbours' place). Similarly, Table 7.11 shows a negative but insignificant effect on the woman feeling recognized as herself in the community. While this was confirmed in the SHG member analysis, the coefficients were not significant at conventional levels.

Table 7.11: Independence from husband, pregnant women sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Feel recognized as myself	-0.0204	0.350	1603	-0.0303	0.182	1529	-0.0162	0.497	3595
	(0.0218)			(0.0227)			(0.0238)		
Allowed to go									
Alone to market	0.0364	0.213	1,605	0.0382	0.197	1,532	0.0302	0.409	3602
	(0.0293)			(0.0296)			(0.0364)		
Alone to health facility	0.0265	0.322	1609	0.0257	0.350	1535	0.0134	0.660	3606
	(0.0268)			(0.0275)			(0.0304)		
Alone to neighbour's home	0.0255	0.443	1608	0.0259	0.439	1534	0.0615	0.244	3605
	(0.0333)			(0.0334)			(0.0527)		
Alone to friends/relatives	0.0248	0.388	1602	0.0249	0.396	1528	-0.0024	0.939	3599
	(0.0288)			(0.0294)			(0.0308)		
Alone to place of worship	0.0214	0.572	1608	0.0230	0.553	1534	0.0388	0.415	3605
	(0.0379)			(0.0388)			(0.0475)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

When we considered the six blocks of our study area, we found diverging results. In Gwalpara there was a robust and mostly significant decrease in the probability of being allowed to go to places alone, while we found significant increases in these indicators for Murliganj and Uda Kishanganj.

Overall, we did not find convincing support for our hypothesis that Gram Varta enabled women to become more independent of their husbands in the household and pregnant women sample. This is a crucial piece of evidence, since it fails to support key channels of change as per the ToC.

Our findings for hypotheses 11 and 12 were complemented by the findings from our qualitative study. Most of the SHG member respondents in the qualitative study were women from

historically marginalized communities. A majority of these women were illiterate. Their mobility, decisions regarding their health, income, and their sense of identity was defined and limited by patriarchal norms. In the earlier rounds of qualitative study, we found that women had very little power in exercising decisions on health and nutrition of their family. Based on our interviews and FGDs, it seemed that women's sense of agency in deciding for their health and making informed actions had changed very little even after Gram Varta's inception:

"*It was a terrible night. I knew I had to go to the hospital to deliver this baby and had tried telling my husband and in-laws. But they insisted that I have this baby in our home. When the labor was taking too much time, they finally called up the Anganwadi sevika and the dai. Upon their insistence, I was taken to the hospital and my life was saved….Being a woman makes you subservient to others, so people here never want to have daughters. Who will pay for the hefty dowry when we have to get them married?" (SHG member, 28 years, treatment village 3).*

Hypothesis 13

Results related to whether Gram Varta enabled women to develop an identity of their own based on the household sample are described in this section. In the difference-in-differences specification, treatment decreased the probability of the woman being known as the husband's wife, at a significance level of 10 percent (table 7.12). Although the coefficient had the same sign in the subsample of SHG members, it was insignificant. These results are in line with our hypothesis that Gram Varta enables women to develop an identity on their own.

Table 7.12: Own identity, household sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Known as husband's wife	-0.0012	0.955	3019	-0.0006	0.978	3015	-0.068	0.072	4891
	(0.0209)			(0.0208)			(0.0376)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

Hypothesis 14

This section describes the results from investigating the hypothesis that Gram Varta increases women's involvement in the community based on the household sample. This relates to another crucial set of assumptions in the ToC (Stage 4): If a majority of participants speak up in the group and indicate their commitment to certain practices, all participants are assumed to have committed to it and to lead to a change of social norms within the community regarding the behaviour. This is assumed to help them once they are back in their household and have to discuss the change in practices with their relatives, as they can refer to other village members who practice the desirable behaviours already.

A few significant effects were found related to women's involvement in the community (Table 7.13). Treatment increased the probability that a woman was acquainted with health staff, at a significance level of 10 percent using endline data and 5 percent using the difference-in-differences specification. Treatment also seemed to increase the probability that a woman was acquainted with government officials at a significance level of 5 percent for the first and third model specification. In the difference-in-differences specification, treatment also significantly increased the probability that the woman was acquainted with school officials or any other officials from the village panchayat or ward committee, at the 10 and 5 percent level respectively. However, this latter effect was not consistent across specifications. The effect on whether the woman identifies her gender as the reason for not being allowed to participate in community activities was positive and highly significant in the first two specifications. Finally, voting in the last national election seemed positively influenced by treatment.

When adjusting for multiple testing, the results from Model 1 and 2 were rendered insignificant. In contrast, while adjusting for multiple testing is potentially a too conservative approach given the positive correlation between many of the hypotheses, the findings from the DiD model that women were increasingly likely to be acquainted with government or other officials remained significant.

Moreover, in the analysis of active SHG members we also found positive and significant effects on being acquainted with government or other officials. Similarly, positive effects on being acquainted with different officials were confirmed in most subsamples when splitting the full sample according to age groups, education level, caste or block.

Results from our qualitative studies present some insight on the mechanisms explaining the increase in community involvement. Gram Varta meetings were reported to have better participation if they were attended by health workers or government officials. Such meetings were generally attended by the entire community and the attendees were introduced to the health workers or government officials. Moreover, the participation of Jeevika officials in the meetings also ensured the contact of SHG members with Jeevika officials other than the facilitator. If the facilitator reported problems in conducting the meetings, she would be accompanied to meetings by a more experienced Gram Varta facilitator or other trained officials. Additionally, the regular activities of the SHG provided members with opportunities to be in contact with their local government officials and health staff.

Table 7.13: Involvement in the community, household sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Acquainted with									
Health staff	0.0324	0.082	3107	0.0265	0.151	3103	0.0643	0.024	4978
	(0.0186)			(0.0185)			(0.0283)		
	0.041	0.043	3100	0.0333	0.093	3096	0.0876	0.002**	4965

Government officials	(0.0202)			(0.0198)			(0.0281)		
School officials	0.0096	0.291	3035	0.0066	0.460	3031	0.0259	0.095	4920
	(0.0091)			(0.0090)			(0.0154)		
Other officials	-0.0024	0.907	3041	-0.0052	0.798	3037	0.1099	0.009*	4928
	(0.0205)			(0.0203)			(0.0413)		
Participation in community activities not allowed	-0.0063	0.719	2994	-0.0058	0.736	2990	-0.0049	0.831	4870
	(0.0176)			(0.0173)			(0.0228)		
Reason Gender	0.1646	0.016	321	0.1693	0.012	320	0.3333	0.167	316
	(0.0681)			(0.0671)			(0.2400)		
Voting	0.0256	0.129	3087	0.017	0.299	3083	0.0384	0.056	4953
	(0.0169)			(0.0164)			(0.0199)		
Attendance in public meeting	0.0059	0.669	3034	0.0066	0.634	3030	0.0075	0.744	4874
	(0.0139)			(0.0140)			(0.0230)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates. Stars indicate that p-value remains significant after multiple testing adjustment.

Taken together, our findings suggest that Gram Varta had a positive influence on women's social capital, allowing us to confirm our hypothesis.

Hypothesis 16

Whether Gram Varta made women self-confident in refusing sexual intercourse with their husband or demanding him to use a condom is discussed in the following section. Results are based on data from the household sample.

Table 7.14: Self-confidence in refusing sexual intercourse and demanding a condom, household sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Wife can refuse intercourse if									
Husband has STD	0.0269	0.438	2660	0.0288	0.405	2656	0.1561	0.001**	4168
	(0.0346)			(0.0347)			(0.0442)		
Husband cheats	0.0622	0.091	2807	0.0643	0.080	2803	0.1554	0.003*	4340
	(0.0368)			(0.0368)			(0.0516)		
Woman is tired	0.0391	0.247	2873	0.0398	0.240	2869	0.1109	0.013*	4376
	(0.0338)			(0.0338)			(0.0444)		

Women can demand condom use	-0.0067	0.813	2290	-0.0026	0.926	2286	0.0472	0.239	3615
	(0.0282)			(0.0284)			(0.0400)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates. Stars indicate that p-value remains significant after multiple testing adjustment.

Table 7.14 shows an overall positive effect on women's self-confidence. Coefficients were consistently positive across specifications for the first three indicators. In the first and second specification, treatment increased the probability of the woman agreeing that a wife can refuse intercourse if the husband has sex with other women, with marginal significance. This effect was highly significant in the third specification. Using the difference-in-differences specification, the intention-to-treat effect also reached the 5 percent significance level for the woman's opinion that the wife can refuse intercourse if the husband has an STD and if the wife is tired or not in the mood. Effect sizes were low in the first two specifications, but increased to about 16 and 11 percentage points in the difference-in-differences specification. These effects were also found in the subsample of SHG members, although the effect sizes differ slightly. When we split the full sample according to the age of the woman, we found that effects were stronger for women aged 18 to 35 compared to women aged 36 and above. In the sample with younger women, all effects were positive and several were significant at the 5 percent level. In contrast to that, effects were more inconsistent across specifications and insignificant for the sample with older women.

Notably, while multiple testing adjustment again rendered results from Models 1 and 2 insignficant, the DiD analysis-based results remained signficiant, possibly due to the higher statistical power of the DiD sample. Overall, the above findings remained largely robust, even when multiple testing was accounted for.

Our findings from the qualitative data collected at midline and endline suggest that Gram Varta meetings acted as a space for women to openly discuss their personal problems. However, in rare cases where the facilitator was unmarried, the respondents indicated difficulty in talking about sexual issues. The following quotes suggest that there was a difference in the nature of participation when the facilitator was married compared to when she was a single woman. Our FGDs also indicated that respondents felt more confident in addressing intimate issues with a married woman.

"*The Didi who comes to tell us about health is a stranger to us in this village. She is very young and is not married. She talks about maintaining the health of pregnant women and children, but we do not feel comfortable talking to her about our problems. We let her talk though… (laughing) ...how can she even understand my health problems?" (Rekha, IDI, pregnant woman (March 2016), 24 years, SHG member.)*

"*Our CM didi is very educated and she has told us many things about keeping ourselves healthy. Initially I thought that these meetings were boring and only attended these meetings just because CM didi had asked us to attend. However, when I started playing these games and activities, I began to have questions and asked her in private." (Seema, IDI, married, 19 years, SHG member).*

This qualitative evidence suggests that the effect on self-confidence, when it comes to refusing sexual intercourse with their husband or demanding him to use a condom, may have been reduced in cases where the facilitators were unmarried. Thus, we expect the effect to be larger when considering SHGs with a married facilitator only.

Overall, we found that Gram Varta does strengthen women's self-confidence regarding sexual behaviour, especially among younger women.

Hypothesis 17

This section describes the results of testing the hypothesis that Gram Varta reduces women's acceptance of domestic violence. Data are from the household sample. This hypothesis ties in with the ToC assumption related to an increase in women's empowerment from attending, participating and learning from Gram Varta activities (Stage 4).

Effects on women's acceptance of domestic violence were overall inconsistent, with coefficients' signs switching across specifications and coefficient sizes being very close around zero (Table 7.15). One exception was the increase in the probability of agreeing that the husband is justified in beating his wife if she burns the food. The direction of this effect was not as expected. With most indicators of acceptance of domestic violence, however, signs were as expected, especially for girls' acceptance of domestic violence, although insignificant.

Table 7.15: Acceptance of domestic violence, household sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Husband justified in hitting wife if she									
Goes out without telling	-0.0092	0.772	2134	-0.0077	0.810	2132	-0.0275	0.624	4218
	(0.0317)			(0.0318)			(0.056)		
Neglects children	0.041	0.261	1759	0.0422	0.246	1757	-0.0018	0.973	3944
	(0.0365)			(0.0364)			(0.0545)		
Argues with husband	-0.0006	0.984	2088	0.0102	0.743	2087	0.0405	0.384	4195
	(0.0312)			(0.0313)			(0.0464)		
Refuses sex	-0.0045	0.870	2370	-0.0039	0.889	2367	-0.0672	0.110	4406
	(0.0277)			(0.0278)			(0.0418)		
Burns food	0.0471	0.179	1409	0.0575	0.094	1408	0.0348	0.545	3649
	(0.035)			(0.0344)			(0.0574)		
Husband justified in hitting wife if she									
Goes out without telling (girl)	-0.003	0.938	602	-0.0004	0.992	602			
	(0.0391)			(0.039)					

Neglects children (girl)	-0.055	0.173	598	-0.0506	0.209	598
	(0.0403)			(0.0403)		
Argues with husband (girl)	-0.0017	0.967	594	0.0064	0.874	594
	(0.0412)			(0.0403)		
Refuses sex (girl)	-0.0268	0.478	557	-0.0218	0.554	557
	(0.0378)			(0.0368)		
Burns food (girl)	-0.0304	0.445	592	-0.0248	0.523	592
	(0.0398)			(0.0388)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

In the subsample of active SHG members, the effect on the indicator about burning food was not confirmed but a different effect was found. Gram Varta reduced the probability that the woman believes a husband is justified to hit his wife if she refuses sex. This coefficient was significant at the 5 percent level in specification 3 and at the 10 percent level in the first two specifications.

Overall, we did not find convincing evidence that Gram Varta reduced women's acceptance of domestic violence.

Hypothesis 18

The effect of Gram Varta on the practice of domestic violence and oppression based on the household and pregnant women samples are described in this section. In the household sample, several effects could be found on indicators of domestic violence (Table 7.16). In all three specifications, treatment significantly reduced the probability that the husband limited the woman's contact with her family. The effect size was almost twice as high in the difference-in-differences specification as in the first two, at about 10 percentage points. The same held for the effect on the probability that the husband insists on knowing where the woman is at all times. Although smaller, the effect was significant and negative in all three specifications. Gram Varta also had an effect on the reporting of kicking, dragging and beating in the past 12 months. In all three specifications, the coefficient was negative and significant at the 5 percent level. Marginal effects of the ordered logit model imply that reporting of regular beating decreased. Similarly, treatment appeared to decrease the probability of the husband forcing the woman to perform non-consensual sexual activities. The coefficients were negative and significant in all three specifications and the direction of marginal effects was in line with expectations. Although not significant, the direction of coefficients also implied a reduction in the remaining two dimensions of domestic violence.

Table 7.16: Domestic violence and oppression, household sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N

Afraid of husband									
Most of time	0.0002	0.976	2672	0.0000	0.998	2670	-0.0769	0.130	4641
	(0.0071)			(0.0072)			(0.0505)		
Sometimes	0.0007	0.976	2672	0.0001	0.998	2670			
	(0.0225)			(0.0224)					
Never	-0.0009	0.976	2672	-0.0001	0.998	2670			
	(0.0296)			(0.0296)					
Husband									
... is jealous	0.0076	0.704	2789	0.0093	0.640	2786	-0.0197	0.546	4696
	(0.0199)			(0.0199)			(0.0326)		
... accuses wife of cheating	-0.0123	0.497	2596	-0.0111	0.538	2594	-0.053	0.026	4547
	(0.0181)			(0.0181)			(0.0236)		
... forbids meeting friends	-0.0109	0.554	2612	-0.0097	0.591	2670	-0.0312	0.311	4547
	(0.0184)			(0.0181)			(0.0307)		
...limits contact with familiy	-0.0574	0.003	2484	-0.0586	0.002	2482	-0.1032	0.001**	4438
	(0.0195)			(0.0193)			(0.0300)		
... insists on knowing whereabouts	-0.0318	0.075	2745	-0.0325	0.067	2742	-0.0755	0.004**	4642
	(0.0179)			(0.0177)			(0.0260)		
Push/shake/throw at wife									
In last 12 months	-0.0184	0.225	2846	-0.0151	0.294	2842	-0.012	0.583	4773
	(0.0151)			(0.0144)			(0.0218)		
Never	0.0182	0.231	2846	0.0147	0.311	2842	-0.0181	0.496	4773
	(0.0152)			(0.0145)			(0.0266)		
Sometimes	-0.0161	0.231	2846	-0.0129	0.312	2842			
	(0.0134)			(0.0128)					
Often	-0.0021	0.231	2846	-0.0018	0.311	2842			
	(0.0018)			(0.0018)					
Kick/drag/beat wife									
In last 12 months	-0.0344	0.015	2536	-0.0309	0.021	2533	-0.0372	0.043	4527
	(0.0142)			(0.0134)			(0.0182)		
Never	0.0339	0.017	2536	0.0301	0.998	2533	-0.0373	0.082	4527
	(0.0142)			(0.0133)			(0.0213)		
Sometimes	-0.0298	0.017	2536	-0.0264	0.024	2533			
	(0.0125)			(0.0117)					
Often	-0.0041	0.021	2536	-0.0037	0.029	2533			
	(0.0018)			(0.0017)					
Choke/burn wife									
	-0.015	0.127	2846	-0.0131	0.175	2842	-0.0189	0.111	4771

In last 12 months	(0.0098)			(0.0096)			(0.0118)		
Never	0.0149	0.129	2846	0.0128	0.998	2842	-0.0192	0.152	4771
	(0.0098)			(0.0096)			(0.0133)		
Sometimes	-0.0124	0.132	2846	-0.0107	0.185	2842			
	(0.0082)			(0.0081)					
Often	-0.0024	0.136	2846	-0.0021	0.187	2842			
	(0.0016)			(0.0016)					
Forced sexual activities									
In last 12 months	-0.0206	0.048	2992	-0.0189	0.065	2988	-0.0248	0.174	4873
	(0.0104)			(0.0102)			(0.0182)		
Never	0.0205	0.049	2992	0.0188	0.998	2988	-0.0297	0.274	4873
	(0.0104)			(0.0102)			(0.0271)		
Sometimes	-0.0175	0.049	2992	-0.016	0.066	2988			
	(0.0089)			(0.0087)					
Often	-0.003	0.062	2992	-0.0028	0.080	2988			
	(0.0016)			(0.0016)					

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

With multiple testing accounted for, we continued to find significant evidence in the DiD model for the husband being less likely to insist on knowing where the woman was and to limit contact to her family. The latter effect was also robust to a change in sample to SHG members only. Treatment seemed to reduce the probability that the husband limits his wife's contact with her family, by about 9 percentage points, similar to the full sample. Moreover, the direction of effects on different forms of domestic violence was confirmed in this subsample analysis.

In the subgroup analysis, we found the expected effect signs, implying a reduction in the practice of domestic violence in the four dimensions. One exception was when we looked at the subgroup of households from Scheduled Castes. In that subgroup, effects were in the opposite direction in the first three dimensions, implying an increase in domestic violence. For this subgroup, we also found a consistent increase in the probability of being afraid of the husband, while marginal effects for this indicator were very close around zero for the full sample.

Table 7.17: Domestic violence, pregnant women sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N

Pushed/shaken/having things thrown at by husband (Yes/no)	-0.0312	0.374	1,293	-0.0308	0.403	1,236	-0.0403	0.349	3,29
	(0.0351)			(0.0368)			(0.0428)		
Kicked/dragged/beaten up by husband (Yes/no)	-0.0314	0.328	1,312	-0.0231	0.505	1,256	-0.0443	0.320	3,309
	(0.0320)			(0.0347)			(0.0428)		
Choked/burned on purpose by husband (Yes/no)	-0.0456	0.105	1,324	-0.0410	0.181	1,269	-0.0578	0.074	3,321
	(0.0281)			(0.0307)			(0.0321)		
Forced to do perform non-consensual sexuality activities by husband (Yes/no)	-0.0480	0.116	1,304	-0.0437	0.170	1,249	-0.0774	0.019	3,301
	(0.0306)			(0.0319)			(0.0325)		
Pushed/shaken/having things thrown at by husband									
Never	0.0331	0.344	1,293	0.0318	0.388	1,236	-0.0732	0.185	3,256
	(0.0350)			(0.0368)			(0.0549)		
Sometimes	-0.0304	0.343	1,293	-0.0289	0.387	1,236			
	(0.0320)			(0.0334)					
Often	-0.0028	0.360	1,293	-0.0028	0.401	1,236			
	(0.0030)			(0.0034)					
Kicked/dragged/beaten up by husband									
Never	0.0340	0.287	1,312	0.0251	0.313	1,256	-0.0840	0.121	3,283
	(0.0320)			(0.0249)			(0.0539)		
sometimes	-0.0302	0.286	1,312	-0.0221	0.311	1,256			
	(0.0283)			(0.0218)					
Often	-0.0039	0.304	1,312	-0.0031	0.334	1,256			
	(0.0037)			(0.0032)					
Choked/burned on purpose by husband									
Never	0.0453	0.106	1,324	0.0407	0.055	1,269	-0.0633	0.082	3,311
	(0.0281)			(0.0213)			(0.0362)		
Sometimes	-0.0395	0.107	1,324	-0.0354	0.056	1,269			
	(0.0245)			(0.0185)					
Often	-0.0058	0.119	1,324	-0.0053	0.062	1,269			
	(0.0037)			(0.0028)					
Forced to do perform non-consensual sexuality activities by husband									
Never	0.0488	0.112	1,304	0.0450	0.062	1,249	-0.1183	0.004	3,294
	(0.0307)			(0.0241)			(0.0407)		
Sometimes	-0.0420	0.112	1,304	-0.0383	0.061	1,249			
	(0.0264)			(0.0204)					
Often	-0.0068	0.130	1,304	-0.0067	0.081	1,249			
	(0.0045)			(0.0039)					

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

In the pregnant woman sample, the coefficients show the expected negative signs on the probability of domestic violence in the four studied dimensions: (1) pushed/shaken/having things thrown at by husband, (2) kicked/dragged/beaten up, (3) choked/burned on purpose, (4) forced to do perform non-consensual sexual activities (Table 7.17). These effects were significant in the difference-in-differences estimation for choked or burned and sexual violence at the 10 and 5 percent significance levels, respectively, with an effect size of 4.1 to 5.8

percentage points. Looking at the frequency of domestic violence, specifications 1 and 2 showed that Gram Varta increased the probability of never being abused while it reduced the probability of being abused sometimes or often. As expected, the difference-in-differences coefficients showed a negative impact of Gram Varta on the frequency of domestic violence. Again, the coefficients for choked or burned and sexual violence were significant at the 10 and 1 percent significance levels. It should be noted however that all effects turned insignificant when multiple testing was adjusted for.

Nevertheless, the overall pattern of the effects on domestic violence was confirmed when looking at the sample of active SHG members, even though none of the effects was significant. Unfortunately, the variation in domestic violence was too small to estimate ordered logit marginal effects of domestic violence frequency in specification 2. Still, specifications 1 and 3 confirmed the negative impact on the frequency of domestic violence. Specification 1 also confirmed the negative impact on the probability of domestic violence.

In the subgroup analysis by block we found a robust and significant increase in domestic violence in Bihariganj. Gram Varta increased the probability of being abused by the husband by 13 to 36 percentage points depending on the type of violence and specification. However, in Murliganj, Madhepura Sadar and Uda Kishanganj we found robust and significant evidence that Gram Varta reduced the probability and frequency of domestic violence exercised by husbands. Depending on the block, specification and type of violence the effect ranged between 10 and 22 percentage points. In the subgroup analysis by caste, we found significant reductions in the probability and frequency of domestic violence for other backward castes and general category castes. Depending on caste and specification the probability of domestic violence dropped by 6.5 to 20.5 percentage points. In the subgroup analysis by education, we found robust evidence that Gram Varta reduced domestic violence. For women with no education, all signs pointed in the right direction while only the effect on the probability of being sexually abused by the husband was significant. However, when we turn to women that completed junior secondary school or higher, all coefficients had the expected sign and most coefficients were significant.

To summarize our results, we found support for our hypothesis that Gram Varta reduces the practice of domestic violence and oppression in general, even though these findings were not robust to multiple testing adjustment. Notably, for women in the Bihariganj block we found that the opposite holds true.

Hypothesis 19

This section describes the results of the investigation that Gram Varta gave adolescent girls a more positive outlook on their future based on the household sample.

Table 7.18: Positive outlook of adolescent girls, household sample

	Model 1			Model 2		
	B/SE	P value	N	B/SE	P value	N

Lack of self-confidence						
No obstacle	0.0548	0.134	275	0.0601	0.116	275
	(0.0366)			(0.0383)		
Somewhat obstacle	0.0433	0.104	275	0.0466	0.086	275
	(0.0266)			(0.0271)		
Definitely obstacle	-0.0981	0.113	275	-0.1067	0.096	275
	(0.0620)			(0.0640)		
Lack of family support						
No obstacle	-0.0002	0.995	223	-0.0021	0.959	223
	(0.0403)			(0.0412)		
Somewhat obstacle	-0.0001	0.995	223	-0.0013	0.959	223
	(0.0262)			(0.0262)		
Definitely obstacle	0.0004	0.995	223	0.0035	0.959	223
	(0.0665)			(0.0675)		
Marriage						
No obstacle	-0.0107	0.713	579	-0.0074	0.795	579
	(0.0291)			(0.0287)		
Somewhat obstacle	-0.0059	0.712	579	-0.004	0.794	579
	(0.0160)			(0.0155)		
Definitely obstacle	0.0154	0.712	579	0.0106	0.795	579
	(0.0419)			(0.0409)		
Domestic responsibility						
No obstacle	-0.0198	0.563	426	-0.02	0.561	426
	(0.0341)			(0.0344)		
Somewhat obstacle	-0.0098	0.568	426	-0.0098	0.565	426
	(0.0172)			(0.0171)		
Definitely obstacle	0.0282	0.563	426	0.0284	0.561	426
	(0.0488)			(0.0489)		
Mobility						
No obstacle	0.0166	0.602	297	0.0191	0.544	297
	(0.0320)			(0.0314)		
Somewhat obstacle	0.0138	0.605	297	0.0154	0.549	297
	(0.0267)			(0.0257)		
Definitely obstacle	-0.0304	0.603	297	-0.0345	0.545	297
	(0.0586)			(0.0570)		
Husband as decision maker	-0.0091	0.869	464	-0.009	0.870	464
	(0.0553)			(0.0547)		
Expected asset ownership						

Less than parents	0.0025	0.877	301	0.0034	0.837	301
	(0.0162)			(0.0167)		
Same as parents	0.0062	0.878	301	0.0083	0.840	301
	(0.0403)			(0.0412)		
More than parents	-0.0087	0.878	301	-0.0118	0.839	301
	(0.0564)			(0.0578)		
Expects to be housewife	0.0068	0.892	336	-0.0014	0.978	336
	(0.0502)			(0.0512)		
Copy parents' role model	-0.0684	0.230	337	-0.0735	0.189	337
	(0.0571)			(0.0559)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values and Model 2 further adds controls. Model 3 could not be estimated as the sample is limited to adolescent girls.

Albeit not being robust to the Benjamini-Hochberg correction, we found some evidence that Gram Varta decreased the probability that lack of self-confidence is seen as a definite obstacle to the girl's career while it increased the probability that it is seen as somewhat of an obstacle (table 7.18). Quantitatively, however, the negative effect was larger by 5 percentage points. Similar results were seen with mobility as a factor in the girl's future, although effect sizes were smaller and not significant. In contrast, the probability that marriage or domestic responsibility is seen as a definite obstacle increased. Another effect in line with the hypothesis was related to the probability that the girl wants to copy her parents' model of men's and women's role. For adolescent girls in households with active SHG members, we found similar effects on what is seen as an obstacle to the girl's future and whether the girl wants to copy her parents, but the coefficients were insignificant.

Overall, we did not find conclusive evidence that Gram Varta gives adolescent girls a more positive outlook on their future based on the household sample.

Hypothesis 20

Gram Varta reduces adolescent girls' preferred number of children based on the household sample—the results of testing this hypothesis are described here. This tests the ToC assumption that targeting of adolescents (Stage 2) was successful and it led to them changing their thinking (Stage 3) about their own health and well-being.

In neither specification nor sample was a significant effect on the adolescent girl's desired number of children found (table 7.19). Adding baseline covariates changed little in the coefficients and standard errors. Moreover, the coefficient was positive, implying an increase in the desired number of children. Splitting the sample for subgroup analyses resulted in very small sample sizes which makes an interpretation difficult.

Table 7.19: Preferred number of children of adolescent girls, household sample

	Model 1			Model 2		
	B/SE	P value	N	B/SE	P value	N
Desired number of children	0.0765	0.226	360	0.0790	0.182	360
	(0.0630)			(0.0589)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values and Model 2 further adds controls. Model 3 could not be estimated as the sample was too small because it was limited to adolescent girls.

Thus, we reject our hypothesis that Gram Varta reduces adolescent girls' preferred number of children based on the household sample.

Hypothesis 21

This section investigates the hypothesis that Gram Varta decreased women's preference for sons based on the household sample. Several related assumptions in the ToC led us to investigate this: that women's critical thinking and the importance they give to gender equality increases (Stage 3).

Table 7.20: Preference for son of adolescent girls, household sample

	Model 1			Model 2		
	B/SE	P value	N	B/SE	P value	N
Preference for boy	-0.1459	0.0170	284	-0.1454	0.0190	284
	(0.0609)			(0.0622)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values and Model 2 further adds controls. Model 3 could not be estimated as the sample was too small because it was limited to adolescent girls.

In both specifications, with and without baseline controls, treatment significantly decreased adolescent girls' preference that her first or next child be a boy (table 7.20). With a reduction of about 15 percentage points, the coefficient was also quantitatively large. In the subsample of exposed adolescent girls, however, the effect size was much smaller and significance disappeared. Moreover, results presented in Table 7.20 were not robust to multiple testing adjustment. We therefore found only weakly supportive evidence that Gram Varta decreased women's preference for sons.

Hypothesis 22

This section investigates the hypothesis that Gram Varta makes it more likely that women desire a higher age at marriage for themselves or their daughters based on the household sample.

Table 7.21: Age of marriage, household sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Opinion of woman									
Preferred age of marriage for daughter	0.0467	0.636	3050	-0.0028	0.975	3046	-0.0940	0.393	4937
	(0.0983)			(0.0880)			(0.1097)		
Preferred age of first birth for daughter	0.0172	0.889	2918	-0.0323	0.767	2914	-0.0546	0.732	4838
	(0.1231)			(0.1088)			(0.1590)		
Opinion of adolescent girl									
Ideal age of marriage	0.2316	0.349	616	0.177	0.446	616			
	(0.2468)			(0.2319)					
Difference in marriage age between boys and girls	0.286	0.087	601	0.2511	0.130	601			
	(0.1664)			(0.1651)					
Believe marry too young	0.0119	0.797	352	0.0129	0.779	352			
	(0.0464)			(0.0460)					
Pressure to marry	-0.025	0.488	584	-0.0213	0.544	584			
	(0.0360)			(0.0351)					
Preferred own marriage age	-0.2019	0.650	157	-0.2572	0.492	157			
	(0.4435)			(0.3723)					

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

No coefficient for any indicator under this hypothesis was significant (table 7.21). Only one coefficient reached marginal significance. In the first model specification without covariates, treatment seemed to increase the difference in ideal marriage ages according to the adolescent girls. Since the difference was calculated as boy's ideal marriage age minus girl's ideal marriage age, an increase in age difference was not in line with the hypothesis.

A similar effect, although insignificant, was found in the subsample of adolescent girls whose mother was a SHG member. In this subsample, we also found a negative treatment effect on

the mother's preferred age at which her daughter should have her first child. Again, the direction of the effect is contrary to the hypothesis.

In the full sample, effects on different indicators seem contradictory. While the ideal marriage age according to adolescent girls increased, their preferred age for their own marriage decreased. This was despite the decrease in the probability that they feel pressured to marry and increase in the probability of believing that girls marry too young.

Based on these results we reject our hypothesis that Gram Varta makes it more likely that women desire a higher age at marriage for themselves or their daughters.

Hypothesis 23

This section describes the results of testing the hypothesis that Gram Varta reduced the likelihood of early pregnancies based on the household sample. The ToC shows how Gram Varta was expected to lead to change in health-related practices (Stage 5), especially about women's health, because the participants were expected to realize the importance of maternal health from a lifecourse perspective.

Table 7.22: Early pregnancies, household sample

	Model 1			Model 2		
	B/SE	P value	N	B/SE	P value	N
Early pregnancy	-0.0539	0.595	72	-0.0305	0.755	72
	(0.1014)			(0.0979)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values and Model 2 further adds controls. Model 3 could not be estimated as the sample is limited to adolescent girls.

In neither model specification did the coefficient on early pregnancy reach significance (Table 7.22). One reason might be the low number of observations for this question, which makes calculations for most subgroups impossible. The coefficient's sign was negative as expected, in both the full and the subsample of girls whose mother was an active SHG member. However, this is not sufficient support to confirm the hypothesis that Gram Varta reduces the likelihood of early pregnancies based on the household sample.

Hypothesis 24

This section investigates the hypothesis that Gram Varta improved attitudes towards and practices of care for daughters based on the household and pregnant women sample.

In the household sample, we found significant effects on whether there is enough food for the daughter. While significance was not robust to the Benjamini-Hochberg correction, marginal effects pointed in the expected direction (table 7.23). Moreover, specifications 1 and 2 showed some marginally significant effects on whether the adolescent girl usually ends up eating last. However, opposite to the expectation, the effect was negative meaning adolescent girls end up eating last more often. We found similar results for the indicator whether the adolescent girl is given less food during her period. The effect of Gram Varta treatment was marginally significant and negative, contrary to the hypothesis. It should be noted however that significance disappeared when multiple testing was accounted for. Moreover, neither of the effects on the importance of giving sufficient food to the daughter, opinions on equal access to education nor the ease of talking with one's mother was significant at conventional levels. Lastly, coefficients for the women's view on food for the daughter and equal access to education were inconsistent.

Table 7.23: Care for daughters, household sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N

Answer of household head									
Importance of food for daugher	-0.0003	0.949	3534	-0.0005	0.911	3528	0.0142	0.179	7396
	(0.0041)			(0.0041)			(0.0105)		
Answer of woman									
Girls and boys same access to education	0.0002	0.982	3117	-0.0004	0.963	3113	0.0154	0.239	4964
	(0.0083)			(0.0082)			(0.0131)		
Answers of adolescent girl									
Enough food for daughter									
Yes	0.0339	0.356	587	0.0388	0.285	587			
	(0.0367)			(0.0363)					
Sometimes	0.0125	0.339	587	0.0141	0.271	587			
	(0.0130)			(0.0128)					
No	-0.0464	0.350	587	-0.0529	0.278	587			
	(0.0496)			(0.0488)					
Girl eats last									
Yes	0.065	0.094	589	0.0672	0.084	589			
	(0.0389)			(0.0389)					
Sometimes	0.0106	0.144	589	0.0109	0.140	589			
	(0.0072)			(0.0074)					
No	-0.0756	0.093	589	-0.0781	0.083	589			
	(0.0450)			(0.0451)					
Less food during period									
Yes	0.0872	0.091	436	0.0858	0.094	436			
	(0.0515)			(0.0513)					
Sometimes	0.0100	0.096	436	0.0098	0.095	436			
	(0.0060)			(0.0059)					
No	-0.0972	0.088	436	-0.0956	0.092	436			
	(0.0570)			(0.0567)					
Meals per day	0.0362	0.565	650	0.033	0.595	650			
	(0.0628)			(0.0619)					
Talk with mother									
Very easy	-0.0194	0.639	541	-0.0217	0.587	541			
	(0.0413)			(0.0400)					
Easy	0.0088	0.642	541	0.0097	0.592	541			
	(0.0189)			(0.0181)					
Difficult	0.0081	0.635	541	0.0091		541			
	(0.0170)			(0.0166)					
Very difficult	0.0025	0.645	541	0.0028	0.596	541			
	(0.0054)			(0.0053)					
Parents pay attention to daughter	0.0401	0.447	269	0.0333	0.515	269			
	(0.0527)			(0.0512)					

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

Similar results were found in the subsample of exposed adolescent girls. The effect on the probability of ending up eating last appeared significantly negative as in the full sample, while no other coefficient reached significance at conventional levels.

Table 7.24: Care for daughters, pregnant women sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Ensuring that daughters get enough food/attention important	-0.0009	0.876	1480	0.0020	0.760	1107	0.0021	0.869	3409
	(0.0056)			(0.0064)			(0.0129)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

In the pregnant women sample, the opinion about whether ensuring that daughters get enough food and attention is important was not influenced by Gram Varta as the coefficients were very small and insignificant throughout all specifications (table 7.24). Among SHG members, too few answered this question to evaluate it.

Overall, we reject the hypothesis that Gram Varta improves attitudes towards and practices of care for daughters.

Perhaps this is explained by the general perception of a lack of value of women observed in the qualitative study. Several women had internalized this:

“Some of us had acquired money from the SHG for my needs, but we were forced to give this money to my husband when he demanded money to buy alcohol. I fear the society if I disobey my Malik (master).” (FGD participant; master=husband)

“*Our jobs in the kitchen and in other domestic activities are not considered valuable by others. However a man’s job brings in money and he is considered indispensable for survival of the family.” (FGD participant)*

These quotes suggest that there is complex connection between gender relations and sense of empowerment. For the women in control and treatment villages, empowerment means the ability to have an education and earn an income and behaving in an appropriate manner that do not challenge the traditional norms and customs. Women interviewed in the qualitative study in both control and treatment areas reported feeling that as long as one can negotiate with the norms rather than directly challenging it, changes will be gradually seen in the

behaviour. Moreover, since gaining a sense of agency in this context is a very complex process that involves addressing the challenges posed by social constraints as well as physical limitations such as a lack of infrastructure, empowering women in these villages to stand up for themselves and make major decisions on HNWASH might take longer than two years.

Hypothesis 25

This section investigates the hypothesis that Gram Varta makes adolescent girls more confident to cook for themselves in order to take care of their own nutrition based on the household sample. A specific story is part of Gram Varta PLA that directly addresses this issue.

The effect on whether the adolescent girl feels confident to cook her own food if not enough is left for her was positive as expected, but insignificant (table 7.25). The same held for most subgroups. While this evidence is indicative it is not sufficient to confirm our hypothesis.

Table 7.25: Adolescent girls taking care of own nutrition, household sample

	Model 1			Model 2		
	B/SE	P value	N	B/SE	P value	N
Confident to cook own food	0.0431 (0.0512)	0.400	474	0.0494 (0.0509)	0.331	474

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values and Model 2 further adds controls. Model 3 could not be estimated as the sample is limited to adolescent girls.

Hypothesis 26

This section investigates the hypothesis that Gram Varta increases the husbands' support for their wives' SHG membership based on the pregnant women sample.

Table 7.26: Husband's support in SHG membership, pregnant women sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Husband supports attending SHG	0.0127 (0.0372)	0.733	375	0.0260 (0.0386)	0.501	361	-0.0501 (0.0723)	0.489	2372
Goes to fewer SHG meetings because of lack of support by husband/ family	-0.0648 (0.0570)	0.256	369	-0.0664 (0.0603)	0.271	355	0.0856 (0.1002)	0.394	875

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

While specifications 1 and 2 showed an (insignificant) increase in the husband's support and a decrease in the incidence of going to fewer meetings because of lack of support, the difference-in-differences estimations show the opposite (table 7.26). Thus we cannot confirm our hypothesis.

The qualitative study confirms that a 'patriarchal bargain' exists between women and men in Madhepura and that the husband's support for Gram Varta activities is less when it comes to 'radical decisions' in this context, such as women making decisions about her own health care. Activities in Gram Varta which involve questioning of traditional norms on gender behavior received very little support from both men and women. The following quote from an SHG member supports this idea:

"I have to follow what my in-laws and my husband tell me to do. I will be considered as a 'bad' woman if I oppose them. I don't want to have any fights with them." (IDI, SHG member, treatment village).

She learned from Gram Varta that she should have regular check-ups during her pregnancy. However, her husband and in-laws did not allow her to go to the hospital. She was forced to follow her family's decisions even though she knew that it was wrong.

Based on our quantitative and qualitative results, we reject the hypothesis that Gram Varta increases the husbands' support for their wives' SHG membership.

Summary

Gram Varta seemed to have a weak but positive influence on income earned by the woman and her decision power about that income (Hypotheses 10 and 11). However, this evidence was not sufficient to confirm our hypotheses. Hypothesis 12 about the independence of women was rejected. Gram Varta showed an improvement in the woman having an own identity, although this effect was not significant when multiple testing was accounted for. Gram Varta did, however, influence women's social capital positively, with some consistent effects remaining even under the Benjamini-Hochberg correction, confirming Hypothesis 14. Consistent effects were also seen with Hypothesis 16, implying that Gram Varta strengthened women's self-confidence when it comes to refusing sexual intercourse.

We observed a consistent and significant reduction in some forms of controlling behaviour by husbands (at least as reported by the interviewed women) and some evidence for a reduction in the incidence and frequency of domestic violence (except for women in Bihariganj) (Hypothesis 18). However, these effects were not consistently significant in the main analysis and never significant in the active SHG member analysis. Further we rejected the hypothesis that Gram Varta reduces women's acceptance of domestic violence (Hypothesis 17).

Among the hypotheses focusing on adolescent girls, hypotheses 19, 20, 22 and 24 could not be confirmed. Indicators for hypotheses 21, 23 and 25, however, were consistent and in the expected direction, although not consistently significant.

We also could not confirm the hypothesis that Gram Varta increases husband's support for SHG membership (Hypothesis 26).

7.4 Results related to HNWASH knowledge and practices

This group of hypotheses is related to the knowledge and practices related to nutrition, awareness and prevention of diseases, risky consumption behaviour, domestic hygiene and sanitation as well as adolescent girls' and women's knowledge on sexuality and contraception. Similar to the analysis related to women's agency and empowerment, the evaluation design allows a causal interpretation of the presented estimates. These hypotheses directly test major assumptions of the ToC (Stage 4) that important health-promoting knowledge is transferred to Gram Varta participants; that personal attitudes towards social norms regarding the desirable practices change; and that Gram Varta activities trigger women's critical thinking and problem solving skills and they individually are willing to commit to the desired behaviours.

Hypothesis 27

This section investigates the hypothesis that Gram Varta increases the intake of micronutrients based on the household sample.

Table 7.27: Intake of micronutrients, household sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Children received the following medicine:									
Vitamin A doses (last child)	-0.1096	0.157	1324	-0.1368	0.094	1323	-0.2884	0.070	1990
	(0.0771)			(0.0813)			(0.1584)		
IFA tablet/syrup (last child)	-0.0489	0.052	1557	-0.0572	0.026	1555	-0.0453	0.378	2464
	(0.0251)			(0.0256)			(0.0512)		
Any vitamin A dose (other child)	-0.0226	0.548	1091	-0.025	0.501	1089			
	(0.0377)			(0.0371)					
IFA tablet/syrup (other child)	-0.0113	0.757	1107	-0.014	0.701	1105			
	(0.0365)			(0.0363)					

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

The coefficient on the number of vitamin A doses received by the last born child was marginally significant in model specifications 2 and 3, but had a negative sign (table 7.27). This implies a reduction of vitamin A doses in treatment areas. Note, that the indicator counts the number of

vitamin A doses for the last born child, but for other children only asks whether they received any vitamin A dose. Similarly, treatment appeared to reduce the probability of the last born child having received an IFA tablet or syrup, at a significance level of 5 percent in model specification 2 and almost at 5 percent in model specification 1. For other children under the age of six, no significant effect was found in the full sample. Among other children of exposed mothers, however, a marginally significant and negative effect was found regarding whether the other child received any vitamin A dose. Again, the coefficient was opposite of expectations. Disregarding significance, almost all coefficients in both samples were negative, implying a worsening of micronutrient intake. This surprising finding was even more prominent for children of younger women (below 36 years of age) and children of women with lower educational background, namely those who completed no education level and those who did not complete education above middle school (8th grade completed). For these subgroups, all coefficients were negative and several significant at the 5 or 10 percent level.

In addition to most coefficients having the opposite sign than expected, they became insignficiant when multiple testing was adjusted for. Based on these results, we reject our hypothesis that Gram Varta increases the intake of micronutrients.

Hypothesis 28

This section investigates the hypothesis that Gram Varta raises awareness of the importance of a balanced diet for the family based on the household and pregnant women samples.

Table 7.28: Importance of balanced family diet, household sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Opinion on balanced nutrition	-0.0008	0.888	3481	-0.0003	0.953	3475	0.0093	0.270	7304
	(0.0054)			(0.0055)			(0.0084)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

For the opinion on a balanced nutrition, no significant effect was found in any of the specifications, independent of the sample, and the coefficient size was very small (table 7.28). This is not very surprising as the share of respondents who believe that a balanced nutrition is important was close to 100 percent at baseline itself.

Table 7.29: Intake of micronutrients, pregnant women sample

	Model 1	Model 2	Model 3

	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
PW or children cut size of meal	0.0278	0.821	96	0.1226	0.354	78	0.7143	0.370	224
	(0.1226)			(0.1324)			(0.7377)		
Ensuring a balanced diet to family is important	-0.0040	0.540	1,562	-0.0017	0.811	1,21	-0.0160	0.180	3,483
	(0.0065)			(0.0070)			(0.0119)		
PW decides about family diet	0.0254	0.453	1,596	0.0375	0.246	1,522	0.1045	0.033	3,569
	(0.0338)			(0.0323)			(0.0486)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

In the pregnant women sample, we found a positive effect on the probability that the woman or children cut the size of the meal while the effect on the probability that the woman thinks that ensuring a balanced diet to the family is important was negative (table 7.29). This was against our expectations. However, all coefficients were insignificant. Unfortunately, the variation among these outcomes in the active SHG sample was too low to test these indicators.

In line with expectations, Gram Varta increased the probability of the woman deciding about the family diet. This difference-in-differences estimate showed a large and significant increase of 10.5 percentage points in the probability of deciding about the family diet. It should be noted, however, that significance disappeared when multiple testing was accounted for. The active SHG member analysis confirmed a large and positive yet insignificant effect on the probability that the woman decides about the family diet.

Overall, the results are mixed and we could not confirm our hypothesis that Gram Varta raises awareness of the importance of a balanced diet for the family.

Hypothesis 29

This section investigates the hypothesis that Gram Varta improves knowledge and attitudes toward proper feeding of newborns based on the household and pregnant women samples.

Table 7.30: Attitudes towards feeding of newborns, household sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Feeding thick breastmilk	0.0162	0.175	1.633	0.0153	0.195	1631	0.0723	0.001	2570
	(0.0119)			(0.0118)			(0.0210)		
First breastfeeding									

Immediately	0.0112	0.699	1.686	0.0104	0.718	1684	-0.2304	0.004*	2657
	(0.0289)			(0.0287)			(0.0786)		
Within 24 hours	-0.0059	0.699	1.686	-0.0055	0.717	1684			
	(0.0152)			(0.0152)					
Within three days	-0.003	0.700	1.686	-0.0028	0.718	1684			
	(0.0079)			(0.0077)					
After three days	-0.0019	0.700	1.686	-0.0018	0.719	1684			
	(0.0050)			(0.0049)					
Never	-0.0004	0.699	1.686	-0.0003	0.718	1684			
	(0.0009)			(0.0009)					
Opinion of complementary feeding									
Not important	-0.0011	0.877	1.672	0.0003	0.970	1670	0.0525	0.448	2577
	(0.0073)			(0.0071)			(0.0691)		
Important	-0.0019	0.878	1.672	0.0004	0.970	1670			
	(0.0122)			(0.0118)					
Very important	0.0030	0.877	1.672	-0.0007	0.970	1670			
	(0.0194)			(0.0190)					
Pre-lacteal feeding	-0.0340	0.181	1680	-0.0356	0.166	1678	-0.0714	0.146	2626
	(0.0254)			(0.0257)			(0.0490)		
Frequency of breastfeeding per day									
2 times or less	0.0003	0.944	977	-0.0008	0.873	975	0.0182	0.912	1598
	(0.0048)			(0.0048)			(0.1646)		
3 to 6 times	0.0016	0.944	977	-0.0037	0.873	975			
	(0.0232)			(0.0231)					
7 to 9 times	0.0001	0.943	977	-0.0003	0.874	975			
	(0.0019)			(0.0018)					
On demand	-0.0013	0.944	977	0.0029	0.873	975			
	(0.0186)			(0.0184)					
10 to 12 times	-0.0005	0.944	977	0.0011	0.873	975			
	(0.0070)			(0.0069)					
More than 12 times	-0.0003	0.944	977	0.0007	0.874	975			
	(0.0044)			(0.0044)					
Duration of breastfeeding	39.5906	0.224	476	38.2016	0.243	475	-8.0345	0.927	804
	(32.4485)			(32.5835)			(87.7974)		
Meals per day	0.0394	0.712	1.024	0.0416	0.699	1023	0.3572	0.086	1946
	(0.1066)			(0.1073)			(0.2066)		
Importance of adding oil	-0.0503	0.166	1.569	-0.0517	0.151	1567	-0.0649	0.305	2362
	0.0363			(0.0360)			(0.0631)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates. Stars mark p-values which remain significant even under multiple testing adjustment.

In the household sample, Gram Varta seemed to increase the probability that feeding colostrum after delivery was seen as important (table 7.30). The difference-in-differences specification suggested a significant increase by 7.2 percentage points, although this was not robust to the Benjamini-Hochberg correction. The significant negative coefficient for timing of first breastfeeding implies that treatment leads to sooner breastfeeding after birth. In the remaining ordered logit models, results were not significant and effect sizes were very small. Gram Varta appeared to have a positive effect on the number of meals the last born child ate on the previous day. It also seemed to negatively affect the probability of giving the last child pre-lacteal feeding on the day of birth.

When looking at the subsample of last born children of SHG members, the results were similar. The effect on the first indicator, importance of feeding the thick, yellowish breastmilk, was also found to be positive and highly significant in the difference-in-differences specification. The signs of coefficients for the second indicator were in line with those found in the full sample, although insignificant. The effect on the number of meals per day eaten by the child, however, was inconsistent across specifications. Two additional significant effects found in that sample were an increase in the duration of breastfeeding and a decrease in the importance of adding extra oil to every meal of small children. Overall, there seemed to be some treatment effects on the feeding of newborns, but these were not robust and not consistent across specifications and not in line with the hypothesis for some indicators.

In the pregnant women sample, the coefficients on whether breast feeding is important for the baby's health were small, insignificant, and changed signs, whereas the coefficients of whether feeding the breast milk of the first three days is important or not were consistently positive, between 1.7 and 2.2 percentage points (table 7.31). While this seems contradicting, all effects were insignificant. The effects on complementary feeding after six months were small and insignificant, and the effects on the importance of adding oil to the child's meal pointed towards a negative yet insignificant effect. Overall there appeared no clear pattern of the effect of Gram Varta on opinions about child feeding. This was confirmed when looking at the active SHG member analysis. The effect on the probability of feeding breast milk after delivery was small, positive, and insignificant, while all other effects appeared inconsistent.

Table 7.31: Attitudes towards feeding of newborns, pregnant women sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Breastfeeding is important	0.0010 (0.0188)	0.959	1334	-0.0007 (0.0196)	0.973	1,275	-0.0047 (0.0208)	0.821	3315
Breastmilk after delivery	0.0169 (0.0203)	0.405	1324	0.0215 (0.0209)	0.303	1,265	0.0221 (0.0335)	0.511	3043
Complementary feeding after 6m.	0.0045 (0.0119)	0.707	1297	0.0016 (0.0126)	0.897	1,19	0.0030 (0.0178)	0.865	3175

Adding oil to meal of children	-0.0077 (0.0354)	0.828	1181	0.0078 (0.0361)	0.829	1,131	-0.0745 (0.0453)	0.102	2984

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

Based on these results we reject the hypothesis that Gram Varta raises awareness of the importance of a balanced diet for the family.

Hypothesis 30

This section investigates the hypothesis that Gram Varta encourages parents to prevent diseases in children, for example through vaccinations and bednets based on the household sample.

Table 7.32: Prevention of diseases in children, household sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
All vaccinations (last child)	-0.0019 (0.0406)	0.962	841	0.0042 (0.0409)	0.919	841	0.0118 (0.0774)	0.879	1456
All vaccinations (other child)	0.0601 (0.0417)	0.150	510	0.0524 (0.0410)	0.201	510			
Vaccination card (last child)	-0.0039 (0.0280)	0.891	1707	-0.0092 (0.0276)	0.738	1705	-0.0497 (0.0414)	0.232	2673
Vaccination card (other child)	0.0347 (0.0380)	0.361	1148	0.0234 (0.0369)	0.526	1147			
Bednet	-0.006 (0.0079)	0.447	1715	-0.0066 (0.0075)	0.378	1713	0.0336 (0.0184)	0.069	2677
Treated bednet	-0.0576 (0.0256)	0.024	1532	-0.0619 (0.0255)	0.015	1530	-0.0928 (0.0330)	0.005*	2315
Deworming drugs (last child)	0.0121 (0.0294)	0.680	1682	0.0036 (0.0292)	0.903	1680	-0.0063 (0.0472)	0.894	2614
Deworming drugs (other child)	0.0272 (0.0364)	0.455	1124	0.0295 (0.0373)	0.429	1122			

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates. Stars mark p-values which remain significant even under multiple testing adjustment.

Overall, effects on indicators under this hypothesis appear inconsistent (table 7.32). While some were consistent across specifications, the effects sometimes went in opposite directions for the last born child and other children. Moreover, effect sizes were small in most cases. One exception was the intention-to-treat effect on the probability that the last born child sleeps under a treated bednet. Treatment seems to decrease this probability significantly, by 5.8 to 9.3 percentage points. In DiD models this even held when multiple testing was accounted for. The negative sign was opposite of what was expected but confirmed in the subsample of SHG members and many subgroups. In the full sample, there was a marginally significant effect on the probability that the last born child sleeps under any bednet in the third specification, but coefficient signs were the opposite in specifications 1 and 2. While effects were positive on the probability of giving other children deworming drugs, these were inconsistent for the same indicator about the last born child. In the subsample of active SHG members, all coefficients for this indicator were positive, but again insignificant.

Based on our results, we reject the hypothesis that Gram Varta encourages parents to prevent diseases in children.

Hypothesis 31

This section investigates the hypothesis that Gram Varta reduces risky consumption behaviour based on the household and pregnant women sample.

Table 7.33: Risky consumption behaviour, household sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
More than 2 cups of chai per day	-0.0058 (0.0158)	0.712	2312	-0.0134 (0.0149)	0.366	2309	-0.0155 (0.0350)	0.659	4353

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates. Stars mark p-values which remain significant even under multiple testing adjustment.

In the household sample, we found the expected coefficient signs for the probability of drinking more than two cups of caffeinated tea per day, however, the effect was insignificant in all specifications (table 7.33).

Table 7.34: Risky consumption behaviour, pregnant women sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N

Smokes	-0.0155	0.483	1,561	-0.0141	0.551	1,488	-0.0754	0.169	2,174
	(0.0221)			(0.0236)			(0.0546)		
Frequency of smoking									
Less than one cigarette	0.0294	0.722	208	-0.0004	0.885	1,53	-0.7333	0.082	330
	(0.0825)			(0.0024)			(0.4172)		
1 to 5 cigarettes	-0.0197	0.718	208	-0.0020	0.884	1,53			
	(0.0547)			(0.0137)					
6 to 10 cigarettes	-0.0065	0.729	208	-0.0032	0.883	1,53			
	(0.0187)			(0.0215)					
More than 10 cigarettes	-0.0032	0.735	208	0.0044	0.884	1,53			
	(0.0094)			(0.0298)					
Chews tobacco	-0.0476	0.158	1,558	-0.0494	0.145	1,486	0.0214	0.753	2,169
	(0.0337)			(0.0339)			(0.0680)		
Frequency of chewing tobacco									
Once a day	0.0330	0.243	610	0.0236	0.437	578	-0.1158	0.519	889
	(0.0283)			(0.0304)			(0.1789)		
1 to 5 times a day	-0.0065	0.394	610	-0.0028	0.607	578			
	(0.0076)			(0.0055)					
6 to 10 times a day	-0.0235	0.237	610	-0.0183	0.427	578			
	(0.0199)			(0.0230)					
More than 10 times a day	-0.0030	0.317	610	-0.0025	0.473	578			
	(0.0030)			(0.0035)					
More than 2 cups coffee/chai	0.0410	0.114	1,59	0.0426	0.103	1,519	0.0051	0.898	3,583
	(0.0259)			(0.0261)			(0.0394)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

In the pregnant women sample, Gram Varta had a negative but insignificant impact on the woman's husband's incidence and frequency of smoking cigarettes (table 7.34). Across all specifications Gram Varta reduced the incidence of smoking. Regarding the frequency of smoking, the ordered logit estimations of specifications 1 and 2 showed a positive or no effect on the probability of smoking less than one cigarette a day and negative although small effects on the probabilities of smoking more cigarettes in a day. In line with this, the difference-in-differences estimate showed that Gram Varta reduced the frequency of smoking significantly at the 10 percent significance level. However, when looking at the sample of active SHG members we could not confirm this finding. The effect on the probability of smoking changed sign across specifications and we found no consistent pattern for the frequency of smoking (while the number of observations in the difference-in-differences specification was too small to estimate the effect).

Regarding chewing tobacco, the logit and ordered logit estimations showed a reduction in the probability and frequency of chewing tobacco, whereas the difference-in-differences estimates showed a positive effect on chewing tobacco and a negative effect on the frequency of chewing

tobacco. None of the effects were significant. When looking at the sample of active SHG members, we observed that the probability of chewing tobacco was consistently reduced across specifications, however, the effect on the frequency was inconsistent.

Finally, Gram Varta increased the husband's probability of drinking more than two cups of caffeinated coffee or tea per day but the effect was not significant at conventional levels and small when considering the difference-in-differences estimate. This was confirmed by the active SHG member analysis.

While some of our results were in the expected direction and were weakly significant, the findings are not conclusive, not allowing us to confirm that Gram Varta reduces risky consumption behaviour. This is especially the case as multiple testing adjustment turned all observed effects insignificant.

Hypothesis 32

This section investigates the hypothesis that Gram Varta improved domestic storage and treatment of water based on the household and pregnant women sample.

Table 7.35: Domestic storage and treatment of water, household sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Any water treatment	0.0136	0.087	3572	0.0118	0.138	3566	-0.0036	0.702	7521
	(0.0080)			(0.0080)			(0.0095)		
Adequate water treatment	0.0159	0.024	3572	0.0143	0.048	3566	0.0048	0.557	7521
	(0.0070)			(0.0072)			(0.0081)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

In the first specification, a marginally significant effect was found on the probability that the household engages in any water treatment (table 7.35). Quantitatively, this effect was small, although positive, implying an increase in 1.4 percentage points. Significant effects were found in the first two model specifications on the probability of engaging in adequate water treatment. Quantitatively, this effect was also small at less than 2 percentage points and not robust to the alternative difference-in-differences specification. The effects were however not robust to multiple testing adjustment and were not confirmed either when looking at the subsample of households with SHG members.

When we split the full sample according to educational levels of the woman, results clearly differed across the three groups. While coefficients were around zero and insignificant for the groups with low and medium education level, they were positive and significant at the 10 or 5 percent level for the group in which the woman completed secondary school or has higher qualification completed. For this subsample, the difference-in-differences specification suggested an increase of 7 percentage points in the probability that the household treats water adequately.

Table 7.36: Domestic storage and treatment of water, pregnant women sample

	Model 1			Model 2		
	B/SE	P value	N	B/SE	P value	N
Treats water adequately	0.0673 (0.0672)	0.316	61	-0.0558 (0.0984)	0.571	24

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values and Model 2 further adds controls. Model 3 could not be estimated as the sample is limited to adolescent girls.

Only 60 and 61 households in the pregnant women sample at baseline and endline, respectively, indicated that they treat their water in any way to make it safer to drink. Based on such a small sample we were not able to estimate our difference-in-differences specification. The logit estimations in table 7.36 show no clear impact of Gram Varta on treating drinking water adequately, however, due to the small sample size we do not intend to interpret these results. Also the analysis for active SHG members was not feasible due to the small sample size.

Overall, we do not find sufficient support for the hypothesis that Gram Varta improved domestic storage and treatment of water.

Hypothesis 33

This section investigates the hypothesis that Gram Varta improves domestic hygiene, including hand-washing and use of toilets, based on the household and pregnant women samples.

Table 7.37: Domestic hygiene, household sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Buying soap									

Never or if provided	0.0270	0.184	3511	0.0362	0.064	3505	-0.1067	0.118	7450
	(0.0203)			(0.0196)			(0.0680)		
More than 2 months	0.0008	0.193	3511	0.001	0.080	3505			
	(0.00069			(0.0006)					
Every 1 to 2 months	0.0005	0.408	3511	0.0004	0.511	3505			
	(0.0006)			(0.0007)					
Once a week	-0.0283	0.184	3511	-0.0377	0.065	3505			
	(0.0213)			(0.0204)					
Improved toilet	0.0107	0.580	3448	-0.0046	0.779	3442	-0.0215	0.092	7401
	(0.0194)			(0.0164)			(0.0127)		
Cleanliness of toilet									
Very dirty	-0.0002	0.986	694	0.0026	0.844	694	-0.0532	0.572	1230
	(0.0129)			(0.0130)			(0.0938)		
Dirty	-0.0005	0.986	694	0.0054	0.843	694			
	(0.0280)			(0.0272)					
Clean	0.0005	0.986	694	-0.0059	0.843	694			
	(0.0304)			(0.0297)					
Very clean	0.0002	0.986	694	-0.0021	0.842	694			
	(0.0105)			(0.0106)					
Stool piles	-0.0237	0.477	3574	-0.0191	0.563	3568	0.0408	0.350	7502
	(0.0332)			(0.0331)			(0.0435)		
Sewage water	-0.0138	0.643	3574	-0.0123	0.680	3568	0.0279	0.433	7497
	(0.0298)			(0.0298)			(0.0355)		
Infant stool disposal	0.0261	0.580	634	0.0158	0.733	634	-0.0194	0.728	3872
	(0.0473)			(0.0463)			(0.0557)		
Practice of ODF	-0.0567	0.035	662	-0.0563	0.033	662	0.0827	0.102	1198
	(0.0269)			(0.0264)			(0.0503)		
Reprehend ODF	-0.0027	0.924	3555	-0.0058	0.836	3549	-0.0205	0.587	7477
	(0.0289)			(0.0279)			(0.0377)		
ODF is health hazard	0.0125	0.387	3418	0.0111	0.449	3412	0.0371	0.044	7278
	(0.0145)			(0.0146)			(0.0183)		
Use toilet or cover with mud	0.0114	0.413	3350	0.0076	0.579	3344	-0.0231	0.296	7096
	(0.0139)			(0.0138)			(0.0221)		
Use soap after toilet	-0.0101	0.700	3577	-0.0217	0.403	3571	-0.0039	0.893	7530
	(0.0263)			(0.0259)			(0.0292)		
Use soap before meal	-0.0136	0.522	3577	-0.0187	0.361	3571	-0.0147	0.641	7530
	(0.0213)			(0.0204)			(0.0315)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

In the household sample, treatment did not seem to improve the frequency of buying hand washing soap (table 7.37). In the second specification, there were marginally significant effects on three of the four possible values of the indicator. However, they went in the opposite direction with high frequency of buying soap decreasing and lower frequency of buying soap increasing. Also contrary to our hypothesis are the negative effects on the use of soap, both after using the toilet and before a meal. According to specifications 1 and 2, treatment decreased the practice of open defecation at a significance level of 5 percent by about 6 percentage points. The effect changed its sign in the third specification, however, and lost significance. In the difference-in-differences specification, the probability increased that open defecation is seen as health hazard in treatment areas by 3.7 percentage points. While the sign on this indicator was the same for the other two specifications, the coefficient was insignificant. It appears contradictory that despite this positive effect, coefficients for whether the household reprehends open defecation were negative in all specifications. In the subsample of households with an active SHG member, effects on the frequency of buying soap and using soap were similar. However, effects on the practice of open defecation, whether open defecation is reprehended or seen as a health hazard were inconsistent.

Table 7.38: Domestic hygiene, pregnant women sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Using a toilet or cover excreta with mud is important	-0.0331	0.222	1,554	-0.0396	0.133	1,481	-0.0979	0.034	3,322
	(0.0271)			(0.0264)			(0.0458)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

Against prior expectations, Gram Varta decreased the probability of having the opinion that using a toilet or covering excreta with mud is important in the pregnant women sample (table 7.38). The difference-in-differences specification showed a significant decrease in the probability by almost 10 percentage points, where 83 percent of the control group think that using a toilet or covering excreta with mud is important. The results changed when considering active SHG members only. Here, the probability of using a toilet or covering excreta with mud increased in the treatment group, although insignificantly.

In the subgroup analysis by block, we observed a robust and significant decrease in the probability of using a toilet or covering excreta with mud in Kumarkhand, while we found no robust pattern for the other blocks. Similarly, for women who had at least completed junior secondary school, we found significant adverse effects. This was also the case for both age groups considered. In the subgroup analysis by caste we found an increase in the probability to use a toilet or cover excreta with mud in scheduled tribes. However, it was significantly reduced by 14.5 percentage points (difference-in-differences) in the subgroup of other backward classes. Similarly, for both age groups, we found that Gram Varta significantly reduced the probability of using a toilet or covering excreta with mud at the 10 percent level for both age groups. This was also true for for women who had at least completed junior secondary school, at the 1 percent significance level.

Overall, we found rather adverse effects on domestic hygiene. In addition, none of the effects found related to hypothesis 33 remained significant when multiple testing was accounted for and we thus reject the hypothesis.

Hypothesis 34

This section investigates the hypothesis that Gram Varta increases women's awareness about infectious diseases such as malaria based on the household and pregnant women samples.

Table 7.39: Awareness about infectious diseases, household sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Seen/heard/read									
messages about malaria/dengue	0.0341	0.090	2988	0.0238	0.235	2984	-0.0244	0.504	4719
	(0.0201)			(0.0200)			(0.0365)		
messages about diarrhoea	-0.0134	0.479	2996	-0.0229	0.227	2992	-0.0879	0.010*	4718
	(0.0190)			(0.0189)			(0.0340)		
messages about ARI	0.0288	0.161	2945	0.0158	0.443	2941	-0.0317	0.349	4679
	(0.0205)			(0.0206)			(0.0338)		
messages about STI	0.009	0.650	2757	0.001	0.960	2753	-0.0461	0.120	4359
	(0.0198)			(0.0192)			(0.0295)		
Knows danger signs of malaria	0.0067	0.751	3153	-0.0016	0.937	3149	-0.0402	0.256	5003
	(0.0211)			(0.0208)			(0.0353)		
Knows danger signs of ARI	-0.0064	0.777	3153	-0.0147	0.507	3149	-0.0796	0.010*	5003
	(0.0225)			(0.0222)			(0.0305)		
Knows treatment of diarrhoea	-0.0068	0.743	3153	-0.0181	0.369	3149	-0.0629	0.045	5003
	(0.0208)			(0.0201)			(0.0311)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates. Stars mark p-values which remain significant even under multiple testing adjustment.

In the household sample, according to specification 1, treatment increased the probability that the woman saw, heard or read messages about the malaria/dengue programme at a significance level of 10 percent (table 7.39). This did not hold for the other two specifications. Gram Varta appeared to decrease the probability that the woman noticed messages about diarrhoea, knows any danger signs of acute respiratory infections (ARI) or knows how to treat diarrhoea adequately. These negative effects were significant at the 5 percent level in the difference-in-differences specification and remained significant when multiple testing was accounted for. However, they were opposite of what would be expected. Moreover, they were confirmed in the analysis of the subsample with SHG members only. Effects were sizable at 10.6 and 12.2 percentage points respectively. However, these effects differed considerably across geographic blocks. In two blocks, we found only negative effects for all indicators, whereas in the four other blocks, intention-to-treat effects were consistent across specifications and positive for some indicators.

Table 7.40: Awareness about infectious diseases, pregnant women sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Seen/ heard/ read messages about malaria/ dengue/ chikungunya	0.0277	0.347	1,554	0.0193	0.520	1,485	-0.0574	0.179	3,494
	(0.0295)			(0.0300)			(0.0425)		
Knows that danger signs of malaria are:									
Fever	0.0230	0.459	1,612	0.0276	0.383	1,538	-0.0018	0.965	3,61
	(0.0310)			(0.0316)			(0.0412)		
Chills	0.0269	0.377	1,612	0.0263	0.400	1,538	-0.0417	0.368	3,61
	(0.0305)			(0.0312)			(0.0462)		
Sweat	0.0280	0.048	1,612	0.0285	0.078	1,538	0.0291	0.060	3,61
	(0.0141)			(0.0162)			(0.0154)		
Headache	0.0031	0.780	1,612	0.0070	0.547	1,538	0.0660	0.064	3,61
	(0.0112)			(0.0116)			(0.0353)		
Vomit	0.0163	0.211	1,612	0.0166	0.215	1,538	0.0139	0.550	3,61
	(0.0130)			(0.0134)			(0.0232)		
Loose motions	-0.0119	0.197	1,612	-0.0126	0.279	1,479	-0.0296	0.036	3,61
	(0.0093)			(0.0116)			(0.0140)		
Knows any danger sign of malaria	0.0302	0.340	1,612	0.0333	0.303	1,538	0.0010	0.980	3,61
	(0.0317)			(0.0323)			(0.0414)		
Knows two danger signs of malaria	0.0532	0.066	1,944	0.0487	0.084	1,853	0.0692	0.155	3,944
	(0.0290)			(0.0282)			(0.0484)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates. Stars mark p-values which remain significant even under multiple testing adjustment.

In the pregnant women sample, only indicators about malaria were included (table 7.40). While Gram Varta significantly increased the probability of knowing the danger signs sweat and headache, it significantly reduced the probability of knowing the danger sign of loose motions at the 5 percent significance level. The composite indices (knowing any or two danger signs of malaria) showed positive effects and were significant for knowing two danger signs in specifications 1 and 2. When controlling for individual and time fixed effects in the difference-in-differences specification, the effects were no longer significant. The remaining coefficients showed mixed and insignificant effects on having noticed the messages and knowing the danger signs chills and sweat. In addition to these inconsistencies, all effects were rendered insiginificant when multiple testing was adjusted for. Moreover, the active SHG member analysis confirmed that there existed no consistent pattern in having noticed the malaria messages or knowing the danger signs of malaria.

Based on the results from the household and pregnant women samples, we reject the hypothesis that Gram Varta increases women's awareness about infectious diseases such as malaria.

Hypothesis 35

This section investigates the hypothesis that Gram Varta improves adolescent girls' and women's knowledge about sexuality and contraception based on the household sample.

Table 7.41: Sexuality and contraception, household sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Number family planning methods used by women	-0.0774	0.305	3153	-0.1087	0.142	3149	0.6261	0.001*	4757
	(0.0752)			(0.0737)			(0.1873)		
Women currently uses contraception	0.0135	0.526	3090	0.0056	0.786	3086	0.0871	0.005*	4712
	(0.0213)			(0.0208)			(0.0306)		
Girl knows where to get									
Pill	-0.0315	0.610	331	-0.0421	0.487	331			
	(0.0618)			(0.0606)					
Emergency contraception	0.0042	0.942	307	-0.0111	0.838	307			
	(0.0577)			(0.0541)					
Condom/nirodh	0.0334	0.477	432	0.0175	0.697	432			
	(0.0470)			(0.0449)					
Female condom	0.0526	0.133	381	0.0451	0.201	381			
	(0.0350)			(0.0353)					
Girl believes									
Condoms are effective contraception	0.223	0.043	64	0.2242	0.030	64			
	(0.1103)			(0.1030)					
Condoms can be used more than once	0.0094	0.939	59	0.0368	0.766	59			
	(0.1224)			(0.1239)					
Okay to suggest condom to partner	0.1857	0.053	74	0.2699	0.006	74			
	(0.0960)			(0.0980)					
Condoms are effective HIV/AIDS protection	0.1635	0.087	64	0.2263	0.015	64			
	(0.0957)			(0.0931)					
Feel embarrassed to buy condoms	0.0798	0.495	71	0.0724	0.558	71			
	(0.1171)			(0.1235)					
	0.2125	0.055	53	0.1852	0.090	53			

Suggesting condoms means no trust in partner	(0.1108)			(0.1093)		
Condoms can slip off and disappear in body	0.0411	0.731	67	0.119	0.272	67
	(0.1193)			(0.1083)		
Condoms effective against STD	-0.0007	0.995	76	0.0281	0.810	76
	(0.1077)			(0.1167)		
Discussed contraception	-0.1075	0.048	502	-0.1071	0.046	502
	(0.0544)			(0.0537)		
Girl believes						
Pregnancy after kissing	-0.0048	0.847	408	-0.0073	0.779	408
	(0.0250)			(0.0259)		
Woman has to bleed at first intercourse	0.0561	0.296	275	0.0637	0.229	275
	(0.0536)			(0.0530)		
Pregnancy possible at first time	0.0232	0.702	273	0.0269	0.662	273
	(0.0606)			(0.0616)		
Woman stops growing after first time	0.0328	0.537	244	0.0358	0.490	244
	(0.0530)			(0.0518)		
Highest pregnancy risk between periods	0.0608	0.362	228	0.0609	0.331	228
	(0.0666)			(0.0626)		
Use of tampons/female pads	0.0398	0.397	645	0.0297	0.471	645
	(0.0470)			(0.0411)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates. Stars mark p-values which remain significant even under multiple testing adjustment.

Using the difference-in-differences specification, treatment has a highly significant positive effect on the number of family planning methods known by the woman and whether the woman currently uses any contraception (table 7.41). Notably, these effects remained significant even when multiple testing was accounted for. The coefficient sizes suggested an increase in the number of known family planning methods by less than one and an increase in use of contraception by 8.7 percentage points. However, only the second of these effects was consistent across specifications. Similarly in the analysis of SHG members, treatment effects on these two indicators were highly significant in the difference-in-differences specification and with the expected sign.

For the adolescent girl data, observation numbers were low for the questions on knowledge about condoms since few answered that they knew what a condom was. Low observation numbers made analyses impossible for several of the subgroups. Nevertheless, a few significant effects were found in the full sample. Treatment significantly increased the probability that an adolescent girl believes condoms are effective contraception. The same held for the statement that it is okay to suggest condom use to the partner and that condoms

are effective protection against HIV/AIDS. Here, slight differences in significance appeared between specifications 1 and 2. Treatment also increased the probability that the girl says suggesting condom use shows a lack of trust in the partner, at a significance level of 10 percent. The direction of the effect was opposite of what was expected. Also contrary to expected effect was the significant negative effect on whether the girl discussed contraception with anyone. This negative effect, however, did not appear as significant in the subsample of girls in households with an active SHG member. Moreover, possibly due to the low number of observations, none of these effects on adolescent girls remained significant with the Benjamini-Hochberg correction.

Overall, the presented evidence leads us to reject our hypothesis that Gram Varta improves adolescent girls' and women's knowledge about sexuality and contraception.

Summary

Hypothesis 27 about the child's consumption of micronutrients showed some consistent effects, but opposite to the hypothesis and overall not significant. Hypothesis 28 and 30 could not be confirmed. Gram Varta seemed to affect knowledge and attitudes toward proper feeding of newborns (Hypothesis 29) as expected for some indicators, but these results were not consistent and especially not in the pregnant woman sample. Gram Varta had a negative impact on the woman's husband's incidence and frequency of smoking cigarettes. However, when looking at the sample of active SHG members we could not confirm this finding. Effects on domestic storage and treatment of water were inconsistent, as were those on domestic hygiene. For indicators under the latter hypothesis we even found coefficients that were often in the direction opposite to our hypothesis. Hypothesis 34 on women's awareness about infectious diseases was also not confirmed, with unexpected signs for the main Gram Varta sample and inconsistent findings in the pregnant women sample. Gram Varta does seem to increase women's knowledge about contraception methods and their use, but did not consistently improve girls' knowledge about contraception.

7.5 Results related to pregnancy

Hypothesis 36

This section investigates the hypothesis that Gram Varta encourages pregnant women to be mindful of their health, to take healthy and sufficient diet, to avoid stress and to avoid health risks based on the pregnant women sample. The ToC assumptions are the same as those identified in the previous section. However, these hypotheses are related specifically to the health of pregnant women, an important target group of Gram Varta.

Table 7.42: Health risks, pregnant women sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N

Meals per day	0.0644 (0.0412)	0.121	1602	0.0679 (0.0415)	0.104	1529	0.1382 (0.0547)	0.013	3597
Child sleeps under bednet	0.0111 (0.0153)	0.471	1437	0.0143 (0.0163)	0.381	1288			
Bednet is insecticide-treated	0.0099 (0.0165)	0.548	1262	0.0097 (0.0157)	0.537	1206			

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates. Stars mark p-values which remain significant even under multiple testing adjustment.

Gram Varta increased the number of meals per day eaten by the woman by 0.06 to 0.13 meals and the coefficient was significant at conventional levels in the difference-in-differences specification (table 7.42). It should be noted, however, that this effect became insignificant once multiple testing was accounted for. The indicator “woman sleeps under a bednet” and “the bednet is insecticide-treated” changed to “whether the child sleeps under a bednet,” thus we could not run a difference-in-differences estimation on this variable. Specifications 1 and 2 showed positive but small and insignificant effects on the bednet indicators. In the subgroup analysis by education, we found a significantly positive effect on the number of meals eaten per day, while the coefficients on the use of bed nets pointed in the right direction, the coefficients were not significant.

Overall, we found weak evidence in support of the hypothesis that Gram Varta encourages pregnant women to be mindful of their health, to take healthy and sufficient diet, to avoid stress and to avoid health risks.

Hypothesis 37

This section investigates the hypothesis that Gram Varta increases the frequency of antenatal care visits as well as their quality based on the pregnant women sample. We split results related to hypothesis 37 into three different tables, where the first presents results on antenatal care use, the second on advice and information received during antenatal antenatal care and the third on measures taken during antenatal care.

Table 7.43: Antenatal care visit, pregnant women sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Saw someone for ANC	0.0152 (0.0347)	0.661	1574	0.0129 (0.0341)	0.705	1502	-0.1166 (0.0501)	0.021	3571
ANC by skilled health personnel	0.0124 (0.0355)	0.728	1561	0.0090 (0.0347)	0.796	1487	0.0535 (0.0515)	0.301	2724
	-0.0296	0.187	780	-0.0416	0.085	681	-0.0299	0.387	2008

ANC in health facility	(0.0224)			(0.0241)			(0.0344)		
Number of ANC visits	-0.3034	0.080	708	-0.3152	0.058	673	-0.7311	0.001*	1918
	(0.1721)			(0.1648)			(0.2245)		
Months until first ANC	-0.1889	0.256	700	-0.1962	0.230	667	-0.0392	0.862	1894
	(0.1656)			(0.1627)			(0.2244)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates. Stars mark p-values which remain significant even under multiple testing adjustment.

Table 7.43 shows rather adverse effects of Gram Varta on antenatal care indicators. In all three specifications, the frequency of antenatal care visits dropped significantly and the result from the DiD analysis remained weakly significant when multiple testing was accounted for. This is contradicting the programme's intentions. Although the effects were not significant, it is puzzling that we observed a positive coefficient on the probability of seeing skilled health personnel for antenatal care but a negative coefiicients for the probability of getting antenatal care at a health facility (given antenatal care is received) as skilled health personnel usually work at health facilities. Perhaps this is explained by one of the findings of our qualitative study. The participation of health staff in Gram Varta meetings as well as home visits by Anganwadi workers ensures that pregnant women are in contact with the health staff. However, our qualitative studies suggest that women's mobility in villages is restricted. Thus, even though women do not go to health facility to receive care, they have contact with the health workers through their home visits.

Looking at the sample of active SHG members, we found positive, insignificant effects on the probability of seeking antenatal care and the number of antenatal care visits. However, the effects on the probability of seeing skilled health personnel, going to a health facility as well as months until first antenatal care remained difficult to interpret.

Table 7.44 informs about the impact of Gram Varta on having received advice or information on various aspects during antenatal care.[8] These indicators reflect the awareness for quality of antenatal care induced by Gram Varta. Women in the treatment group are expected to remember better what information they did receive when this was also discussed during Gram Varta SHG sessions.

Table 7.44: Information received at antenatal care visit, pregnant women sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
During antenatal care woman received advice on									
	-0.0320	0.488	709	-0.0559	0.234	675			

[8] Note that the variable "Woman received advice on postpartum depression" was not included at baseline.

Post partum depression	(0.0461)			(0.0469)					
Work during pregnancy	0.0383	0.365	722	0.0072	0.868	688	-0.0139	0.816	1950
	(0.0422)			(0.0436)			(0.0597)		
Breastfeeding	0.0168	0.733	717	0.0035	0.947	684	0.0971	0.303	1945
	(0.0492)			(0.0526)			(0.0939)		
Keeping the baby warm	0.0441	0.269	708	0.0401	0.324	674	0.0772	0.209	1936
	(0.0398)			(0.0407)			(0.0611)		
Cleanliness at delivery	-0.0027	0.932	714	0.0030	0.928	680	0.0281	0.682	1942
	(0.0323)			(0.0335)			(0.0685)		
Family planning for spacing	-0.0338	0.505	676	-0.0166	0.738	643	0.0203	0.765	1904
	(0.0507)			(0.0498)			(0.0679)		
Family planning for limiting	-0.0013	0.978	686	-0.0107	0.828	653	0.0393	0.543	1914
	(0.0466)			(0.0494)			(0.0645)		
Maternal nutrition	0.0112	0.604	733	-0.0009	0.968	699	0.0709	0.263	1961
	(0.0217)			(0.0223)			(0.0631)		
Child nutrition	0.0315	0.345	704	0.0209	0.532	672	-0.0480	0.407	1932
	(0.0333)			(0.0335)			(0.0576)		
Need for institutional delivery	-0.0164	0.736	677	-0.0460	0.373	644	-0.0423	0.625	1905
	(0.0486)			(0.0516)			(0.0862)		
Danger of malaria infection	0.0098	0.863	664	0.0206	0.727	632	0.6235	0.369	1892
	(0.0568)			(0.0589)			(0.6925)		
During antenatal care woman was informed about									
Bleeding	0.0027	0.956	726	-0.0042	0.936	693	0.0304	0.703	1954
	(0.0488)			(0.0516)			(0.0795)		
Convulsions	-0.0345	0.478	722	-0.0364	0.456	688	-0.0090	0.897	1950
	(0.0487)			(0.0489)			(0.0695)		
Prolonged labour	-0.0409	0.394	725	-0.0497	0.318	691	0.0446	0.530	1953
	(0.0479)			(0.0498)			(0.0709)		
Preterm labour	0.0122	0.791	712	-0.0062	0.903	679	0.0900	0.214	1940
	(0.0461)			(0.0511)			(0.0721)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

Across all three specifications there was an increase in receiving advice on breastfeeding, keeping the baby warm and the danger of a malaria infection while there was a decrease in receiving advice on the need for institutional delivery and information on convulsions (table 7.44). However, none of the coefficients were significant.

For the sample of active SHG members, the difference-in-differences specification showed a negative impact on advice received for all topics except for advice about keeping the baby warm. However, only the topics breastfeeding, family planning for spacing, family planning for limiting and maternal nutrition were consistently negative across specifications whereof the effect on the probability of receiving advice on maternal nutrition was the only significant one.

Turning to the probability of having received certain information, information about convulsions was consistently negative across specifications as it was in the main analysis. Additionally, the effect on the probability of receiving information on bleeding was consistently negative while it was positive for information about preterm labour.

Taken together, no clear pattern emerged for having received advice or information on different aspects of pregnancy.

Gram Varta was expected to induce women to specifically pay attention to whether measures were taken during antenatal care in comparison to women in the control group. Table 7.45 presents different measures taken during antenatal care. It appears that women in treatment groups reported less often having had their blood pressure taken, weight measured, a urine sample, a blood sample or an ultrasound, at least when looking at specifications 2 and 3. However, none of the effects were significant, indicating that Gram Varta did not have the expected effect. Similarly, in the sample of active SHG members we found mostly insignificant effects on the probability of getting different measures taken during pregnancy.

Table 7.45: Measures taken at antenatal care visit, pregnant women sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
BP measured at ANC	0.0024	0.948	741	-0.0184	0.626	708	-0.0248	0.670	1967
	(0.0374)			(0.0378)			(0.0582)		
Weight measured at ANC	0.0049	0.908	742	-0.0081	0.840	709	-0.0400	0.532	1,97
	(0.0424)			(0.0404)			(0.0638)		
Urine taken at ANC	0.0032	0.944	741	-0.0271	0.552	709	-0.0266	0.704	1968
	(0.0457)			(0.0455)			(0.0699)		
Blood taken at ANC	-0.0168	0.691	741	-0.0404	0.324	709	-0.0055	0.932	1968
	(0.0423)			(0.0409)			(0.0646)		
Ultrasound during ANC	-0.0013	0.978	740	-0.0336	0.490	708	-0.0295	0.534	2728
	(0.0483)			(0.0486)			(0.0473)		
Tetanus injection during pregnancy	-0.0118	0.523	1427	-0.0117	0.573	1279	-0.0773	0.035	3424
	(0.0185)			(0.0207)			(0.0362)		
Numner of tetanus injections	0.0378	0.284	1299	0.0344	0.389	1235	-0.0558	0.299	2759
	(0.0351)			(0.0398)			(0.0536)		
Number of iron & folic acid tablets	0.2720	0.898	992	-0.2365	0.909	948	-6.3067	0.046	2813
	(2.1099)			(2.0712)			(3.1261)		
Number of iron & folic acid bottles	-0.0507	0.933	817	0.0851	0.894	783	-0.3160	0.637	2626
	(0.5977)			(0.6384)			(0.6680)		
Days of iron & folic acid intake	-2.0942	0.507	531	-2.4500	0.431	507	-12.345	0.061	854
	(3.1494)			(3.1023)			(6.5414)		
Other nutritional supplements during pregnancy	0.0025	0.957	1389	-0.0027	0.953	1326	-0.0138	0.784	3353
	(0.0452)			(0.0466)			(0.0503)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

Surprisingly, Gram Varta showed mixed or adverse effects on the probability of having received a tetanus injection, the number of tetanus injections, the number of iron and folic acid tablets or bottles, the days of iron and folic acid intake, as well as the intake of other nutritional supplements during pregnancy. The effects on the probability of receiving a tetanus injection and the days of iron and folic acid intake were significant at conventional levels, as long as no multiple testing correction was performed. Similarly, in the sample of active SHG members we found mixed and rather adverse effects on tetanus injections and nutritional supplements.

Taken together, our results lead us to reject our hypothesis that Gram Varta increases the frequency of antenatal care visits as well as improves their quality.

Hypothesis 38

This section investigates the hypothesis that Gram Varta increases support to pregnant women for obtaining antenatal care based on the pregnant women sample.

Table 7.46: Support for antenatal care, pregnant women sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Accompanied to antenatal care by									
anyone	-0.0361	0.289	744	-0.0582	0.097	711	-0.1143	0.049	1972
	(0.0340)			(0.0351)			(0.0576)		
husband	0.0606	0.258	621	0.0474	0.376	594	-0.0466	0.604	1413
	(0.0535)			(0.0535)			(0.0897)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

Table 7.46 shows the results related to the family support received by the woman when going to antenatal care. Gram Varta significantly reduced the probability of being accompanied to antenatal care by any household member at the 5 percent significance level (specification 3), although this result was rendered insiginificant by the Benjamini-Hochberg correction. Looking at the difference-in-differences estimation, Gram Varta reduced the probability of being accompanied to antenatal care by any family member by 11.4 percentage points.

This was confirmed in the SHG member sample with a consistently negative effect across specifications and also significant in case of the difference-in-differences effect. The effect on the probability of being accompanied by the husband was not consistent across specifications and insignificant in the main and the SHG member analysis. In the block Gwalpara Gram Varta significantly reduced the probability of being accompanied to antenatal care by anyone or the

husband. Similarly, in the subgroup analysis by education we found a clear pattern that Gram Varta reduced the probability of being accompanied to antenatal care by anyone or the husband for women with no education completed. However, the effect was only significant for being accompanied by anyone in specification 2.

Overall, we reject the hypothesis that Gram Varta increases support to pregnant women for obtaining antenatal care.

Hypothesis 39

This section investigates the hypothesis that Gram Varta increases pregnant women's satisfaction with antenatal care based on the pregnant women sample.

In theory Gram Varta may create awareness of quality thereby decreasing satisfaction with bad antenatal care but potentially increasing it with respect to good antenatal care (table 7.47). However, the effect of Gram Varta on satisfaction with antenatal care was inconsistent across specifications. We observed similar mixed results in the sample of active SHG members.

In the subgroup analysis by caste, we found a positive impact of Gram Varta on antenatal care satisfaction for the scheduled caste subgroup, which was robust across specifications and significant in the difference-in-differences estimation. However, the presented evidence does not allow us to unequivocally confirm the hypothesis that Gram Varta increases pregnant women's satisfaction with antenatal care.

Table 7.47: Satisfaction with antenatal care, pregnant women sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Satisfied with antenatal care	-0.0169 (0.0250)	0.500	751	-0.0165 (0.0269)	0.540	717	0.0589 (0.0377)	0.120	1938

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

Hypothesis 40

This section investigates the hypothesis that Gram Varta encourages women to accept their pregnancy, making them more optimistic about their situation based on the pregnant women sample.

Table 7.48: Acceptance with pregnancy/life, pregnant women sample

	Model 1	Model 2	Model 3

	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
How often stressed or strongly worried in the past 12 months									
Never	0.0242	0.047	1596	0.0237	0.053	1524	-0.0936	0.038	3557
	(0.0122)			(0.0123)			(0.0446)		
Sometimes	0.0133	0.095	1596	0.0131	0.100	1524			
	(0.0080)			(0.0080)					
Often	-0.0376	0.053	1596	-0.0368	0.059	1524			
	(0.0194)			(0.0195)					
Satisfied with family life	0.0389	0.270	1499	0.0355	0.325	1436			
	(0.0353)			(0.0361)					

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

Table 7.48 shows that Gram Varta successfully reduced the frequency of feeling stressed during the past 12 months. While specification 3 showed a significant reduction in the frequency of being stressed, the ordered logit estimation in specifications 1 and 2 showed a significant increase in never feeling stressed and a significant decrease in feeling stressed often, while there was also a significantly positive but smaller effect on the probability of feeling stressed sometimes.

We only have information about being satisfied with family life at endline. The effect was positive although insignificant in specifications 1 and 2. These findings were confirmed in the SHG member analysis, however, the coefficients were not significant. In the subgroup analysis by age, we found that Gram Varta significantly reduced the frequency of feeling worried. This effect was robust and significant across all specifications.

Overall, our results suggest that Gram Varta encourages women to accept their pregnancy, making them more optimistic about their situation. It should be noted, however, that the significance of estimates disappeared once multiple testing is accounted for. The evidence in support of the hypothesis is therefore only weak.

Summary

Overall the evidence for pregnancy related outcomes is mixed. We found weak evidence in support of the hypothesis that Gram Varta encourages pregnant women to be mindful of their health, to take healthy and sufficient diet, to avoid stress and to avoid health risks (Hypothesis 36), which was mostly driven by an increase in the number of meals women eat per day. We found mixed and inconclusive results about women's frequency of antenatal care visits as well as their quality (Hypothesis 37), thus we reject this hypothesis. Contrary to expectations, our results suggest that Gram Varta reduced the probability of pregnant women receiving support for obtaining antenatal care and therefore we reject hypothesis 38. While we found some evidence that Gram Varta increased pregnant women's satisfaction with antenatal care (Hypothesis 39), it was not sufficient to unequivocally confirm our hypothesis. Finally, our results suggest that Gram Varta encouraged women to accept their pregnancy, making them

more optimistic about their situation (Hypothesis 40). This was mostly driven by a significant reduction in the frequency of feeling stressed or strongly worried over the past 12 months. With that said, overall these results are vulnerable to multiple testing adjustment and should therefore be interpreted with caution.

7.6 Results related to Anganwadi centers

This set of hypothesis evaluates the effects of Gram Varta on quality and utilization of Anganwadi services. While results based on the Anganwadi survey should not be taken as causal evidence due to the lack of a control group, they provide valuable information to put the main results into context. These hypotheses are related to the ToC assumptions that Gram Varta participants speak up to demand improvements in service provision in the community (Stage 4), and that HNWASH front line workers are more responsive and supportive (Stage 5).

Hypothesis 42

This investigates the hypothesis that Gram Varta increases the use of Anganwadi centers (AWC) based on the household sample.

In table 7.49 the coefficients are similar for the first two specifications and higher in absolute terms in the third. However, the sign of the coefficients are negative, opposite of what was expected, although not significant for any of the specifications. The treatment effect on this indicator seems similar for the subsample of SHG members.

Table 7.49: Use of Anganwadi centers, household sample

	Model 1			Model 2			Model 3		
	(1)	P value	N	(2)	P value	N	DiD	P value	N
Visits to Anganwadi center	-0.2258	0.296	3020	-0.24	0.262	3.016	-0.4947	0.179	4694
	(0.2156)			(0.2132)			(0.3667)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

We also addressed this hypothesis in our survey of AWWs. The average number of pregnant women registered by the AWW decreased by 1.4 (paired t test p value=0.004, N=195). Notably, the number of pregnant women who were registered because the AWW approached them decreased by 1.24 (paired t test p value=0.017, N=168). The number of pregnant women who approached the AWW themselves decreased by 1.66, but this was not statistically significant. Interestingly, the AWW reported that the number of pregnant women living in the area of her responsibility during the previous month increased by an average of 4.46 (paired t test p value=0.006, N=188). A similar pattern was observed with lactating mothers as well. The

AWWs reported that the number of children under age 2 in their registers decreased by about 2.26, the ones they themselves approached decreased by 1.60 on average, and that the number that were brought to the center out of the caregivers' interest increased by 4.52. However, none of these findings were statistically significant. Notably, the number of children in their area during the previous month who belonged in this age group reportedly increased by 18.79 (paired t test p value<0.001). The AWWs reported that on average, the number of adolescent girls (aged 10–19 years) in their register, the ones they approached themselves and the ones who approached the AWW on their own, all decreased. However, none of these changes were statistically significant. Similarly, they reported that on average, the number of adolescent girls who lived during the previous month in the area they were in-charge of increased, but this too was not statistically significant.

This general pattern of reported decrease in the number of beneficiaries registered, decrease being reported even among beneficiaries themselves approaching the AWW, and the perception that the number of target beneficiaries in her area is increasing over time does not support our hypothesis that Gram Varta caused an increase in the use of the AWCs. Our results from the household sample imply a decrease in the number of visits to the AWC, thus we reject our hypothesis.

Hypothesis 43

This section investigates the hypothesis that Gram Varta improves malnutrition treatment and prevention, using data from the AWW survey. As per the AWWs, the number of severely underweight children identified during the previous 12 months, in the area that they were responsible for, decreased by 6.15 on average. However, this finding was not significant. The corresponding figure for the previous month was reported to be an increase of about 0.82, again not significant. On average, about 8.19 fewer children were reported to have recovered from severe malnutrition during the previous 12 months after they started attending that particular AWC (paired t test p value=0.018). While the corresponding figure for the previous month was an average increase of 3.08, this was not significant. A notable finding was that the average number of days in the previous month when hot cooked meals were served increased by an average of 5.77 (paired t test p value<0.001). Corroborating this finding was the result that the average number of meals eaten at the center during the previous month by children aged three to six years increased by about 14.22 (paired t test p value<0.001).

Another effort to prevent undernutrition by the AWW is her discussion with expectant and nursing mothers about supplementary nutrition. 91 AWWs had answered "once a week" at baseline when asked how frequently they discuss supplementary nutrition with this group of mothers. However, 47 of them answered "once a week" at endline, with 37 of them reporting "once a month." Among the 84 AWWs who had said "once a month" at baseline, 36 reported discussing this topic "once a week." Almost all AWWs had answered yes, at both baseline and endline, when asked whether they provide mothers with locally available food recipes for supplementary nutrition of children in the six months to six years age group. This pattern of close to 100 percent "yes" at both points in time was also seen when they were asked whether they were able to provide all the items as per the menu of hot cooked meals and whether the take-home rations were provided as per norms.

Taken together, these findings suggest that Gram Varta improved the efforts to prevent malnutrition by providing hot cooked meals and ensuring that three to six year old children consume their meals at the AWC itself. However, we found no evidence that Gram Varta improves the treatment of malnutrition.

Hypothesis 44

This section investigates the hypothesis that Gram Varta improves the quality of child weighing practices.

We enquired about the types of weighing machines available at the AWC of the AWW interviewed in our survey. At baseline 143 out of 198 AWCs had Salter scales, 157 had tray scales, and 151 had platform (floor-based) scales. The remaining ones had no weighing scales. This changed to 139, 160 and 145 out of 198 at the time of the midline, and 30 AWCs were without any type of weighing scales at the endline.

At baseline, when asked which tasks they performed as an AWW regarding keeping track of the children's weight, 95 out of 193 mentioned “recording the weight of children aged 5 and below every month on growth charts for the mother and in AWC register”. 142 mentioned recording the weight of newborns and 52 mentioned “plotting of weight of children in a growth monitoring chart.” The corresponding figures at endline were 132, 129 and 65 of 193 respectively.

Based on these results, we do not find any evidence to support the hypothesis that Gram Varta improves the quality of child weighing practices.

Hypothesis 45

This section investigates the hypothesis that Gram Varta improves the cleanliness of and hygiene practices at the AWC. In response to the question whether they treat the water at the AWC to make it safer to drink, of the 185 of the total 201 who had answered yes at endline, 170 continued to say yes at endline; among the 16 who had said no at baseline, 9 had switched to yes at endline. However, these differences were not statistically significant. Roughly the same proportion of AWWs (about 95 percent) reported storing water in a closed container during the baseline and endline. At baseline 62 of 182 AWWs reported boiling and cooling water before storage and 82 of 182 reported filtering before storage. At endline, this changed to 102 and 45 respectively. Out of 198 AWCs, 169 had no toilets during baseline, and of these 169, 28 reported having toilets during endline (7 of the 29 who had reported having toilets at baseline switched to “no” at endline) (McNemar's chi square p value=0.001). Of the 20 AWCs for whom we have baseline and endline data on how clean the toilet was (based on enumerator's inspection), 17 were either clean or somewhat clean at both baseline and endline.

The enumerators checked to see if soap was available at the AWC for handwashing and whether it was being used. For 200 AWCs data on this was available at both baseline and endline. Of the 108 AWCs where soap was available and was being used during baseline, soap was not available during the endline at 24 AWCs. On the other hand, of the 88 where soap was not available at baseline, soap was available and used at 62 AWCs.

More than 95 percent of the AWWs at baseline and endline reported that the floors in and around the AWCs get cleaned everyday. Of the 107 of 199 AWCs which were recorded as having “okay” level of clean floors at baseline, 46 were reported to have clean floors during the endline. At baseline about 91 percent were at “okay” or “clean” levels of clean floors, while this changed to about 84 percent at endline.

While the results related to availability of and use of soap supports the hypothesis, the evidence is not consistently supportive of the hypotheses across different outcomes related to cleanliness and hygiene.

Hypothesis 46

This section investigates the hypothesis that Gram Varta improved the quality of work and activities related to preschool children. There was no significant increase in the number of children enrolled in the preschool, the number of children with disabilities attending preschool, or the number of parents who attended the last parent-teacher meeting held by the AWW.

At baseline, 42 (N=196) AWWs had reported that none of the preschool children used the ICDS-prescribed workbooks. Of these 42, 36 changed their answers at endline to “all” or “some of them” when asked about workbook use. Of the 104 who had responded to this question with “some of them” at baseline, 57 changed to “all.” While 98 percent of the AWWs (N=194) reported inviting parents for a meeting or discussion either once a week or once a month at baseline, the corresponding proportion at endline was about 95 percent.

There is thus some support for our hypothesis on the basis of the increased proportion of preschool children who use workbooks at endline versus baseline.

Hypothesis 47

This section investigates the hypothesis that Gram Varta improves counselling of pregnant and lactating women. One of the hypothesized outcomes of Gram Varta was increased demand, and consequently, increased utilization of services. We specifically looked at whether greater numbers of pregnant and lactating women received counseling by the AWW. The average number of pregnant and lactating women counseled in the previous two months decreased by roughly 2 (paired t test p value=0.003) for pregnant women and 3 (paired t test p value<0.001) for lactating women. Based on these results we reject our hypothesis.

Hypothesis 49

This section investigates the hypothesis that Gram Varta improves AWWs' health knowledge. We enquired about a variety of health issues that the AWW is expected to know about. Out of 200 AWWs, 96 percent said that antenatal checkups are important for pregnant women, which changed to 98.5 percent at endline. At baseline 202 out of 212 AWWs responded that polio vaccination was very important to protect health. This changed to 199 out of 212 at endline. A big change was observed when a similar question was asked about measles vaccination. While 176 out of 196 had responded at baseline that they did not know whether measles vaccine was important to protect health, 177 out of 196 said that it was very important at endline. All AWWs reported that feeding colostrum was either somewhat important (4 of 200) or very important (196 of 200) at baseline and this changed to 100 percent reporting “very

important" at endline. Similarly, while 94.3 percent of 192 AWWs at baseline said that bottled water should be given to children at age six months, this figure changed to 98.4 percent at endline. Almost all AWWs, at both baseline and endline, responded that they believed that open defecation or open sewage water represents a health hazard to themselves, their families or their community. While 14 of 200 AWWs had reported at baseline that using a toilet for defecation was not important, all 200 reported that it was very important at endline. A bigger change was observed when they were asked whether covering excreta with mud (in case no toilet was available) was important or somewhat important or not important. All 8 who answered "not important" and all 24 who answered "somewhat important", at baseline, changed to "very important" by endline (N=195). Almost all AWWs at baseline and endline said that ensuring that women get a good balance of foods was very important. A similar trend was seen in response to the questions on how important it was to ensure that adolescent girls get a good balance of foods; on ensuring that adolescent girls get enough food, attention and care; the importance of IFA supplements for adolescent girls; and ensuring that children get a good balance of foods.

While 97 percent of 200 AWWs at baseline said that a good balance of foods would protect families from diseases and malnutrition, this proportion dropped to 88 percent at endline (with 7 percent answering "not true" and 5 percent saying "somewhat true"). This constitutes a counter-intuitive finding. About 14 percent of AWWs (N=199) responded with "somewhat important" at baseline, when asked about starting complementary feeding of babies after they are six months old (the rest said "very important"). This changed to 3.5 percent for "somewhat important" and 96 percent for "very important." A similar pattern was observed with the question on the importance of adding extra oil to every meal given to children in the 6 to 59 months age group with most of the "somewhat important" and "not important" responses at baseline changed to "very important" at endline. However, among the 150 of 195 who had responded "very important" at baseline, 32 changed to "somewhat important" and 11 to "not important" at endline.

A notable shift in responses from baseline to endline was seen when the AWWs were asked how much food a woman having her period should be given. While a majority (157 of 198) answered "usual amount of food" at baseline, among the 27 who said "less food than usual" 25 changed to "usual amount of food at the endline. Similarly, 9 of the 10 who said "less food than usual sometimes" at baseline changed to "usual amount of food" at endline.

These findings, especially the results related to measles vaccination, covering excreta with mud, starting complementrary feeding at six months, and feeding usual amount of food to menstruating women, support our hypothesis that Gram Varta improved the health knowledge of AWWs.

Hypothesis 50

This section investigates the hypothesis that Gram Varta improves immunization practices. The average number of pregnant women (during the previous month) who did not complete the immunization schedule prescribed to them decreased by 0.34 (not significant, N=66) between baseline and endline. The similar statistic for children was an increase of 1.71 (paired t test p value=0.012, N=86). On the other hand, the average number of pregnant women whom the AWW motivated during the previous month to get immunized increased by 4.24 (not significant, N=136). The corresponding statistic for children was an average increase of 3.2

(not significant, N=139). Data from 163 AWWs shows an average increase of 6.44 when it comes to children of six years or younger registered at the AWC and fully immunized (paired t test p value<0.001). Taken together, there appears to be inconsistent evidence to support the hypothesis that Gram Varta improves immunization practices.

Hypothesis 51

This section investigates the hypothesis that Gram Varta improves facilitation of routine check-ups. We asked AWWs whether they helped people in their area access public health services. Because 98 percent of them (N=199) said yes at both baseline and endline, we are unable to find evidence supporting this hypothesis.

Hypothesis 52

This section investigates the hypothesis that Gram Varta increases the job satisfaction of AWWs. When asked how much they enjoyed their job as an Anganwadi worker, 28 of 200 AWWs at baseline, had said “not very much.” Of these, 24 changed to “very much” and 3 to “somewhat” at endline. At baseline 164 of 200 said “very much” while at endline 180 of 200 said “very much.” We also asked, “Do people come to you or the AWC on their own or do you have to go out and persuade them to take advantage of the services provided at the Anganwadi center?” Of the 12 who responded “people come on their own” at baseline, 10 changed to “both” at endline. Of the 67 who said “I have to go out” at baseline, 48 switched to “both” by endline. The majority (121 of 200) said “both” at baseline, and of them 22 switched to “people come on their own” at baseline, while 31 switched to “both.”

At baseline,113 out of 200 AWWs said “yes, mostly” and 79 out of 200 said “yes, somewhat” in response to the question, “Do you think people listen to the advice that you give them as an Anganwadi worker?” This changed to 163 out of 200 saying “yes, mostly” and 35 out of 200 saying “yes, somewhat” at the endline. We also asked, “If you point out a problem or give a recommendation in your function as Anganwadi worker, does the community act on it?” At baseline, 91 out of 199 said “yes, always”, 49 said “yes, most of the time”, and 58 said, “yes, sometimes.” This changed at the endline to 168 out of 199 saying “yes, always”, 13 saying “yes, most of the time” and 16 saying “yes, sometimes.”

Taken together, these results support our hypothesis about increased job satisfaction of AWWs.

Hypothesis 53

This section investigates the hypothesis that Gram Varta improves participation of AWWs in community health events. Compared to baseline, AWWs at the endline reported organizing one average 0.13 fewer events with the Mahila Mandal during the previous two months (N=115). However, this was not statistically significant. Similarly, they reported organizing 0.22 fewer events during the previous two months, on average, with the Mata Samiti (N=121), and this too was not significant. The number of counseling sessions under the Kishori Shakti Yojana also showed a similar pattern. While 25 out of 154 reported at baseline that they had organized a Kishori Swasthya Mela during the previous three months, this increased by 1 AWW only (26 of 154) at endline. We therefore find no evidence to support our hypothesis that Gram Varta improves participation of AWWs in community health events.

Hypothesis 54

This section investigates the hypothesis that Gram Varta improves perception of and respect for AWWs by the community. At baseline,113 out of 200 AWWs said “yes, mostly” and 79 said “yes, somewhat” in response to the question, “Do you think people listen to the advice that you give them as an Anganwadi worker?” This changed to 163 out of 200 saying “yes, mostly” and 35 saying “yes, somewhat” at the endline. We also asked, “If you point out a problem or give a recommendation in your function as Anganwadi worker, does the community act on it?” At baseline, 91 out of 199 said “yes, always”, 49 said “yes, most of the time.” and 58 said, “yes, sometimes.” This changed at the endline to 168 out of 199 saying “yes, always”, 13 saying “yes, most of the time” and 16 saying “yes, sometimes.” Based on these findings, we find support for our hypothesis that Gram Varta improves the perception of respect for AWWs by the community.

Summary

The present evaluation has demonstrated that Gram Varta did not lead to an increase in the utilization of Anganwadi health centers, despite the demand-side focus of the intervention. While there is some weak evidence in favour of improvements in nutrition-related services, we found no evidence regarding improvement of malnutrition treatment. Similarly, Gram Varta was not consistently associated with better hygiene and cleaning practices, but we found weak support for the hypothesis that quality of work and activities with pre-school children improved. Interestingly, while there is evidence for better health knowledge among AWWs, counselling of lactating and pregnant women did not improve. This is again interesting in relation to the theory of change. Even though knowledge improved, this did not fully translate into action, similar to what was found for CMs. Moreover, while AWWs were more satisfied with their jobs and felt more respected by the community, their participation in community health events did not increase.

7.7 Results related to health outcomes

A crucial part of the evaluation of Gram Varta is the effect of the programme on observable and self-assessed health outcomes. This corresponds with Stage 6 of our ToC figure, and tests the theory that due to improved HNWASH practices, service demand and availability and solved problems on the village level, women and children are healthier and less malnourished. As data from both treatment and control villages is available for these indicators, we are able to obtain causal effect estimates.

Several of the indicators used under these hypotheses are self-reported. They are based on questions about how the woman respondent rates her own health and on the mother’s report of her children’s recent illnesses. It is evident that respondents can potentially misreport health outcomes. They might underreport ill health in order to make a good impression or if they did not recognise symptoms of a disease. They might also exaggerate ill health if they expect to receive help or any benefits. Systematic misreporting in either direction could lead to a bias in the estimation results.

Self-reported information about diarrhoeal episodes or recent acute respiratory infections of children was not crosschecked with data from health facilities. First of all, these data were not available. Secondly, our data showed that sick children were not always (not even most of the time) taken to a doctor or health facilities. As an objective measurement of health outcomes, anthropometric measurements were taken from all household members. Trained enumerators took measurements of height and weight from all household members. They also did blood tests to record the haemoglobin concentration as an indicator of anaemia. Children below the age of five at baseline were tested for oedema and their mid upper arm circumference was measured.

Hypothesis 55

This section investigates the hypothesis that Gram Varta improves women's health based on the household and pregnant women sample.

Table 7.50: Woman's self-assessed health, household sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Self-assessed health									
Very bad	-0.0009	0.732	3134	-0.0005	0.851	3130	0.0044	0.951	4992
	(0.0025)			(0.0025)			(0.0725)		
Bad	-0.0049	0.731	3134	-0.0026	0.850	3130			
	(0.0141)			(0.0140)					
Moderate	-0.0015	0.731	3134	-0.0008	0.850	3130			
	(0.0043)			(0.0042)					
Good	0.0064	0.731	3134	0.0035	0.850	3130			
	(0.0185)			(0.0183)					
Very good	0.0008	0.730	3134	0.0005	0.850	3130			
	(0.0024)			(0.0024)					
Feels chronically tired	-0.0215	0.227	3230	-0.0203	0.262	3126	0.0539	0.096	4988
	(0.0178)			(0.0181)			(0.0322)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates. Stars mark p-values which remain significant even under multiple testing adjustment.

In the household sample, Gram Varta did not show any significant effect on women's self-assessed health (table 7.50). Although coefficients were in the right direction, they were close to zero and insignificant. In the difference-in-differences specification, there was a positive, marginally significant treatment effect of about 5 percentage points on the probability of feeling chronically tired. However, specifications 1 and 2 point in the opposite direction and significance of the DiD effect disappeared once multiple testing was accounted for. We found very similar results when looking at active SHG members but none of the effects was significant. An interesting result appeared when we split the full sample according to the woman's age. For young women aged 18 to 35 we found a negative intention-to-treat effect on health. For older woman above 35 years the effect was positive. While p-values were considerably lower when splitting the sample, still no effect was significant at conventional levels.

Table 7.51: Health of main woman, household sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Weight (kg)	0.7796	0.051	2120	0.5581	0.129	2118	0.102	0.814	5521
	(0.3967)			(0.3654)			(0.4343)		
Haemoglobin level (g/dl)	0.0058	0.941	2030	-0.0023	0.977	2028	0.1225	0.227	5358
	(0.0783)			(0.0785)			(0.1013)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

Table 7.51 presents results of the woman's non-self-reported health measures for the household sample. Note, that height was not measured for adults older than 21. We found a positive effect on women's weight with an increase of 102 to 780 grams in the treatment group relative to the control group. However, none of the coefficients were significant at the 5 percent level. The results for haemoglobin do not allow a clear interpretation as the effect sign changed across specifications and all coefficients were insignificant. When considering the sample of active SHG members we found a positive and significant increase in women's weight in all three specifications, confirming the general trend identified in the main sample. Gram Varta increased active SHG members' weight by 1.03 to 1.65 kilograms. We also detected a positive but insignificant effect on haemoglobin.

Table 7.52: Health of all women in reproductive age (20-49), household sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Weight (kg)	0.3871	0.331	2600	0.199	0.595	2596	-0.2061	0.586	5976
	(0.3968)			(0.3732)			(0.3786)		
Haemoglobin level (g/dl)	0.0117	0.879	2467	0.0091	0.904	2463	0.1454	0.102	5763
	(0.0764)			(0.0753)			(0.0889)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

Table 7.52 considers all women in reproductive age, not only the main woman. This means that specification 3 presents a difference-in-differences estimate that is based on two cross-sections and controls for household fixed effects but not individual fixed effects. The effect of Gram Varta on the weight of women in reproductive age pointed in different directions

depending on the specification and none of the effects was significant. The coefficients on haemoglobin were positive but insignificant. In the sample of active SHG members we observed a positive effect on women's weight and the effect was significant in specifications 1 and 2 and ranged between 0.5 and 1.32 kilograms. The effect on haemoglobin was positive, but not significant.

Table 7.53: Woman's self-assessed health, pregnant women sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Self-assessed health									
very bad	-0.0003	0.884	1603	-0.0004	0.885	1530	0.0347	0.708	3601
	(0.0022)			(0.0024)			(0.0926)		
bad	-0.0019	0.883	1603	-0.0020	0.884	1530			
	(0.0130)			(0.0137)					
moderate	-0.0032	0.882	1603	-0.0032	0.883	1530			
	(0.0214)			(0.0215)					
good	0.0043	0.882	1603	0.0044	0.884	1530			
	0.0288			0.0298					
very good	0.0011	0.883	1603	0.0011	0.883	1530			
	0.0078			0.0079					
Feels chronically tired	-0.0162	0.686	1598	-0.0086	0.833	1525	0.0464	0.347	3594
	(0.0400)			(0.0405)			(0.0491)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

In the pregnant women sample, results for self-assessed health were very similar to those based on the household sample, although the coefficient of the difference-in-differences specification was somewhat larger (table 7.53). These findings were confirmed when looking at the sample of SHG members. However, none of the effects was significant.

The coefficient sign on the probability of feeling chronically tired turned positive in the third specification of the main analysis, while all effects were negative for the SHG sample and the coefficients of specifications 1 and 2 are even significant. In the subgroup-analysis by blocks, we found that Gram Varta had a significant negative impact on self-reported health in Gwalpara. However, for Madhepura Sadar the effects rather pointed in the direction that self-reported health improved, but coefficients were only significant in specification 1 and the sign of the estimate for feeling chronically tired was inconclusive.

In table 7.54 we investigate the non-self-assessed health of women in the pregnant women sample and find no convincing evidence that Gram Varta has positive effects on women's health. We found negative but insignificant effects on weight and positive but insignificant

effects on haemoglobin. In the sample of SHG members we found insignificant adverse effects on both weight and haemoglobin.

Table 7.54: Health of woman, pregnant women sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Weight (kg)	0.0573 (0.4162)	0.891	1279	-0.1545 (0.3885)	0.691	1223	-0.5616 (0.5397)	0.300	3244
Haemoglobin level (g/dl)	0.1257 (0.0994)	0.208	1242	0.0929 (0.1012)	0.360	1189	0.1177 (0.1266)	0.354	3145

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

Overall, we found small, positive and insignificant effects on women's self-assessed health, but the effects for feeling chronically tired were often negative. Further, we observed positive but insignificant effects on haemoglobin and mixed, insignificant results for weight.

Hypothesis 56

This section investigates the hypothesis that Gram Varta improves husbands' health based on the household and pregnant women sample.

Table 7.55: Husband's self-assessed health, pregnant women sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Self-assessed health									
very bad	0.0006 (0.0006)	0.333	1580	0.0004 (0.0007)	0.532	1507	-0.1917 (0.1110)	0.086	2224
bad	0.0086 (0.0090)	0.339	1580	0.0058 (0.0095)	0.543	1507			
moderate	0.0242 (0.0256)	0.343	1580	0.0154 (0.0255)	0.547	1507			
good	-0.0248 0.0262	0.343	1580	-0.0162 0.0268	0.546	1507			
very good	-0.0086 0.0090	0.341	1580	-0.0054 0.0089	0.341	1507			
Feels chronically tired	-0.0038 (0.0370)	0.918	1572	-0.0048 (0.0383)	0.900	1501	0.0188 (0.0750)	0.802	2206

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

We found weak evidence that Gram Varta reduces self-assessed health for husbands from the pregnant women sample (table 7.55). This can be seen in the negative and statistically significant effect at the 10 percent level on self-assessed health in the difference-in-differences estimation, although it should be noted that the significance disappeared once multiple testing was accounted for. Further, the ordered logit estimation showed positive but insignificant effects on self-assessed health being bad and very bad and positive effects on self assessed health being good and very good. As for women, the effect on the probability of feeling chronically tired was ambiguous. In the SHG member sample, specifications 1 and 3 confirmed the findings from the main analysis, while specification 2 did not.

In the subgroup analysis by blocks, in Gwalpara and Uda Kishanganj we found a significant and robust worsening of self-reported health, where self-assessed health significantly decreased and the probability of feeling chronically tired significantly increased. In the subgroup analysis by caste, we found adverse effects on self-reported health in the subgroup of general classes. While self-assessed health was reduced, the probability of feeling chronically tired was increased. Both effects were significant in the difference-in-differences specification.

Table 7.56: Health of all men, 20-49 years, pregnant women sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Weight (kg)	1.1836	0.192	450	1.3591	0.136	426	-0.5142	0.611	901
	(0.9032)			(0.9070)			(1.0092)		
Haemoglobin level (g/dl)	0.2253	0.244	422	0.2438	0.202	398	0.1868	0.553	847
	(0.1927)			(0.1899)			(0.3139)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

As shown in table 7.56 we found no significant effect on husbands' weight and the coefficient sign differed across specifications. The effect on haemoglobin was positive but insignificant. For husbands of active SHG members we observed positive but insignificant effects for both weight and haemoglobin.

Table 7.57 reports results related to non-self-reported health measures of male household members in the age of 20 to 49 in the household sample. We observed a positive effect on weight and haemoglobin of 155 to 590 grams and 0.03 to 0.07 gram per decilitre, respectively. However, the effects were not significant. This was confirmed in the sample of men living in a household with an active SHG member, where effects were also positive, considerable larger (0.61 to 1.50 kilograms and 0.05 to 0.32 gram per decilitre), but also not significant.

Overall, we found adverse and partially significant effects on men's self-reported health and rather positive but insignificant effects on not self-reported measures, leaving us with the conclusion that Gram Varta did not have the hypothesized positive effect on men's health.

Table 7.57: Health of all men, 20-49 years, household sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Weight (kg)	0.5902	0.379	1062	0.2074	0.740	1061	0.1554	0.940	4511
	(0.6684)			(0.6233)			(2.0657)		
Haemoglobin level (g/dl)	0.0665	0.596	1006	0.0308	0.811	1005	0.0632	0.790	2817
	(0.1252)			(0.1281)			(0.2376)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

Hypothesis 57

This section investigates the hypothesis that Gram Varta improves child health based on the household and pregnant women sample.

Table 7.58: Reported child health, household sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Recent diarrhoea (last child)									
Last 24h	-0.0041	0.173	1685	-0.0039	0.200	1683	-0.0233	0.753	2640
	(0.0030)			(0.0030)			(0.0741)		
Last 2 weeks	-0.0126	0.162	1685	-0.0118	0.190	1683			
	(0.0090)			(0.0090)					
Last 3 months	-0.0138	0.164	1685	-0.0129	0.191	1683			
	(0.0099)			(0.0098)					
No diarrhoea	0.0305	0.161	1685	0.0286	0.189	1683			
	(0.0218)			(0.0218)					
Recent ARI (last child)									
Last 24h	-0.0005	0.562	1676	-0.0004	0.200	1674	-0.0366	0.562	2614
	(0.0008)			(0.0008)			(0.0630)		

Last 2 weeks	-0.0043	0.551	1676	-0.0036	0.613	1674
	(0.0073)			(0.0072)		
Last 3 months	-0.0069	0.547	1676	-0.0059	0.609	1674
	(0.0115)			(0.0114)		
No ARI	0.0118	0.548	1676	0.0099	0.610	1674
	(0.0196)			(0.0194)		
Recent diarrhoea (other child)						
Last 24h	-0.0048	0.459	1128	-0.0049	0.468	1126
	(0.0065)			(0.0067)		
Last 2 weeks	-0.0061	0.468	1128	-0.0061	0.478	1126
	(0.0084)			(0.0086)		
Last 3 months	-0.0079	0.453	1128	-0.0078	0.464	1126
	(0.0105)			(0.0107)		
No diarrhoea	0.0188	0.458	1128	0.0188	0.468	1126
	(0.0253)			(0.0259)		
Recent ARI (other child)						
Last 24h	-0.0002	0.863	1121	-0.0002	0.830	1119
	(0.0011)			(0.0011)		
Last 2 weeks	-0.0015	0.859	1121	-0.0018	0.824	1119
	(0.0083)			(0.0079)		
Last 3 months	-0.0027	0.859	1121	-0.0032	0.825	1119
	(0.0155)			(0.0144)		
No ARI	0.0044	0.859	1121	0.0052	0.825	1119
	(0.0249)			(0.0234)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

In table 7.58 based on the household sample, the coefficients for recent diarrhoea and respiratory infection showed the hypothesized sign. However, the effects were small, at least in specifications 1 and 2, and insignificant. In the subsample of children of SHG members, we found significant effects on the probability of the last born child experiencing a diarrhoeal episode in specification 2, while specifications 1 and 3 showed the expected signs but the effects were not significant. For acute respiratory infections of the last born child and other children as well as diarrhoea of other children the effects did not consistently point in the expected directions and the effects were small and insignificant.

Table 7.59 presents results of the non-self-reported health measures of the last born child in the household sample. Note, that this included the same children that we observed at baseline and endline, therefore allowing us to control for individual fixed effects in specification 3. The signs of the effects on weight did not show a consistent direction and the effects were insignificant. Against expectations, the coefficient for height was negative in all specifications and significant at the 5 percent level in the difference-in-differences specification (although significance disappeare with the Benjamini-Hochberg correction). The negative effect in the difference-in-differences specification was extremely large: last born children in the treatment group grew 3.6 cm less than children in the control group. The effect on haemoglobin levels was positive but insignificant. Note, at baseline haemoglobin was only measured for individuals older than five, not allowing a difference-in-differences estimation for this indicator. The effects for oedema and the middle upper arm circumference were positive and insignificant.

Table 7.59: Health of last born child, household sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Weight (kg)	0.0998	0.646	639	-0.06	0.729	639	0.0395	0.897	1985
	(0.2167)			(0.1728)			(0.3042)		
Height (cm)	-0.3171	0.742	645	-0.7639	0.330	645	-3.5842	0.042	2470
	(0.9606)			(0.7811)			(1.7626)		
Haemoglobin level (g/dl)	0.0989	0.477	475	0.0430	0.738	475			
	(0.1388)			(0.1281)					
Oedema	0.0058	0.476	651	0.0095	0.306	651	0.0238	0.448	2478
	(0.0081)			(0.0093)			(0.0314)		
Middle upper arm circumference (cm)	1.7115	0.288	542	1.6763	0.307	542	1.2059	0.676	1958
	(1.6030)			(1.6336)			(2.8788)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

The results for the non-self-reported health measures of the last born child of active SHG members were similar. While we found mixed evidence for weight and height, the effects were positive for oedema and arm circumference. However, none of the coefficients was significant.

Table 7.60 presents results of the non-self-reported health measures of all children under five in the household sample. Specification 3 controlled for household but not individual fixed effects. The regression results showed a negative but insignificant effect of Gram Varta on children's weight. For last born children, the effect on children's height was negative and highly significant in specifications 2 and 3. Other children of the woman respondent grew on average 1.6 to 2.9 cm less than children in the control group. Notably, the DiD effect remained

significant even with multiple testing adjustment. The coefficient in the haemoglobin estimation was positive, while the effect on oedema was negative in all three specifications. However, none of the coefficients was significant at the 10 percent level, nor were those for the arm circumference. For children under the age of five in households of active SHG members we did not find consistent effects of Gram Varta on weight, height, haemoglobin, oedema and arm circumference across specifications.

Table 7.60: Health of all children under the age of five, household sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Weight (kg)	-0.0337	0.841	1582	-0.1887	0.144	1582	-0.123	0.651	4839
	(0.1674)			(0.1286)			(0.2722)		
Height (cm)	-1.1221	0.198	1701	-1.6066	0.013	1701	-2.9198	0.004*	6037
	(0.8686)			(0.6372)			(1.0174)		
Haemoglobin level (g/dl)	0.0962	0.396	1048	0.0534	0.622	1048			
	(0.1131)			(0.1081)					
Oedema	-0.0046	0.492	1758	-0.0042	0.529	1758	-0.0056	0.656	5900
	(0.0067)			(0.0067)			(0.0126)		
Middle upper arm circumference (cm)	1.2951	0.277	1345	1.1125	0.365	1345	-0.0211	0.989	4350
	(1.1881)			(1.2243)			(1.5298)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates. Stars mark p-values which remain significant even under multiple testing adjustment.

In the pregnant women sample, women were pregnant at baseline (Table 7.61). Therefore, we only observed last born children at endline and only estimated specifications 1 and 2. We found positive yet insignificant effects on weight, height and haemoglobin. However, when we looked at last born children of active SHG members we found positive effects on weight, negative effects on height and mixed effects on haemoglobin, though none of the coefficients were significant.

Table 7.61: Health of last born child, pregnant women sample

	Model 1			Model 2		
	B/SE	P value	N	B/SE	P value	N
Weight (kg)	0.0827	0.551	1201	0.0124	0.930	1150
	(0.1383)			(0.1412)		
Height (m)	0.4669	0.406	1194	0.2801	0.621	1144
	(0.5604)			(0.5650)		
Haemoglobin level (g/dl)	0.0247	0.838	893	0.0309	0.799	857
	(0.1201)			(0.1210)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values and Model 2 further adds controls.

For children under the age of five in pregnant women households, we found positive but insignificant effects on weight, height, and haemoglobin (table 7.62). When we looked at the sample of under five-year-old children in active SHG member households, we found mixed results for weight, negative effects for height and positive effects for haemoglobin, and none of the coefficients was significant.

Table 7.62: Health of all children under the age of five, pregnant women sample

	Model 1			Model 2		
	B/SE	P value	N	B/SE	P value	N
Weight (kg)	0.1181	0.382	1985	0.0559	0.674	1904
	(0.1346)			(0.1328)		
Height (m)	0.6260	0.318	1981	0.3864	0.516	1898
	(0.6241)			(0.5934)		
Haemoglobin level (g/dl)	0.1401	0.193	1461	0.1091	0.307	1402
	(0.1072)			(0.1065)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values and Model 2 further adds controls.

Overall, we found some indication that child diarrhoea of the last born child was reduced. However, this was not true for other children under the age of five in the household. We did not find convincing evidence that confirms our hypothesis for all the other indicators. We even found a negative and significant impact of Gram Varta on children's height in the household sample.

Summary

We hypothesized that Gram Varta has a positive impact on women's, husband's and children's health under the age of five. However, we did not find statistically significant and consistent evidence to confirm these hypotheses. We found a few positive and consistent effects on women's weight, but the results were only significant in the sample of active SHG members. Further we found consistent yet insignificant evidence that Gram Varta improved self-assessed health of women but that it reduced self-assessed health of husbands (Hypothesis 55). Moreover, we found some evidence that Gram Varta reduced the probability of diarrhoea in the last born children of the respondent but not in her other children. Further, our results showed adverse effects on children's height (Hypothesis 57).

Based on our results we cannot conclude that Gram Varta has improved women's, husband's or children's health. Especially since most of the significant results were not robust to the Benjamini-Hochberg correction. As such, they could be driven by chance.

7.8 Results related to social cohesion

Lastly, we investigated the effect of Gram Varta on social cohesion. The cooperation between women SHGs and the community was expected to increase a sense of partnership within villages. In the ToC this is indicated by the assumptions that the community comes together in solidarity; it supports all members in the health-promoting decisions they make; and that participants speak up to demand improvements in service provision in the community (Stage 4). A crucial assumption in the ToC is that all of these dialogues take place and that in the process the community is truly mobilized to jointly take action and solve the identified problems. The expectation is that beliefs regarding obstacles and constraints preventing the desired behaviours are removed. Indicators of trust hence form an important part of the impact evaluation of Gram Varta. Again, given the randomized design, these results can be given a causal interpretation.

Hypothesis 58

This section investigates the hypothesis that Gram Varta increases mutual trust within the community based on the household sample.

Table 7.63: Mutual trust in the community, household sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Most people can be trusted									
Strongly agree	0.0125	0.416	3039	0.0141	0.357	3035	-0.0648	0.319	4715
	(0.0154)			(0.0153)			(0.0648)		
Somewhat agree	0.0001	0.901	3039	0.0001	0.887	3035			
	(0.0006)			(0.0006)					
Somewhat disagree	-0.0072	0.414	3039	-0.0082	0.355	3035			
	(0.0088)			(0.0088)					
Strongly disagree	-0.0054	0.418	3039	-0.0061	0.359	3035			
	(0.0066)			(0.0066)					
Take advantage of me									
Strongly agree	-0.0102	0.360	3006	-0.011	0.322	3002	0.1575	0.042	4682
	(0.0111)			(0.0111)			(0.0771)		
Somewhat agree	-0.0108	0.364	3006	-0.0117	0.327	3002			
	(0.0119)			(0.0119)					
Somewhat disagree	0.0109	0.361	3006	0.0118	0.322	3002			
	(0.0119)			(0.0119)					
Strongly disagree	0.0101	0.364	3006	0.0109	0.326	3002			
	(0.0111)			(0.0111)					
No trust in money matters									

Strongly agree	-0.0172	0.124	3001	-0.0158	0.155	2997	0.2049	0.003*	4697
	(0.0112)			(0.0111)			(0.0686)		
Somewhat agree	-0.0161	0.130	3001	-0.0149	0.162	2997			
	(0.0106)			(0.0106)					
Somewhat disagree	0.0177	0.123	3001	0.0164	0.155	2997			
	(0.0115)			(0.0115)					
Strongly disagree	0.0155	0.131	3001	0.0143	0.163	2997			
	(0.0103)			(0.0103)					

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates. Stars mark p-values which remain significant even under multiple testing adjustment.

For indicators under this hypothesis, coefficients consistently pointed in the expected direction. In the difference-in-differences specification, two effects were found to be significant at the 5 percent level (table 7.63). Treatment increased the probability that the woman disagrees that people in the neighbourhood are likely to take advantage of someone like her. It also increased the probability of disagreeing that people in the village do not trust each other in matters of borrowing and lending money. The effects were quantitatively large, with about 16 and 20 percentage points respectively, remained significant for the DiD model when multiple testing is adjusted for, and were in line with the hypothesis of increased mutual trust.

In the subsample of active SHG members, the direction of Gram Varta's impact was confirmed, although significance was lost and effect sizes in the difference-in-differences specification considerably reduced.After splitting the full sample into different subgroups, although effects were insignificant, the pattern of increased mutual trust was confirmed with very few exceptions. Overall, we found weak evidence in favour of the hypothesis that Gram Varta increases mutual trust within the community.

Hypothesis 59

Using data from the household sample, this section investigates the hypothesis that Gram Varta reduces tensions in the neighbourhood.

Table 7.64: Tensions in the neighbourhood, household sample

	Model 1			Model 2			Model 3		
	B/SE	P value	N	B/SE	P value	N	B/SE	P value	N
Problems due to religion	0.0051	0.862	2964	0.0028	0.923	2960	0.0484	0.387	2928
	(0.0294)			(0.0295)			(0.0558)		
Problems due to caste	-0.0101	0.934	2944	-0.0240	0.844	2940	0.0704	0.284	2909
	(0.1219)			(0.1217)			(0.0655)		

The table above reports marginal effects with standard errors below coefficients and separate columns for p-values and sample sizes. Model 1 is an adjusted comparison of endline values, Model 2 further adds controls and Model 3 provides linear difference-in-difference estimates.

We found no significant effects for the full sample, although the coefficient for the first indicator implies an increase in problems due to religion in all three specifications (Table 7.64). In the subsample of SHG members, treatment seemed to increase the probability that problems due to caste occur. The coefficient was positive and sizable, implying an increase of 19.1 percentage points. However, it was not consistent across specifications. Based on this evidence we reject the hypothesis that Gram Varta reduced tensions in the neighbourhood.

Summary

This hypotheses group investigated the impact of Gram Varta on social cohesion. Results for indicators on mutual trust (Hypothesis 58) showed positive effects of Gram Varta, but were mostly insignificant. We found only weak evidence in favour of hypothesis 58. Treatment effects on problems due to caste or religion were often negative and inconsistent and therefore we reject the hypothesis that Gram Varta reduced tensions in the neighbourhood (Hypothesis 59).

7.9 Summary of hypotheses tests

The main results of the quantitative evaluation are summarized in the table below. Even with multiple testing adjustments some beneficial effects of Gram Varta were observed on women's agency and empowerment. However, the results were overall inconsistent and statistically insignificant for most hypotheses. We discuss potential reasons for the lack of findings on most dimensions in Section 8 of this report.

Table 7.65: Summary of main findings

Hypotheses group	Main findings	Quality of evidence
Women's self-help groups	• No increase in participation in SHG groups • Despite more discussion on health/nutrition and better knowledge on health and health finance, no increase in uptake of government health services or financial services	• Not causal, due to lack of control group
Agency and empowerment	• Not sufficient evidence to confirm positive impact on (financial) independence • Consistent and significant effects appear on a number of indicators related to women's social capital and self-confidence • Despite no reduction in women's acceptance of domestic violence, some evidence for reduction in frequency and intensity of domestic violence commited by husbands • Consistent evidence for reduction in controlling behaviour of husbands • No consistent and significant effects on adolescent girls' outlook, child and marriage preferences, care	• Causal effects

HNWASH knowledge/practice	• Effect on HNWASH indicators overall inconsistent and almost always insignificant	• Causal effects
Pregnancy	• Some evidence that women are more accepting and mindful of their pregnancy but inconsistent results on actual antenatal care take-up and satisfaction • Overall, results are very vulnerable to multiple testing adjustment	• Causal effects
Anganwadi centers	• Evidence suggests improvement in Anganwadi workers' health knowledge as well as job satisfaction • However, no consistent evidence in favour of actual improvements in service take-up and quality	• Mostly not causal, due to lack of control group
Health outcomes	• Overall, no consistent evidence for actual health improvements • Time frame of evaluation potentially too short to uncover such effects	• Causal effects
Social cohesion	• Weak evidence in favour of improved trust in the community but no evidence for reductions in tensions between groups	• Causal effects

Results of the quantitative analysis for active SHG members were very similar to the main results. In most cases they confirmed the results of the full sample, but with lower levels of significance. This might be due to the smaller sample size leading to lower power. For some indicators the hypotheses could not be tested due to the limited number of observations and low variance. Just as the results for the full sample can be interpreted as intention-to-treat effects, those for the sample of exposed households (the active SHG member households) can be interpreted as the treatment-on-the-treated effect. Table 7.66 summarizes the findings. This analysis was performed using data only from the household and pregnant women samples.

Table 7.66: Summary of findings for active SHG members

Hypotheses group	**Main findings**	**Quality of evidence**
Agency and empowerment	• Slightly stronger evidence for positive impact on financial independence, but none on non-financial independence • Some significant, positive effects on a number of indicators related to women's social capital and self-confidence • Despite no consistent reduction in women's acceptance of domestic violence, some evidence for reduction in frequency and intensity of domestic violence committed by husbands although mostly insignificant • Consistent evidence for reduction in controlling behaviour of husbands, but mostly insignificant • No consistent and significant effects on adolescent girls' outlook, child and marriage preferences, care	• Causal effects

HNWASH knowledge/practice	• Effect on HNWASH indicators overall inconsistent and almost always insignificant • Some positive effects on knowledge about family planning methods among adult women	• Causal effects
Pregnancy	• Some evidence for increased antenatal care take-up and adverse impact on quality, yet insignificant • Weak evidence that women are more accepting and mindful of their pregnancy	• Causal effects
Health outcomes	• Overall, no consistent evidence for actual health improvements, except weak evidence for positive impact on women's weight • Time frame of evaluation potentially too short to uncover such effects	• Causal effects
Social cohesion	• Much weaker evidence in favour of improved trust in the community and no evidence for reductions in tensions between groups	• Causal effects

Summary results complementing the impact timeline

Order of change	Expected change	Summary of results
1	Change in knowledge	• The knowledge of facilitators seems to have improved (H8) and similarly for AWWs (H49). • Women's knowledge of childhood diseases does not appear positively affected by Gram Varta (H34). • There is some evidence that adult women's knowledge about family planning methods increased due to Gram Varta (H35).
2	Increased community solidarity	• There is some indication as reported by facilitators that cooperation between SHGs and the community improved (H2). • Gram Varta also seems to slightly increase women's involvement in the community (H14). • AWWs feel more respected by the community (H54), although Gram Varta does not appear to improve participation of AWWs in community health events (H53).
2	Change in attitude	• We do not find effects on acceptance of domestic violence (H17), on adolescent girls' attitude toward family size (H20), perceived ideal marriage age (H22), attitude toward care for daughters, opinion about balanced nutrition (H28), and open defecation (H33). • Pregnant women do not receive increased family support for obtaining antenatal care (H38). • There is weak evidence that the son preference of adolescent girls is reduced (H21). There is weak evidence for an improved attitude toward feeding colostrum and early breastfeeding (H29) and also for pregnant women to feel less stressed and worried (H40).
2	Increased self-confidence	• Gram Varta does not seem to strengthen women's decision-making power in the household (H11). • Gram Varta seems to increase women's self-confidence when it comes to refusing intercourse (H16).

3	Change in practices regarding hygiene	• There is no evidence for increased water treatment (H32). • We do not find positive effect on hygiene behaviour such as handwashing (H33).
4	Increased non-financial independence	• Gram Varta does not seem to increase women's independence from their husbands (H12, H13). • Gram Varta increases the likelihood that women have non-agricultural income.
4	Increased use of services	• There is no evidence from the CM, the household or the AWW survey that use of government health services improved (H5, H42).
5	Change in practices regarding sanitation, diet	• There is no indication of a positive effect on micronutrient supplementation for children under 5 (H27) or a reduction of risky consumption behaviour (H31). • There is no evidence of improved sanitation practices (H33).
6	Change in pregnancy-related behaviour, child care	• There is no evidence that the probability of early pregnancy is reduced (H23). • There is no evidence for increased prevention of diseases in children under 5 (H30). • There is only weak evidence that Gram Varta encourages pregnant women to be mindful of their health (H36). However, the effect on antenatal care practices is rather adverse (H37). • Evidence that Gram Varta increased women's use of contraception. • Gram Varta increased the likelihood of pregnant women making decisions regarding their family's diet.
7	Increased financial independence	• There is no evidence that Gram Varta encourages women to acquire paid work and become more economically independent (H10). Some evidence that it increased the likelihood of having non-agricultural income.
7	Increased trust and reduced tensions	• There is weak evidence of increased trust within communities (H58), but no reduction of tensions (H59).
8	Increased quality and accountability of services	• There is no supporting evidence that Gram Varta increased the quality of antenatal care visits (H37). • Results from the AWW survey suggest improved efforts to prevent malnutrition, but no improvement of its treatment (H43). • There is no evidence of improvement in weighing practices (H44), cleanliness and hygiene (H45), counseling of pregnant and lactating women (H47), AWW activities related to preschool children (H46), immunization practices (H50), and routine check-ups (H51).
9	Change in health outcomes	• Weak evidence of better self-assessed health of women (not men), but no consistent impact on non-self-reported health (H55, H56). • No consistent impact on health of children, self-reported and non-self-reported (H57). • Weight of active SHG members' increased by about 1.03 to 1.65 kilograms.Weak evidence that haemoglobin levels increased.

Summary results complementing the ToC

Stage in ToC	Summary of results

1. Trained facilitator conducts social mapping, prepares for meetings	• The knowledge levels of facilitators appear to have increased after training (H8).
2. Trained facilitator sets up meetings	• Qualitative study and other notes from the field suggest that a few PLA meetings were not held in a few areas. • We have no data about whom the facilitators actively invited to the meetings.
3. Participation in meetings (SHG members, community and target groups actively participate in meetings)	• Overall participation in SHGs seemed to decrease in treatment areas (H1) according to the data from the CM survey. Our analysis does not find a differential trend in participation between treatment and control villages. • We do not have data on who the active participants in meetings were and to what extent members were engaged and involved. • Monthly progress reports show low attendance of men and frontline workers.
4. Participatory learning occurs. As a result: Awareness about HNWASH increases; critical thinking and problem solving skills improve; members engage much more with community and front-line workers; members feel more empowered; and community solidarity is observed.	• Qualitative evidence reports the use of picture cards, games, stories, and demonstrations. The appeal and recall value of games and stories was confirmed in our qualitative studies, however the demonstrations related to sanitation seemed to be the most memorable. • Gram Varta does not seem to strengthen women's decision-making power in the household and independence from their husbands (H11, H12, H13). • Gram Varta does strengthen women's self-confidence when it comes to sexual negotiations with their hurbands (H16). • Gram Varta reduces the likelihood of reporting domestic violence. • Results suggest (weak evidence) that Gram Varta decreases young women's preference of sons. • There seem to be weak improvements in trust within the community (H58). Self-reports from CMs suggest that there was greater cooperation between SHG members and the community in efforts to solve problems. • Gram Varta reduces stress among pregnant women. Also, some evidence that it increases the chance that they accept the pregnancy and feel optimistic about it. • Facilitators self-reported that provision of information on health care increased (H3). However, women's health-related knowledge increased only in few areas (H34, H35). • Attitudes seem to change regarding some feeding practices such as feeding colostrum (H29), but not in other topics (H17, H20, H22, H28, H33, H38). • Gram Varta increases awareness and use of contraceptive methods among women. • Gram Varta increases the number of meals consumed per day by pregnant women.
5. Community mobilization, action at individual and community level and evaluation	• We do not have data on action plans developed. • Action at individuallevel: Gram Varta increased the likelihood that the pregnant woman herself plans the diet consumed by her family (H28). • There is (qualitative) evidence that participants did not discuss topics within their households much. • Self-reports from CMs do not provide evidence that the demand for services increased.
6. Indicators of proximate determinants of HNWASH outcomes change and ultimate outcome indicators improve	• We find little evidence for improvement in HNWASH practices such as handwashing or use of soap; diets; and antenatal care. We also find little evidence of improved health.

	• Gram Varta increased the body weight of women. Weight of active SHG members' increased by about 1.03 to 1.65 kilograms.Weak evidence that haemoglobin levels increased.

As hypotheses within the same group are likely to be highly correlated, it should be noted that the multiple testing correction applied to the above analysis is a rather conservative approach and may in fact be too pessimistic for the format of this evaluation. Nevertheless, the fact that the main take-aways of the quantitative analysis do not crucially depend on whether the Benjamini-Hochberg procedure is applied or not, raises confidence in the results. The following sub-section discusses the implication of our findings for the cost effectiveness of Gram Varta.

7.10 Cost-effectiveness analysis

For a number of hypotheses, we found consistent, albeit not always significant, effects in the expected direction throughout subgroups. We conducted a cost-effectiveness analysis for indicators under these hypotheses. Unfortunately, we did not find comparable results in the literature in the limited time given. Therefore, these results have to stand on their own, but please note that they should be compared to the cost-effectiveness of other programmes influencing the same indicators.

We present cost-effectiveness in two different ways. Firstly, we express it as US dollars (USD) needed in Gram Varta for achieving one effect. For example, how many dollars are needed for one treated woman to stop experiencing domestic violence? Secondly, we express it as effect per 100 USD invested, which is simply the inverse of the first multiplied with 100. For example, how many treated women less were affected by domestic violence after investing 100 USD in Gram Varta?

Our method of calculating cost-effectiveness provides a rather optimistic account of the true costs, as we only considered actual expenditure on Gram Varta. In addition to these monetary costs, the time invested by SHG members and the community as a whole may have resulted in non-neglible intangible costs. This may account for the lower participation by men than we hoped.

Included in the cost-effectiveness analysis were (1) hypothesis 14 on social capital where we only included the indicators of being acquainted with certain officials, since this was where we found consistent effects; (2) hypothesis 16 on sexual behaviour or refusing intercourse; (3) hypothesis 18 on the practice of domestic violence, both for the household and the pregnant women sample; (4) hypothesis 21 on son preference of adolescent girls; (5) the indicator 'meals per day' in hypothesis 36; (6) hypothesis 40 on women's attitude toward their pregnancy; and (7) hypothesis 58 on mutual trust.

Hypothesis 13 seemed confirmed in the full sample and the subsample of active SHG members, but was not included due to severe inconsistencies in the subgroup analysis. Although there were a few exceptions in the subgroup analysis with mixed results, Hypothesis 40 was included . Although we found adverse effects on domestic violence in one of the blocks,

similar reasoning was used with hypothesis 18, as the results were very strong in most subgroups,

We used results from the difference-in-differences specification to capture the intention-to-treat effect of Gram Varta with the highest precision. The only exception was hypothesis 21, for which we used the result from specification 2, as there was no difference-in-differences specification for adolescent girl data. The total expenditure of Gram Varta implementation used for our calculations was 7.97 crore (1 crore is 10 million INR), which is 1,185,620 USD with the current exchange rate of 0.01488. Across all SHGs and blocks, 46133 women and 8001 adolescent girls participated in each PLA session (this is calculated based on the average attendance in a PLA meeting). Given that one woman might attend several PLA meetings, we cannot calculate participant numbers across meetings, but only across SHGs and blocks.

Table 7.67: Cost-effectiveness

	USD per effect	Effect per 100 USD
Hypothesis 14, household sample		
Gained acquaintance with health staff	399.69	0.2502
Gained acquaintance with government officials	293.38	0.3409
Gained acquaintance with school officials	992.28	0.1008
Gained acquaintance with other officials	233.85	0.4276
Hypothesis 16, household sample		
Gained refusal of intercourse if husband has STD	164.64	0.6074
Gained refusal of intercourse if husband cheats	165.38	0.6047
Gained refusal of intercourse if woman is tired	231.74	0.4315
Gained demand of condom use	544.49	0.1837
Hypothesis 18, household sample		
Averted case of pushing/shaking/throwing at wife	2,141.67	0.0467
Averted case of kicking/dragging/beating wife	690.86	0.1447
Averted case of choking/burning wife	1,359.79	0.0735
Averted case of forced sexual activities	1,036.29	0.0965
Hypothesis 18, pregnant women sample		
Averted case of pushing/shaking/throwing at wife	637.72	0.1568
Averted case of kicking/dragging/beating wife	580.14	0.1724
Averted case of choking/burning wife	444.64	0.2249
Averted case of forced sexual activities	332.04	0.3012
Hypothesis 21, household sample		
Averted son preference	1,019.15	0.0981
Hypothesis 36, pregnant women sample		
Gained meal per day	185.96	0.5377
Hypothesis 40, pregnant women sample		
Averted case of being stressed or worried	274.57	0.3642
Hypothesis 58, household sample		
Gained agreement to ‘People can be trusted’	396.61	0.2521
Averted agreement to ‘People take advantage of me’	163.17	0.6128
Averted agreement to ‘No trust in money matters’	125.43	0.7973

Table 7.67 presents results of the cost-effectiveness analysis. We found that for a woman to gain one acquaintance, be it health staff or any type of official, between 234 and 993 USD have to be invested in Gram Varta. For a treated woman to believe she is allowed to refuse intercourse in certain cases required a lower investment of 165 to 232 USD. About 544 USD needed to be invested in Gram Varta for one woman to change her mind about demanding her husband to use a condom.

Looking at hypothesis 18, we found that Gram Varta was more cost-effective in the pregnant women sample. While only 332 USD needed to be invested in Gram Varta to avert one case of a pregnant woman being forced to nonconsensual sexual activities, it is 1036 USD for the household sample. The difference was smaller for averted cases of the wife being kicked, dragged or beaten.

Hypothesis 21 included only adolescent girls. One needed to invest 1019 USD to ensure that one adolescent girl does not prefer a boy as her first or next child. This is likely to be the upper bound, as girls who did not participate in the PLA meeting might be affected by Gram Varta, in addition to, girls living in a household with a woman who participated in the PLA meeting.

One extra meal per day for a pregnant woman costs about 186 USD, if the amount is invested in Gram Varta. In other words, 100 USD invested in Gram Varta enabled one woman to have half a meal more per day than without the investment. To avert one case of a pregnant woman being stressed or worried, about 275 USD needed to be invested in Gram Varta.

Mutual trust was comparatively less expensive to achieve. For one treated woman, more to disagree with the statement that people in the village do not trust each other in money matters, 125 USD needed to be invested in Gram Varta. For 1000 USD invested in Gram Varta, more than 6 additional treated women would disagree with the statement that in the village someone is likely to take advantage of her.

Considering that costs per PLA meetings are at about 1.32 USD (one could support a full 20-meeting cycle in 5 SHGs for 132 USD), the cost-effectiveness appeared to be quite poor for these indicators. This is one more indication that Gram Varta did not prove to be very effective in Madhepura district. However, with these calculations we did not consider dependencies between effects. The calculations were done separately for each indicator.

8. Discussion

8.1 Concerns regarding internal validity

Sample comparison

One important aspect in determining the validity of the results is the validity of the experiment itself. Given that participation in SHGs is not compulsory, we are not able to obtain unbiased estimates of the average treatment effect on the treated, assuming that more motivated or interested women are more likely to participate. Instead, the main emphasis of the evaluation is the effect of being assigned to a treatment gram panchayat on various interest groups, i.e.

the intention-to-treat effect (ITT). To obtain unbiased estimates of ITT it is necessary that living in a treatment gram panchayat is unconfounded with other characteristics that affect the outcomes of interest.

A randomized field experiment has the advantage that it makes treatment assignment independent of potential outcomes; eliminating bias resulting from imbalance in observable and unobservable characteristics. However, this property only holds on average (i.e. assuming the experiment would be sufficiently often repeated) and substantial differences between both groups may still be present in any single experiment (Imai et al., 2008). To assess whether our randomization was successful, we calculated descriptive statistics for both the treatment and control group based on baseline data and calculated the deviation in means between both groups. We expressed differences in terms of standard deviations to put them into perspective.[9] We defined differences below 0.2 standard deviations as negligible, between 0.2 and 0.5 as small, between 0.5 and 0.8 as medium and above 0.8 as large.[10] Results of these calculations were presented extensively in the baseline report of this evaluation. Statistics were compared for several indicators regarding household characteristics, health, knowledge, opinions and empowerment of women, adolescent girls, and anthropometric measurements including stool and blood tests. Since pregnant women were intentionally oversampled, this subsample was analyzed separately.

Overall, the comparison of samples indicated a strong similarity between the treatment and control households, as well as within the pregnant women sample. Randomization seems to have led to a reasonably good balance in indicators. Hence we do not expect any strong confounding bias from any of these observable characteristics. Strong differences were rare and mostly limited to outcomes. To account for potential confounding, we added covariates from baseline data as control variables and made use of a standard difference-in-differences approach in two model specifications.

Spill-over across villages, contamination, John Henry, Hawthorne effects, self-reporting bias

The core of the intervention design is the participatory learning and action approach, which encourages women to see themselves as a powerful group and come together to make changes within their households and communities. A sense of empowerment and agency is expected to strengthen as the meeting cycle progresses. By design, the programme actively involves and mobilizes the community. Therefore, over time, the impact is expected to spill over to the rest of the community. It is unlikely that the feeling of empowerment and community cohesion would spill over to a counterfactual community, even if there is contact (visits/sharing) between treatment and counterfactual communities. However, it is possible that health knowledge spreads across communities due to these kinds of connections. Our sampling design ensured that treatment and counterfactual communities are separated by sufficient physical distance, thus reducing the probability of spill-over. This was achieved by assigning treatment by panchayat, since villages have close geographic proximity. Additionally, the

[9] In accordance with the CONSORT statement (Moher, 2010) – a collection of guidelines supported by leading medical journals – we do not perform any hypothesis tests, as they constitute a population statistic rather than a sample statistic. A further problem, in addition to the lacking theoretical foundation, is the sensitivity of t-tests to changes in sample sizes. For an in-depth discussion of the topic, see Imai et al. (2008).

[10] This is based on the idea that using the full sample standard deviation as a means to normalization is closely related to Cohen's d, according to which a small effect size is defines as normalized differences in means of 0.2, a medium effect sizes as 0.5 and a large effect size as 0.8 (Cohen, 1988).

cultural context and patriarchal system prevents women from frequent or any travelling to other villages and gram panchayats, further reducing the likelihood of spillovers from treatment to control areas. Programme implementation was strictly aligned with treatment assignment. Thus, contamination was only possible if control group participants travelled to other gram panchayats. For the same reasons that we disregard spill-over across villages, we also do not expect bias from contamination.

In control areas 7.2 percent of woman respondents in the household sample and 0.9 percent of women in the pregnant women sample had heard of Gram Varta. Although this is a very low rate, it is only 3.5 percentage points lower than the rate in treatment areas. Even if the name 'Gram Varta' was known to some individuals in the control area, it does not show any spread of health knowledge across communities.

To prevent John Henry and Hawthorne effects, enumerators were trained to explain the purpose of the surveys to the participants in a manner that makes it clear that they are participating in a study. The enumerators did not reveal that effectiveness of Gram Varta is of central interest for the study. Through emphasizing the survey as a health survey, participants were not aware that the survey was part of an experiment, nor were they aware of the experimental group that they belonged to. Gram Varta is implemented over a long time period and the evaluation team had very brief contact with the study participants, therefore we expect that John Henry and Hawthorne effects did not play a big role (because the novelty of being part of a study to wanes over time). Further, SHG members in treatment areas might not be aware that the sessions are part of the Gram Varta program (see discussion below on lack of publicity of the name "Gram Varta").

Due to the design of Gram Varta, we expected a potential for self-reporting bias in the follow-up of the treatment group. Individuals in the treatment communities might feel that they are being tested on whether they practice what was discussed in the meetings. There is a possibility that treatment group members feel judged when their answers do not conform with Gram Varta's recommended practices. To reduce this bias, we recorded behaviour by collecting self-reported data and also verifying the responses using control questions (such as the amount of soap bought versus how often they wash their hands) and hard data (such as requesting to see the soap and observing toilet facilities). Enumerators were rigorously trained to behave neutrally and not judge respondent answers. We used standard techniques to minimize or account for enumerator bias.

Qualitative data suggest that while levels of knowledge about expected behaviours might have increased, norms had not changed to the extent that respondents might feel judged. For instance, FGDs indicate that people had prioritized hygiene and lack of toilets over all other health problems in that village. Despite that, in our interviews and FGDs we found a pessimistic attitude towards health and hygiene which respondents did not hesitate in sharing. One of the family members of the facilitator responded:

"*Jeevika has a number of activities other than these (Gram Varta). She (his wife) does not have enough time to manage SHGs, conduct these meetings and continue her household activities. We don't need to start cleaning drives. It is never going to work for this village… people are not well-educated and they won't be able to retain whatever they have learned for a long time. They have other things to do." (IDI, man, treatment village).*

Selective attrition

Households drop out of studies for various reasons. For example, they migrate to other areas, members are unavailable at the specific date and time of the interview, they refuse to participate in the survey, and in some cases die. If the households that drop out are systematically different from the households that do not drop out, the sample becomes unbalanced, which can lead to bias. To assess whether we experienced selective attrition, we present descriptive statistics for four groups: (1) Households in control areas that were part of the endline survey; (2) Households in control areas that dropped out and were not part of the endline survey; (3) Households in treatment areas that were part of the endline survey; (4) Households in treatment areas that dropped out and were not part of the endline survey. We present means of main control variables and several indicators used in the analysis of hypotheses as recorded at baseline. In addition, we added differences in terms of standard deviations between groups 1 and 2 and groups 3 and 4. For both treatment and control group, this allows us to see whether there are systematic differences between the lost households and those that were surveyed at baseline and endline. Similar to the idea in the assessment of randomization, we interpreted small effect sizes as normalized differences in means between 0.2 and 0.5, medium effect sizes as between 0.5 and 0.8 and large effect sizes above 0.8. In the following section we consider attrition in the household sample and in the pregnant women sample. Because community mobilizers and Anganwadi workers were only interviewed in treatment regions, we do not present statistics on selective attrition for these samples.

a) Household sample

Households that dropped out of the survey in control areas had a smaller household size than those that remained in the survey by 0.53. This is a small effect size in terms of standard deviations. Notably, no such difference appeared in the treatment group. Livestock ownership also appeared to be less frequent in the lost households in control areas with the difference reaching the small effect size. Since those two variables were among the covariates added to the model in the second specification, these differences should not be a concern. No other large differences were found between groups among the control variables. Among the indicator variables tested, we observed only two differences in terms of standard deviations that are higher than the cutoff for small effect sizes. The number of vitamin A doses received by the last born child appeared higher in the lost households in the control group with a difference of 0.23, which is small in absolute terms. Also the probability of knowing correct danger signs of malaria seemed higher in this group. However, the difference in terms of standard deviations was only just above the cutoff for small effect sizes. Overall, few surprising differences were found between the four groups. Selective attrition does not seem to be a problem for the household sample.

Table 8.1: Selective attrition in the household sample

	Control			Treatment		
	In endline	Dropped	Difference	In endline	Dropped	Difference
	Mean	Mean	in SD	Mean	Mean	in SD
Household	5.28	4.75	0.2600	5.63	5.51	0.0569

size						
Children under 6	0.73	0.67	0.0686	0.90	0.98	-0.0794
Highest education	3.79	3.57	0.1114	3.86	3.68	0.0961
Education of woman	1.80	1.87	-0.0474	1.90	1.89	0.0034
Age of woman	37.65	39.53	-0.1463	36.34	38.33	-0.1565
Land ownership	1.53	1.57	-0.0715	1.52	1.54	-0.0378
Land area owned	359.40	339.55	0.0208	341.50	307.17	0.0347
Livestock ownership	0.93	0.73	0.2779	0.91	0.88	0.0543
Cattle owned	1.59	1.50	0.0602	1.61	1.63	-0.0036
Asset ownership	4.31	4.40	-0.0488	4.57	4.34	0.1117
Number of assets owned	6.58	6.96	-0.0906	7.10	6.47	0.1342
Ever taken loan	0.11	0.14	-0.0864	0.09	0.11	-0.0731
Health care decisions	0.91	0.92	-0.0382	0.87	0.88	-0.0415
Alone to market	0.13	0.10	0.0998	0.13	0.12	0.0507
Voted in election	0.89	0.89	-0.0180	0.85	0.82	0.0754
Demand condom	0.89	0.89	-0.0180	0.85	0.82	0.0754
Hit if neglect children	0.42	0.44	-0.0516	0.46	0.54	-0.1732
Afraid of husband	2.32	2.40	-0.1251	2.35	2.38	-0.0409
Importance care for daughter	2.79	2.81	-0.0368	2.74	2.68	0.1109
Vitamin A doses	0.68	0.91	-0.2263	0.84	0.75	0.0617
Balanced nutrition	2.82	2.85	-0.0606	2.78	2.73	0.0929
Feeding thick breastmilk	2.86	2.80	0.1327	2.73	2.64	0.1539
Sleep under bednet	0.94	0.92	0.0448	0.89	0.89	0.0085
Cups of chai per day	0.40	0.42	-0.0439	0.45	0.40	0.1118
Improved toilet	0.11	0.17	-0.1645	0.15	0.17	-0.0587
Danger signs of malaria	0.37	0.47	-0.2016	0.41	0.42	-0.0178
Use of contraceptio n	0.57	0.51	0.1292	0.50	0.43	0.1493
Visits	1.15	1.32	-0.0278	1.35	0.67	0.1095

Anganwandi						
Chronically tired	0.70	0.70	0.0171	0.65	0.65	0.0120
Trust in neighbourhood	1.72	1.66	0.0686	1.76	1.68	0.0920
Problems due to caste	0.20	0.22	-0.0530	0.22	0.17	0.1238

We also present the statistics for attrition across geographic blocks. Due to survey logistics we progressed through the survey block by block and panchayat by panchayat in the same block. Since during certain times of the survey farming activities were more frequent than at other times, it is possible that we followed up fewer households in a block that we visited during an increased farm work period. This would reduce the comparability of observed and lost households. We were prone to lose more people in one panchayat than the other for the same reason, potentially resulting in different attrition across experimental groups. Table 8.2 shows that attrition was slightly higher in treatment than in control areas. However, all differences were small in absolute terms. Block 6 experienced the highest attrition in the treatment group and the highest difference in attrition between treatment and control group. The differences were similar for all other blocks.

Table 8.2: Attrition across blocks in the household sample

	Control		Treatment		Difference
	N	% Attrition	N	% Attrition	Abs.
Block 1	154	0.14	176	0.15	-0.02
Block 2	286	0.12	176	0.14	-0.02
Block 3	660	0.11	748	0.10	0.01
Block 4	220	0.14	345	0.17	-0.03
Block 5	396	0.09	352	0.13	-0.04
Block 6	264	0.11	176	0.19	-0.08

b) Pregnant women sample

In both treatment and control group lost households appear to have only about half as many assets (unweighted asset index) as households interviewed at endline. This is a large difference in standard deviations. Further, we observe small differences in standard deviations between households interviewed at the endline and households not interviewed at endline, specifically in the woman's age in the treatment group and in the completion of primary education in the control group. In the estimation specification 2, we control for all three variables to address this imbalance. Lastly, we observe a small standardized difference in the treatment group in the share of women feeling recognized as themselves.

Table 8.3: Attrition in the pregnant women sample

	Control			Treatment		
	Mean	Mean	Difference in SD	Mean	Mean	Difference in SD
Interviewed at endline	No	Yes		No	Yes	
Scheduled caste	0.38	0.36	0.0440	0.29	0.28	0.0236

Scheduled tribe	0.04	0.03	0.0553	0.02	0.04	-0.1000
OBC	0.54	0.56	-0.0468	0.58	0.59	-0.0273
General caste	0.05	0.05	-0.0268	0.11	0.08	0.0927
Hindu	0.92	0.92	-0.0137	0.78	0.81	-0.0820
Asset index	2.35	5.03	-1.0993	2.67	4.94	-0.9453
BPL	0.71	0.72	-0.0249	0.65	0.69	-0.0776
Own land	0.40	0.45	-0.1051	0.41	0.41	-0.0023
Own livestock	0.50	0.54	-0.0752	0.55	0.54	0.0306
Work as farmer	0.10	0.10	-0.0212	0.09	0.08	0.0018
HH work	0.92	0.91	0.0293	0.81	0.82	-0.0280
Age	23.38	23.56	-0.0407	22.84	23.89	-0.2233
Primary	0.06	0.15	-0.2659	0.16	0.15	0.0242
Junior secondary	0.08	0.08	0.0078	0.10	0.10	-0.0037
Senior secondary	0.10	0.09	0.0575	0.07	0.07	-0.0128
Higher education	0.09	0.04	0.1954	0.06	0.07	-0.0402
Can read SMS	0.32	0.29	0.0597	0.32	0.33	-0.0218
Age at first birth	12.18	12.38	-0.0206	10.79	12.32	-0.1580
ANC at health facility	0.90	0.93	-0.1148	0.93	0.94	-0.0477
ANC by skilled health personnel	0.94	0.91	0.0879	0.90	0.88	0.0544
Breastfeeding is important	0.98	0.98	0.0443	0.98	0.99	-0.0074
Went for ANC	0.53	0.53	-0.0124	0.67	0.66	0.0194
Number of ANC visits	1.54	2.65	-0.1163	4.61	3.14	0.1169
Decides about family diet	0.72	0.67	0.1062	0.55	0.59	-0.0778
Feels chronically tired	0.73	0.75	-0.0548	0.67	0.69	-0.0511
Using a toilet or covering excreta with mud is important	0.75	0.72	0.0526	0.81	0.79	0.0560
Knows a danger sign of malaria	0.57	0.58	-0.0040	0.57	0.61	-0.0736
Feels recognized as herself	0.09	0.05	0.1489	0.13	0.05	0.3238
Pushed/shaken/having things thrown at by husband	0.09	0.12	-0.0911	0.15	0.15	-0.0081
Kicked/dragged/beaten up by husband	0.07	0.08	-0.0286	0.12	0.11	0.0429
Choked/burned on purpose by husband	0.02	0.02	0.0224	0.07	0.03	0.1595
Forced to do perform non-consensual sexuality activities by husband	0.05	0.05	0.0082	0.08	0.08	0.0202

Based on the descriptive attrition statistics of the household and pregnant women samples, we do not believe that the internal validity of our results is threatened by attrition.

Specificity and sensitivity of the results

Regarding the sensitivity of our results, we have discussed in detail the robustness of our results across three different specifications as well as subsamples. Our specifications vary in the degree of controlling for observable and unobservable characteristics taking into account imbalances that were or were not detected in the baseline characteristics balance analysis or selective attrition analysis. We were conservative in the conclusions drawn from our analysis as effects needed to be consistent across specifications and subgroups to confirm our hypotheses.

Alignment between quantitative and qualitative findings

The qualitative findings are in line with the findings of the quantitative evidence. Focus group discussions and in-depth interviews with different community members indicated that participation in Gram Varta had not created tangible transformations in women's sense of agency, empowerment, and decision-making on health related matters. The baseline and midline qualitative findings suggest that, in general, women had internalized the patriarchal concept of themselves as weak and inferior. Gram Varta participation may have helped women gain some information on HNWASH, but most of the underprivileged women did not share that there was an increase in their bargaining power.

The qualitative study was able to shed light on the key drivers of change identified in the ToC (women's empowerment, critical thinking, decision-making) in Madhepura. There did not appear to be major changes in these key constructs. In fact, some of the women felt that there were gains to be had in maintaining the traditional status quo, rather than challenging social norms. One illustrative quote from the endline :

"*This is a village in Bihar, didi. These programmes come and go here. Things go on like this here. We heard about getting together to stand up for ourselves in this programme, but I ask you, can women stand up for themselves? We have to get permission to go out, we have to maintain untouchability with other males, and we have no authority on ourselves! Men won't say anything if we start SHGs, as they also get income from us. But these programmes ask us to go out of our homes and start taking decisions which might be against traditions. We don't think this is going to happen for a long time." (FGD, treatment village, SHG member)*

The qualitative study found that some of the participants reported that they had gained new information and had the opportunity to learn about different health problems, whereas some other members felt that these meetings would not bring forth any change. The following quotes illustrate the perceptions of Gram Varta programme in the villages:

"*My friend had lost one of her children in the last year. It was during the Gram Varta meetings that I learnt that the baby could have been saved if we had followed what the CM didi had said. I found these meetings to be useful in caring for my children. (IDI, female, 25 years, SHG member)*

"*The CM didi called a meeting one day and told us many things about health (laughing). She told us that we should not go to the fields for defecating and should build pits or toilets. We did not pay much attention to her words as I don't feel these things are important." (FGD, SHG member)*

"The CM didi had called for a meeting near one of the school auditoriums. We were in the field listening to the meetings. We did not want to go near the meeting site as they were talking about issues such as maintenance of health. We feel that these things should be cared by women and we stopped listening after a few minutes." (FGD, men)

"*Our CM didi is very educated and she has told us many things about keeping ourselves healthy. Initially I thought that these meetings were boring and had only attended these meetings just because CM didi had asked us. However, when I started playing these games*

and activities, I began to have questions and asked her in private." (Seema, IDI, married, 19 years, SHG member)

In general, the qualitative study results suggested that the impact of Gram Varta was at best weak. This aligns with the quantitative findings. Qualitative interviews in March 2016 indicated that the central messages of Gram Varta activities were not understood by a majority of the community members except for teenage girls. However, at endline, almost 55 per cent of the people who reported attending meetings (with whom the qualitative interviewer conversed either through interviews or FGDs or informal conversations) mentioned that they were able to understand the central messages of these activities. In the FGDs most of the respondents who said they did not know about the messages of these activities were older women. While the messages from meetings on sanitation were fresh in their memories (they were the last meetings to be held), other meetings such as 'breaking the malnutrition cycle,' and the meeting on reproductive health appeared to have been received with considerably less enthusiasm. More than 85 percent of the respondents of the qualitative study (N=107), from both treatment and control villages, were not familiar with the term '*kuposhan*' (malnutrition).

Our qualitative study suggests that certain activities and messages were accepted more by younger women than by older women (which corresponds with the finding on empowerment among adolescents). The following was a typical finding of FGDs in treatments areas at endline:

"FGD facilitator: So, how do you assure safe pregnancies in this village?
Respondent 1: I have attended the births of my grandchildren. I always tell my daughter-in-laws do not eat much during the pregnancy. If you have a full stomach, the child will grow very big and it will be a difficult delivery. I always tell them.
Respondent 2: No, no, didi, this is very wrong. Didn't the CM didi tell that pregnant women should eat very well. They should eat fruits, green vegetables and should eat well. It will be beneficial to both mother and child.
Respondent 3: Yes, yes. We used to think like grandma, but CM didi and other didis told us that we should eat well to maintain our health as well as the baby's health. We will also have higher chances of baby living well if we do that."

The oppressive nature of social institutions such as caste, coupled with poverty, and the lack of an enabling environment, may have discouraged them from being independent. The activities of Gram Varta have resulted in expanding these women's social circles, enabling them to be in contact with health staff and other members of the community and other SHG members.

We also found that active participation in Gram Varta was hindered by women's attitudes towards health related matters. We found that health, especially their own health, was not considered as an important social concern in the villages. Thus, even though the community members became aware of the ill effects of their health habits, they were not motivated adequately to make transformations in their lives to take action. This was further complicated by the view that Gram Varta meetings were not different from regular SHG meetings, perhaps an unintended effect of using CMs as facilitators. A frequent response in FGDs was one of disappointment that Gram Varta meetings did not have any financial benefit.

Additionally, we found that problems in implementation of Gram Varta had influenced women to give low priority to Gram Varta. For instance, in all of the villages where our qualitative study was carried out, Gram Varta meetings were held during the harvest season. During this time most of the women were very busy and could not attend meetings continuously. The social identity and perceptions of the facilitators also influenced women's attitude towards involvement in Gram Varta. Our findings suggest that there is a sustained favourable response to Gram Varta when the facilitator is seen as motivated, empathetic and a familiar insider. For instance:

"*The Didi who comes to tell us about health is a stranger to us in this village. She is very young and is not married. She tells about maintaining the health of pregnant women and children, but we do not feel comfortable talking to her about our problems. We let her talk though (laughing). How can she even understand my health problems?" (Rekha, IDI, pregnant woman (March 2016), 24 years, SHG member.)*

Because the qualitative data collection was carried out thrice over the two year period, we expected to be able to describe the process of change in the selected villages. However, because very few PLA meetings had been implemented between our first and second round of qualitative data collection, these observations are mainly based on comparison between the second and third rounds.

The qualitative investigator observed that active involvement in Gram Varta had, to a limited extent, helped in busting some myths and traditional ideas on nutrition and care of pregnant women and children. During the first two rounds of qualitative data collection, responses suggested that women continued to believe in traditional myths, for instance, that pregnant women were to be denied certain food items and were discouraged to have a full nutritious meal, because doing so was believed to increase the risk of difficulties during the delivery. However, during focus group discussions in the third round of qualitative data collection, a woman shared that they used to follow this practice but then they learned from Gram Varta meetings that such practices were wrong.

Respondents who reported attending meetings regularly showed an improved understanding of a few HNWASH practices (such as handwashing and use of toilets). However, they reported that while they did change their behaviours initially, they did not continue to follow the healthier practices in their daily life. The reasons cited for giving up on the practises were: difficulty in changing everyday habits, lack of interest, and motivation to follow these habits regularly, and a lack of collective social support.

From an FGD, Round 3:
"*We learnt that we should not go to the fields to attend nature's call. The didi taught us that we should at least cover the faeces with soil and wash our hands after going... we did these things for a couple of days...after that we stopped doing this ..we couldn't see the point in doing these things as everyone in the village has been going to the field for years ..life in villages will always be like this*."

While this is not a change, this shows how the recent demonstrations on the effects of open defecation and other hands-on sessions were better recollected during the third round of qualitative data collection rather than meetings held weeks or months earlier which had included stories or talks. Please note that the third round of qualitative study was carried out a

month after the Gram Varta meetings had been completed and the meetings on sanitation were the last few meetings in the cycle.

"*Let me see.....the didi here called us for meeting and told us many things on health.....(Interviewer : What did she tell you?)....She told us many things..I don't remember them... Yes..I remember the games we played..had so much fun..*

Respondent 2: I remember that there was a public meeting: they demonstrated the effects of open defecation...I cannot forget it. ...it was so vivid. I don't remember much about what didi told us..but I cannot forget these..” FGD, Round 3

Among the indirect effects of Gram Varta were the changes reported by the facilitators. Several of them confided in the qualitative investigator that they had noticed an improvement in their own level of confidence and leadership skills.

Another change noticed by the qualitative investigator was that several SHG members belonging to historically oppressed caste groups had withdrawn from attending Gram Varta meetings. Perhaps this reflected the realities of hierarchial caste relations and tensions between caste groups in the villages visited.

Anganwadi and ASHA workers interviewed by the qualitative investigator reported increased demand for their services. However, this was not supported by the findings in the quantitative analysis. It is possible that these were among the few service providers who noticed an uptick in demand in their areas.

Our qualitative investigator found that the younger women and teenagers (versus older women) in the villages exhibited greater level of curiosity and engagement with Gram Varta meetings from the beginning. They had exhibited better recall of messages during the second round of qualitative inquiry. However, by round 3, it was observed that while the younger women demonstrated better understanding of Gram Varta messages, they also reported difficulty in breaking traditional patriarchal norms on decision-making.

In conclusion, the qualitative study found evidence of some health-promoting impact of Gram Varta (which fits with the regression results being in the right direction but not statistically significant), but the entire potential of Gram Varta was not realized in the villages included in the qualitative study sample. We posit that PLAs, in conjunction with regular home visits, might bring a better outcome in the context of resource-poor areas such as Madhepura.

8.2 Concerns regarding external validity

We have viewed monthly progress report data from Jeevika’s implementation of Gram Varta across Madhepura. The data from Jeevika gives us an idea of the distribution of attendance by various stakeholders such as women, men, members of various castes, communities, men, elderly persons, and frontline service providers. This data fits our impression regarding the lack of attendance of some major stakeholders. In some cases, it supports the lack of impact we have seen, for example in the engagement by frontline workers and their service delivery.

We also conducted an extensive round of qualitative data collection from districts other than Madhepura. We interviewed Gram Varta implementers and beneficiaries at the district, block and village levels across implementing areas of WDC and Jeevika. A separate report of this qualitative study has been previously submitted as part of our qualitative research report. A few salient findings which were consistently reported in several interviews are reproduced here:

1) Attendance of men and elderly people was reported to be poor by most other implementers and beneficiaries. The main reason cited by the respondents was that the SHG is considered to be a strictly female domain where male participation is not desired.
2) Service providers across these implementing areas were reported to be unaware of the Gram Varta implementation and their role in it.
3) In a few areas, funding delays and unplanned changes in the implementation had led to breaking of the continuum of the series of meetings and in mixing the meetings without adequate recaps. This could be a possible reason why the meetings have not had as great an impact as desired.

Our data collection from other districts in Bihar allows us to triangulate our findings from our primary survey, and more importantly, allows us to include detailed interactions with officials who have overseen the implementation of Gram Varta across major districts and blocks since its inception in those districts. The findings from both of our qualitative studies are similar, supporting the fact that men were not a part of the process and were reluctant to allow women to take the lead in undertaking major change initiatives in the community. Some elders also expressed discomfort regarding women attending these meetings and challenging the conventional wisdom of the elders in terms of tending to children and taking care of them.

The quantitative monitoring data we obtained from government departments such as ICDS, the health department, and the education department, did not meet certain conditions which would have allowed their use in evaluating the impact of Gram Varta in districts other than Madhepura. In some cases, the data was not at the level of detail we expected. In some other cases, for example the ICDS, the data we obtained was for the period until March 2013 and not meaningful for our analysis, as the Gram Varta intervention began in Bihar after March 2013.

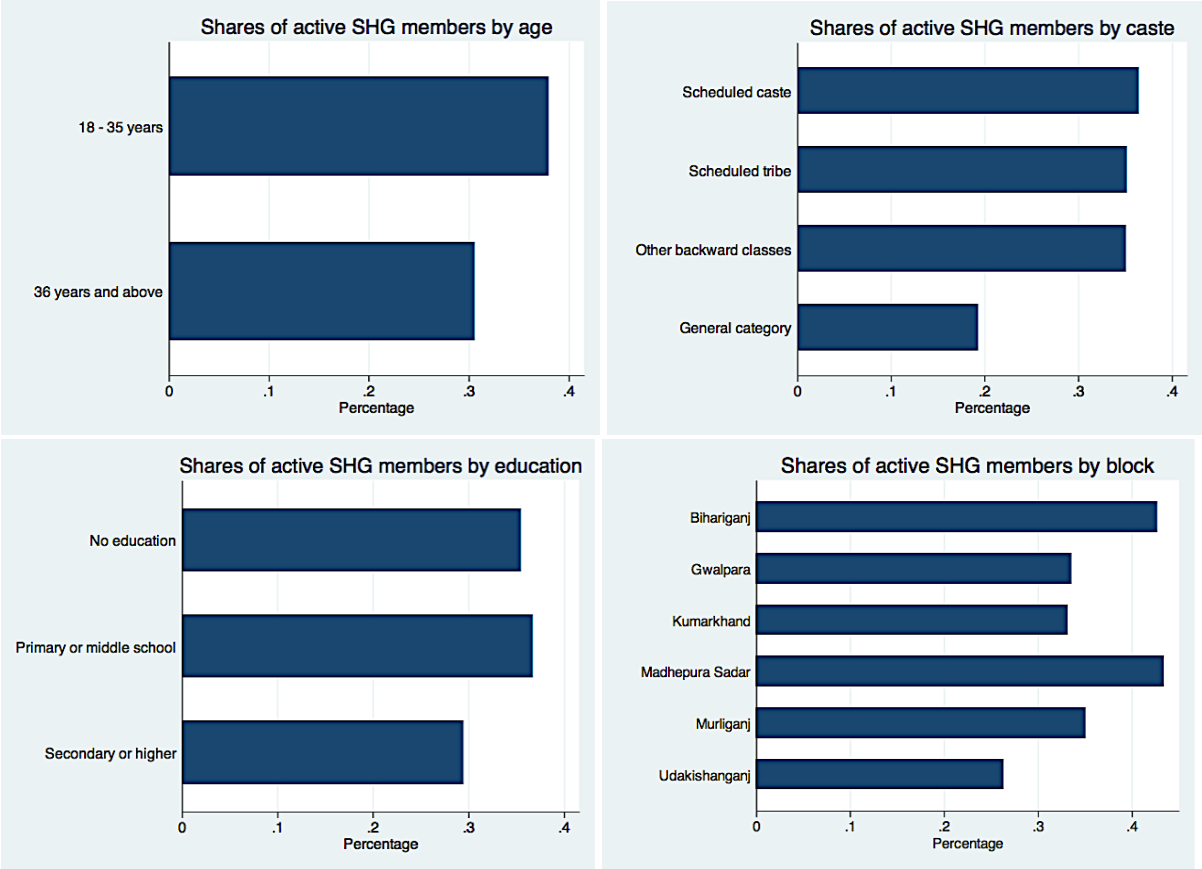
8.3 Concerns regarding implementation and uptake

In the household sample, out of 3,577 household heads (or their proxy respondents), only 182 stated that they had heard of Gram Varta in the endline survey. This is just above 5 percent of respondents and the rate is not much higher for the women respondents. Only 281 out of 3153 respondents had heard of Gram Varta, a rate of 8.9 percent. Given that the first part of the questionnaire was answered by the household head, who is not necessarily a woman, it is reasonable to find a slightly higher rate among respondents of the second part. Since the programme is based on women's SHGs, women can be expected to be more aware of the programme than men. Comparing treatment and control areas, the awareness rates are 10.7 percent and 7.2 percent respectively.

Given these low rates one might suspect that SHGs are unimportant or rare in the setting of the implementation and therefore they might be the wrong vehicle to spread messages.

However, 1360 out of 3151 respondents of the woman questionnaire (43 percent) claimed to be members of SHGs, of which 92 percent were Jeevika SHGs, both in treatment and control areas. About sixty-one percent of SHG members say they attend meetings regularly, underlining that participation in these groups is an integral part in these respondents' lives. Only less than 14 percent of respondents reported attending meetings rarely. Figure 8.1 shows the share of active Jeevika SHG members out of all women respondents by different categories.

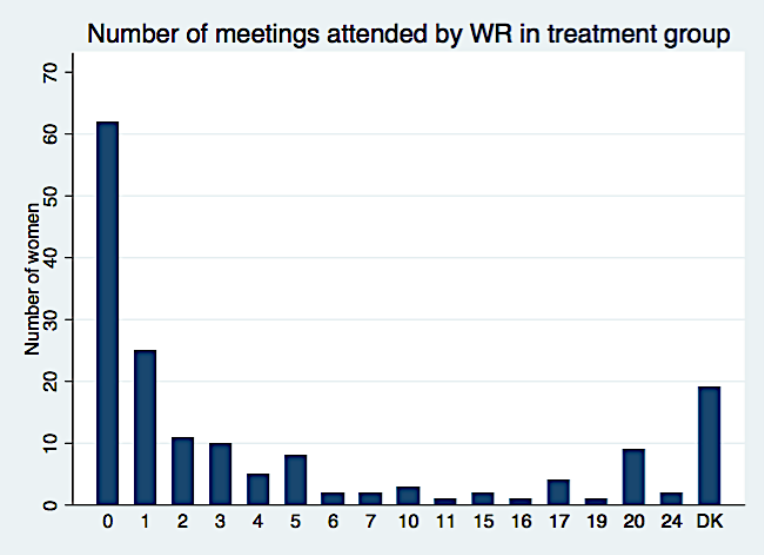
Figure 8.1: Shares of active Jeevika SHG members



One explanation for the observed low awareness of Gram Varta could be that during implementation, the name of the programme was not used everywhere or promoted in any way. This was corroborated by Jeevika officials in Madhepura. Jeevika did not have sufficient funding to publish materials such as banners, posters, and flyers, which could have improved the visibility of Gram Varta. Another concern they shared was that during most government programmes, the District Magistrate, who is the administrative head of the district, usually inaugurates the programme in the presence of relevant department officials. Such a joint event, with all relevant stakeholders, such as health department officials, district ICDS officials and others, could have created a more conducive atmosphere for successfully publicizing and implementing Gram Varta. Notably, the facilitators of the pre-structured sessions who called for the meetings were SHG members themselves. Participants might not have been aware that a specific programme was being implemented or might not have attributed these sessions to the name 'Gram Varta.' Enumerators were specifically trained not to explain the term "Gram Varta" during the interview. Furthermore, out of those who had heard of the program, 64 did not attend any meetings. Of the household head repondents, only 8 claimed that they attended

all 20 sessions, while 42 per cent attended one to five meetings. One hundred and ten female respondents did not attend any meetings, and 98 (35 percent) attended one to 5 meetings. Figure 8.2 shows the number of women respondents attending a given number of meetings.

Figure 8.2: Number of meetings attended



Qualitative findings also suggest problems in implementation of Gram Varta in Madhepura. Ideally, there should have been a 15 day gap between meetings during PLAs 1 to 15, and the entire cycle meetings were supposed to have been conducted in about eight months. However, the responses from the facilitators imply that these meetings were conducted in 3 to 6 months, in a highly erratic manner. Long gaps mean that participants may lose their enthusiasm about the programme, while short intervals between meetings do not leave enough time for participants to reflect and grasp the meetings' messages. This irregularity was mainly due to major changes in decisions about who would facilitate Gram Varta meetings in these villages. Implementing personnel were changed twice before finalizing the community mobilizers as the facilitators. Additionally, there were delays due to the need to replace male CMs with female CMs. This resulted in postponements of implementation, erratic training, and resultantly irregular implementation. Three FGDs in one tola in a treatment village indicated that some of the PLA meetings were not conducted in some SHGs. Moreover, in one treatment village, PLA 1 to 3 were avoided altogether as the CM was not trained for these meetings. Thus, the villagers in this area shared in the FGDs that they did not have a clear idea of Gram Varta. Further increasing this problem, district Jeevika officials in Madhepura reported that there were severe delays in remunerating the field cadres, which continued after the completion of the project rollout in treatment areas. This may have led to low levels of motivation among the field cadres to implement their meetings with the enthusiasm they merited.

The stated goal was that meetings would be conducted at a time that was convenient for community members. However, delays in implementation meant that Gram Varta meetings were conducted at the time of harvesting, one of the busiest times of the year for women (a finding in our qualitative interviews with women). This also discouraged women from attending the meetings as they preferred to earn income through harvesting rather than attend the meetings.

The attitudes of the facilitors also seem to have influenced community's interests in Gram Varta participation. If the facilitator was perceived as not very motivated, then the community members did not engage with Gram Varta very much. For instance, in one of the FGDs, the following response was noted:

"*The CM had called us for the meeting on discussing health issues and had asked us to assemble near the school building by 1 pm. However, she came to the meeting only by 4 pm, by which most of the people had left the meeting." (FGD, SHG member, Village 1)*

Another critical factor is the lack of convergence among and participation of frontline workers such as the ASHA, ANM, and AWW (as evidenced by reports from Jeevika officials in qualitative studies and the numbers included in the monthly progress reports). As several of these workers were unaware of this implementation, they were unlikely to respond to their concerns with the promptness needed. This led to a piecemeal implementation where a critical link of the causal chain and the increased demand leading to increased uptake of services remained weak. The community was likely demotivated by the lack of response from these service providers, which may have discouraged community members from seeking services.

Taken together, we have quite severe concerns about implementation fidelity caused by inconvenient meeting times and delays in implementation due to structural and financial problems in the implementing agency Jeevika and the inadequate quality of work by facilitators. Based on our quantitative and qualitative evidence, treatment group participants have scarcely heard of Gram Varta and facilitator performance was not adequate.

Among implementers, we presented our findings to the district team at Jeevika in Madhepura (Feb 8, 2017), to the state project officials of Jeevika (Feb 10, 2017), and to WDC (Feb 9, 2017). Although most of them were anticipating a larger magnitude of impact in some indicators, the most prominent one being sanitation; in general they agreed with the findings (this was before we accounted for multiple testing). The implementers expressed surprise that sanitation-related indicators had not changed significantly, since the Swachh Bharat Abhiyan was also targeting sanitation indicators. The Jeevika team in Madhepura expressed surprise at the lack of impact in the area of prenatal vaccines especially tetanus toxoid. However, the implementers had anticipated that a one-time intervention was unlikely to have a drastic impact on deep-rooted and intricate social practices. Moreover, the implementers also noted two major factors which may have limited the effectiveness of Gram Varta. One was the lack of involvement of men and other decision-makers in the family and the second was the lack of convergence at all levels and departments of government. The Madhepura Jeevika team explained that because Gram Varta meetings were held along with regular SHG meetings and were facilitated by a regular Jeevika staff (CM) and not a dedicated Gram Varta facilitator (such as a Jeevika Saheli), they anticipated that the impact would be lower than what might be expected.

Since Gram Varta aims to make a long-term impact and enable participants to think critically about their health challenges to initiate a change in behaviour, the implementers and the evaluation team strongly believed that an additional survey is needed to capture the long term impacts. They think that this should be conducted approximately a year after the implementation ended. This is supported by the table summarizing the results according tot the stages depicted in the ToC. While there are a few encouraging results on empowerment

and a suggestion of change related to awareness and practices, evidence related to the last stage, the ultimate impact on health outcomes, was not visible.

This study has been able to answer some major questions related to the implementation of a PLA-based programme in a large area with the government as the implementer. Such a large-scale RCT-based evaluation of Gram Varta had not been conducted after the Ekjut trial. As our study was a randomized, longitudinal survey, we were able to assess the changes made by Gram Varta in comparison to other control areas. Although a long term impact was not assessed, as the final survey was conducted only about one to two months after implementation, we have reported findings on the immediate effects of Gram Varta in many areas. The Ekjut trial had the major motive of reducing IMR and NMR, and Gram Varta too was expected to contribute to improvements in these indicators. However, to effectively assess the long-term impact, future studies will have to include a survey about a year after the implementation is completed. Lastly, Gram Varta in Madhepura was met with serious implementation hurdles, which may have reduced its impact. Future studies might be better able to measure the impact of such a treatment which has high implementation fidelity.

8.4 Concerns regarding programme design

Gram Varta focused on a wide range of topics related to HNWASH, while more successful PLA interventions focused specifically on maternal and perinatal care. As a result of this multitude of topics, not all can be covered to the same extent in one meeting cycle. There might not be enough time to clarify doubts of participants, address their concerns or questions, and use a pace of learning appropriate for less educated rural women. Participants might have difficulties internalizing many messages regarding different topics over one meeting cycle. Moreover, the action plan developed by participants is likely to prioritise specific problems. Groups are bound to choose different activities and address different local concerns, which means that the overall impact on all HNWASH indicators may be small.

Gram Varta was implemented in existing women SHGs. In contrast, other trials of PLA programs, created women's groups specifically for the intervention. Gram Varta's approach has the advantage that no parallel system is created. The operational cost of the programme is therefore low and the programme is potentially sustainable because the groups are well established in the community and are likely to continue their activities even after the implementation is wrapped up. On the other hand, the creation of new groups has the advantage that it might raise more awareness and curiosity in the villages at the beginning of the programme. Existing groups might reflect social hierarchies and discrimination that is prevalent in the communities. They may have their regular members that might be unwilling to include outsiders, for example men, adolescent girls or women from other castes. Newly established groups with the main purpose of implementing the PLA meeting cycle present a fresh start and can draw members from all population groups in the community. This could lead to a more diverse membership in the meetings which facilitates mobilization of the entire community.

The Gram Varta programme did not include a component of strengthening the supply side. One proposed outcome of the theory of change is that demand and utilization of services increases. This necessitates that service providers are actually able to manage this demand and are sufficiently equipped and staffed. If this is not the case, the community's trust in the

service providers might be low, resulting in reduced willingness to use the services or demand changes.

9. Specific findings for policy and practice

9.1 Programme and implementation

The programme mainly seeks to influence women's empowerment and critical thinking, to inspire them to work with their families and community to bring about changes in HNWASH. We saw some evidence of impact on indicators of theorized drivers of change. Such as an increase in women's social capital and self-confidence; reduction in controlling behaviour by husbands; reduction in frequency and intensity of violence by husbands; increase in women's acceptance and mindfulness of their pregnancy; and weak evidence of improved trust in the community. Qualitative data suggest that several PLA activities were memorable and that some participants benefited from the health information, resulting in them making changes in their health practices. The qualitative inquiry also uncovered the improved sense of empowerment and confidence reported by the CMs.

The key learnings that emerge from our mixed methods study are that a PLA approach, implemented through existing SHGs and supported by the state, is able to change women's social capital, self-confidence, and reduce the likelihood of them experiencing controlling behaviour by husbands or even domestic violence. Such an approach can lead to an increase in participants' sense of trust in their communities. All of these effects are likely to be very beneficial in low-resource settings with historically high levels of patriarchy. However, whether such changes lead to improved health knowledge, practices and outcomes, can only be known by investigating the long-term impact of Gram Varta participation.

The lack of a consistent effect on a majority of health indicators is perhaps explained by low levels of implementation fidelity (resulting in ineffective programme delivery), and the need for greater time to observe changes in health and other outcomes. The instances where we do find an impact could be interpreted as the beginning of an impact that might manifest fully in the longer term.

Our qualitative study suggests that a longer-term follow-up might reveal some impact of Gram Varta (regardless of statistical significance), especially in indicators such as feeding young children a balanced diet, maintaining cleanliness, and stopping open defecation. Our observations and interviews suggested that participants considered the health of their children (as opposed to their own health) as important. Moreover the work by Anganwadi Sevikas and sustained political propaganda is likely to keep the momentum going. Moreover, even if there were changes triggered in these or other indicators targeted by Gram Varta, it is likely to take a few months before measurable changes are noted in several health indicators.

Our evaluation results suggest some lessons related to bottlenecks along the causal chain. These findings are based on data from monthly progress reports and our qualitative interviews with stakeholders and participants. If these bottlenecks can be addressed and implementation

therefore improved, we expect that similar programs can achieve better results and greater success.

Implementing agencies faced a lack of participation by men from the community. The participation of men is an important component of the program, since change in behaviour and change in social norms are at its core. Some women might need the agreement of their husband to participate in SHG meetings and they need their agreement and support for implementing changes in their households. Men who participate in PLA meetings might be more willing to work for a change with their wives and families compared to men who do not participate, and are therefore less aware of the topics discussed and lessons learned through PLA. Facilitators were trained to also inform men about PLA meetings and encourage their participation. However, this does not appear to have been very successful. It is unclear whether the reason is a lack of information and encouragement by facilitators or a lack of interest by the men. Facilitators training should better highlight the need for encouraging men to participate.

SHG members showed a lack of trust or willingness to go to the government health centres and Anganwadi centres. At stage 5 of the theory of change, SHG members need to mobilize their communities to make changes and take action on priorities identified in the PLA meetings. For several priorities these activities are likely to require coordination with and cooperation of government health services. Without the trust and willingness to work together, plans might not be implemented and the expected results will not be realized. Involvement of service providers and frontline workers in the Gram Varta process could help increase community members trust in these services. It could also improve the providers willingness to cooperate if they are acknowledged as key actors and engaged and informed about the Gram Varta process.

There was a lack of coordination among the government departments, such as the public health engineering department, ICDS, department of health, and the implementation agencies, who together set up a Steering Committee to monitor the implementation of Gram Varta. District officials of Jeevika and WDC reported that there was no regular communication between the relevant departments. The Steering Committee could constitute a useful platform to engage in regular exchange about the status of the implementation and potential ongoing challenges that need to be addressed. We recommend a regular exchange through routine updates instead of need-based communication.

Monthly progress reports showed gaps between PLA meetings, which are larger and shorter than the prescribed fortnightly gap and implementation was considerably slower than planned. This might have caused a loss of continuity. Long breaks mean that weeks can pass between inputs on different topics discussed, especially since SHG members do not attend every meeting. The lack of regularity might reduce the impact that PLA meetings have on the hypothesised feeling of empowerment and development of critical thinking skills. However, this empowerment is crucial for change to be implemented and results to be realized. On the other hand, the delay forced later meetings to be held in short succession. As a result, participants may not have enough time to reflect on and process the messages of the meetings. It is difficult to enforce regularity of PLA meetings, since participants are meant to decide themselves about the place and time of every meeting. The more people who are willing to participate, the more difficult it might be to agree on a time. However, giving the participants decision power about the timing of meetings increases their ownership of the process. To increase continuity of the

programme and discussions, facilitators could be encouraged to initiate conversations about scheduled topics with SHG members between meetings of the whole group (see success story in qualitative report). Conversations outside of the meetings could also help disseminate messages to community members who are not able to participate in the meetings due to other obligations; including adolescent girls attending school. Facilitators should encourage Gram Varta participants to discuss the scheduled topics with other family members and friends between meetings.

Regarding the programme design, we make the following suggestions. Depending on which health, nutrition, hygiene or sanitation issue seems most pressing, it might be useful to focus a meeting cycle on a specific topic. The same platform could be used for further meeting cycles on other topics, but evidence suggests that a lack of focus weakens the programmes impact. Furthermore, if existing SHGs are used for implementing the PLA meetings, it is crucial to encourage participation of non-members. This includes adolescent girls, men, and service providers. It should be made clear that Gram Varta is not a programme exclusively offered to SHG members, but that messages should be disseminated to all community members, stimulating community solidarity. A component of strengthening the supply side of health care could support the programmes success. This may include training for frontline workers and staff in health care facilities or provision of equipment and medicine.

9.2 Policy

The actual implementation of Gram Varta in Madhepura differed from the implementation plan and from our understanding of the Gram Varta implementation in other disctricts. Some of these differences that arose were due to lack of coordination between implementing agencies and government offices. Instead of a need-based communication, we recommend regular exchange between implementing agencies and relevant government officials. This ensures that service providers are on track about the status of the implementation, that shortcomings or barriers to implementation can be addressed without major delay, and there is a continued feeling of support for the programme without loss of enthusiasm.

In general, Gram Varta showed potential for empowering women, but did not (at least when measured right after the implementation) succeed in improving HNWASH outcomes. One explanation for the lack of consistent effects on HNWASH and other indicators could be that Gram Varta was not successful in bringing health providers and husbands along, who are the ones that – despite improvements in female empowerment – make relevant decisions for the entire household. Similar programmes like Gram Varta are recommended to engage Anganwadi workers, other stakeholders and men early with the core concepts of the programme and ensure their openness to change.

Online appendixes

Note to the reader: Online appendixes are provided as received from the authors. This appendix has not been copy-edited or formatted by 3ie.

Online appendix 1: Tables can be accessed here.
http://www.3ieimpact.org/sites/default/files/CPW01-GV-Appendix-1.pdf

Online appendix 2: Sample design can be accessed here.
http://www.3ieimpact.org/sites/default/files/CPW01-GV-Appendix-2.pdf

Online appendix 3: Descriptive statistics can be accessed here.
http://www.3ieimpact.org/sites/default/files/CPW01-GV-Appendix-3.pdf

Online appendix 4: Results of SHG member analysis can be accessed here.
http://www.3ieimpact.org/sites/default/files/CPW01-GV-Appendix-4.pdf

Online appendix 5: Statistical power can be accessed here.
http://www.3ieimpact.org/sites/default/files/CPW01-GV-Appendix-5.pdf

Online appendix 6: Cost data can be accessed here.
http://www.3ieimpact.org/sites/default/files/CPW01-GV-Appendix-6.pdf

Online appendix 7: Pre-analysis plan can be accessed here.
http://www.3ieimpact.org/sites/default/files/CPW01-GV-Appendix-7.pdf

References

Azad, K., S. Barnett, B. Banerjee, et al. 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: 1193–1202.

Barnard, S., P. Routray, F. Majorin, et al. 2013. Impact of Indian Total Sanitation Campaign on latrine coverage and use: a cross-sectional study in Orissa three years following programme implementation. *PLoS ONE*, 8: e71438.

Bihar Rural Livelihoods Promotion Society, 2002. *Poverty and Social Assessment. A Districtwise Study of Bihar.* Available at: http://www.brlp.in/documents/11369/125465/Poverty+and+Social+Assessment+-+A++districtwise+study+of+Bihar.pdf/8a3c0c07-7021-45c8-8dbb-cb5d6190a9e3.

Bommer, C., M. Subramanyam, and S. Vollmer, 2015a. Evaluation of Gram Varta in rural Bihar – A mixed methods study. Analysis plan. Internal document.

Bommer, C., M. Subramanyam, and S. Vollmer, 2015b. Evaluation of Gram Varta in rural Bihar – A mixed methods study. Baseline report. Internal document.

Cohen, J., 1988. *Statistical power analysis for the behavioral sciences*, 2nd edition., Hillsdale: Lawrence Erlbaum Associates.

Colbourn, T., B. Nambiar, A. Bondo, et al. 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 randomised controlled effectiveness trial. *International Health*, 5: 180-195.

Curtis, V., L. Danquah, and R. Aunger, 2009. Planned, motivated and habitual hygiene behaviour: an eleven country review*. Health Education Research*, 24: 655–673.

Daniel, E.E., R. Masilamani, and M. Rahman (2008). The effect of community-based reproductive health communication interventions on contraceptive use among young married couples in Bihar, India. *International Family Planning Perspectives*, 34(4): 189-197.

Datta, U., 2015. Socio-Economic Impacts of JEEViKA: A Large-Scale Self-Help Group Project in Bihar, India. *World Development*, 68: 1–18.

Directorate of Census Operations, B., 2012. *District Census Handbook – Madhepura.* Available at: http://www.censusindia.gov.in/2011census/dchb/1011_PART_B_DCHB_MADHEPURA.pdf.

Duflo, E., Glennerster, R. and Kremer, M. 2007. Using randomization in development economics: a toolkit, *Handbook of Development Economics (4)*, 3895-3962.

Fink, G., McConnell, M., Vollmer, S. 2014. Testing for Heterogeneous Treatment Effects in Experimental Data: False Discovery Risks and Correction Procedures. *Journal of Development Effectiveness*, 6(1): 44-57.

Fottrell, E., K. Azad, A. Kuddus, L. Younes, S. Shaha, and T. Nahar, 2013. The effect of increased coverage of participatory women's groups on neonatal mortality in Bangladesh: a cluster-randomised trial. *JAMA Pediatrics*, 167(9): 816-825.

Freire, P. 1973. *Education for Critical Consciousness.* New York: Seabury Press.

Government of India, Central Statistical Organization, Ministry of Statistics and Programme Implementation, 2015. *Millennium Development Goals India Country Report 2015.* Available at: http://mospi.nic.in/sites/default/files/publication_reports/mdg_2july15_1.pdf, accessed February 17, 2017.

Government of India, Ministry of Women and Child Development, 2016. *National Policy for Women 2016: Articulating a Vision for Empowerment of Women.*

Government of India, 2013. *Annual Health Survey 2012-13. Bulletin Bihar.* Available at: http://www.censusindia.gov.in/vital_statistics/AHSBulletins/AHS_2nd_Updation_Bulletins.html.

Imai, K., King, G. and Stuart, E. A. 2008. Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society,* Series A 171(2), 481–502.

International Institute for Population Sciences (IIPS) and Macro International, 2016. *National Family Health Survey (NFHS-4), 2015-16: State Fact Sheet, Bihar*. Mumbai: IIPS.

Kar, K. 2003. Subsidy or Self-Respect? Participatory Community-Led Total Sanitation in Bangladesh. *IDS Working Paper*, 184, Brighton, UK.

Kar, K., and K. Pasteur, 2005. Subsidy or Self-Respect? Participatory Community-Led Total Sanitation. An update on recent developments. *IDS Working Paper*, 257, Brighton, UK.

Lewycka, S., C. Mwansambo, M. Rosato, et al. 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: 1721–1735.

Manandhar, D.S., D. Osrin, B. Shrestha, et al. 2004. Effect of a participatory intervention with women's groups on birth outcomes in Nepal: cluster-randomised controlled trial. *The Lancet*, 364: 970–79.

Minkler, M. and K. Cox, 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.

Nair, N., P. Tripathy, A. Prost, A. Costello, D. Osrin, 2010. Improving Newborn Survival in Low-Income Countries: Community-Based Approaches and Lessons from South Asia. PLoS Med 7(4): e1000246. doi:10.1371/journal.pmed.1000246. Available at: http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1000246

O'Rourke, K., L. Howard-Grabman, G. Seoane, 1998. Impact of community organization of women on perinatal outcomes in rural Bolivia. *Rev Panam Salud Public,* 3: 9-14. Available at: http://www.scielosp.org/scielo.php?pid=S1020-49891998000100002&script=sci_arttext

Population Census, 2011. http://www.census2011.co.in/

Prost, A., T. Colbourn, N. Seward, et al., 2013. Women's groups practicing participatory learning and action to improve maternal and newborn health in low-resource settings: a systematic review and meta-analysis. *The Lancet,* 381(9879): 1936-1746. Available at: http://www.sciencedirect.com/science/article/pii/S0140673613606856

Shah, M.S., A. Ahmad, N. Khalique, et al. 2011. Home-based management of acute diarrhoeal disease in an urban slum of Aligarh, India. *The Journal of Infection in Developing Countries*, 6(02): 137-142.

Tamang, S., N. Mesko, B.P. Shrestha, et al. 2001. A Qualitative Description of Perinatal Care Practices in Makwanpur District, Nepal. *Contributions to Nepalese Studies*, 28: 10-19.

Tripathy, P. et al., 2010. Effect of 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), pp.1182–1192. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0140673609620420.

United Nations Development Programme, 2009. *Kosi Floods 2008. How we coped! What we need? Perception Survey on Impact and Recovery Strategies.* UNDP, New Delhi, India.

Vijayalakshmi, N., Sharma, N. & Srivastava, S., 2016. Gram Varta: Village Dialogue: A Women's Group Led Community Empowerment Model for Improved Nutrition, Health, Sanitation and Hygiene outcomes in Bihar. *CAPAM*. Available at: https://www.capam.org/files/2016IIApresentations/GramVarta_VillageDialogue-India.pdf.

Wallerstein, N. and E. Bernstein, 1994. Introduction to community empowerment, participatory education, and health. *Health Education Quarterly*, 21(2): 141-148.

Wallerstein, N. and E. Bernstein, 1988. Empowerment education: Freire's ideas adapted to health education. *Health Education and Behavior*, 15(4): 379-394.

Watts, R.J., M.A. Diemer, and A.M. Voight, 2011. Critical consciousness: Current status and future directions. *New directions for child and adolescent development*, 134: 43-57.

Wetmore, S.B. and F. Theron, 1998. Community development and research: Participatory learning and action-a development strategy in itself. *Development Southern Africa*, 15(1): 29-54.